ECRIN Metadata Schema for Clinical Research Data Objects, Version 4 (September 2020)

Appendices – Categorised Questions

Steve Canham, 14/09/2020

Many of the fields in the ECRIN metadata scheme are categorised, i.e. they are constrained to hold one of a predefined set of possible values. The following tables list these values.

In an implementation of the schema, e.g. using a database to hold the categories as lookup tables, each option would have an id, normally an integer, and it would be the id rather than the name which would appear in the various tables making use of these categories.

Contents

[Study types 3](#_Toc50989148)

[Study statuses 4](#_Toc50989149)

[Gender Eligibility Types 5](#_Toc50989150)

[Feature Types 5](#_Toc50989151)

[Feature Values 6](#_Toc50989152)

[Study relationship types 11](#_Toc50989153)

[Object classes 13](#_Toc50989154)

[Object types 15](#_Toc50989155)

[Object filter types 24](#_Toc50989156)

[Object access types 26](#_Toc50989157)

[Object instance types 27](#_Toc50989158)

[Object relationship types 28](#_Toc50989159)

[Contribution types 30](#_Toc50989160)

[Dataset consent types 34](#_Toc50989161)

[Dataset de-identification levels 35](#_Toc50989162)

[Dataset record key types 35](#_Toc50989163)

[Date types 36](#_Toc50989164)

[Description types 38](#_Toc50989165)

[Identifier types 39](#_Toc50989166)

[Language codes 41](#_Toc50989167)

[Resource types 44](#_Toc50989168)

[Time Units 45](#_Toc50989169)

[Title types 46](#_Toc50989170)

[Topic types 48](#_Toc50989171)

[Topic vocabularies 49](#_Toc50989172)

## Study types

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Interventional | A clinical trial. | ClinicalTrials.gov |
| Observational | Any form of non-interventional research. | ClinicalTrials.gov |
| Observational Patient Registry | Collecting data for a designated registry. | ClinicalTrials.gov |
| Expanded access | Off label usage of a new product for individuals. | ClinicalTrials.gov |
| Funded programme | With a single or linked series of grants. | ClinicalTrials.gov |
| Not yet known | Dummy value supplied by default on entity creation. | ClinicalTrials.gov |

## Study statuses

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Withdrawn | Study halted prematurely, prior to enrolment of first participant. | ClinicalTrials.gov |
| Available | Recruitment status available on request but no details provided. | ClinicalTrials.gov |
| Withheld | Recruitment status not provided. | ClinicalTrials.gov |
| Recruiting | Participants are currently being recruited, whether or not any participants have yet been enrolled. | ClinicalTrials.gov |
| Active, not recruiting | Study is continuing, meaning participants are receiving an intervention or being examined, but new participants are not currently being recruited or enrolled. | ClinicalTrials.gov |
| Not yet recruiting | Participants are not yet being recruited. | ClinicalTrials.gov |
| No longer available | Recruitment status no longer available. | ClinicalTrials.gov |
| Suspended | Study halted prematurely but potentially will resume. | ClinicalTrials.gov |
| Enrolling by invitation | Participants are being (or will be) selected from a predetermined population. | ClinicalTrials.gov |
| Approved for marketing | The study has concluded and the medical agent that was the focus of the study has been approved for marketing (in the US). | ClinicalTrials.gov |
| Completed | The study has concluded normally; participants are no longer receiving an intervention or being examined (that is, last participant’s last visit has occurred). | ClinicalTrials.gov |
| Terminated | Study halted prematurely and will not resume; participants are no longer being examined or receiving intervention. | ClinicalTrials.gov |
| Other | Status given as "other" in source data. | WHO |
| Ongoing | Trial is ongoing but recruitment status is unclear. | EU CTR |
| Unknown status | Status information not provided. | ClinicalTrials.gov |

## Gender Eligibility Types

Studies may be open only to female participants, or only male, or both. This categorisation reflects that.

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| All |  | ClinicalTrials.gov |
| Female |  | ClinicalTrials.gov |
| Male |  | ClinicalTrials.gov |
| Not provided |  | ClinicalTrials.gov |
| Unknown status |  | ClinicalTrials.gov |

## Feature Types

The feature types available are listed in the Feature\_Types table, listed below:

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Phase | For a clinical trial of a drug product (including a biological product), the numerical phase of the clinical trial. | ClinicalTrials.gov |
| Primary purpose |  | ClinicalTrials.gov |
| Allocation type | The method by which participants are assigned to arms in a clinical trial. | ClinicalTrials.gov |
| Intervention model | The strategy for assigning interventions to participants. | ClinicalTrials.gov |
| Masking | The party or parties involved in the clinical trial who are prevented from having knowledge of the interventions assigned to individual participants. | ClinicalTrials.gov |
| Observational model | The Primary strategy for participant identification and follow-up. | ClinicalTrials.gov |
| Time perspective | For observational studies, describes the temporal relationship of observation period to time of participant enrollment. | ClinicalTrials.gov |
| Biospecimens retained | Indicates whether samples of material from research participants are retained in a biorepository.' | ClinicalTrials.gov |

## Feature Values

Within the database the available feature categories are all stored within a single table ('study\_feature\_categories'), and are distinguished there by a 'type' field, equivalent to the id of each type. The available categories for each of the feature types are listed below. All are derived from the definitions given by ClinicalTrials.gov.

#### **Phase (interventional trials only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Not applicable | Trials without phases (for example, studies of devices or behavioural interventions). | ClinicalTrials.gov |
| Early Phase 1 | Exploratory trials, involving very limited human exposure, with no therapeutic or diagnostic intent (e.g., screening studies, microdose studies). | ClinicalTrials.gov |
| Phase 1 | Initial studies to determine the metabolism and pharmacologic actions of drugs in humans, side effects, and to gain early evidence of effectiveness; may include healthy participants and/or patients. | ClinicalTrials.gov |
| Phase 1/Phase 2 | Trials that are a combination of phases 1 and 2. | ClinicalTrials.gov |
| Phase 2 | Controlled clinical studies conducted to evaluate the effectiveness of the intervention for a particular indication in participants with the disease or condition under study, and to determine the common short-term side effects and risks. | ClinicalTrials.gov |
| Phase 2/Phase 3 | Trials that are a combination of phases 2 and 3. | ClinicalTrials.gov |
| Phase 3 | Trials conducted after preliminary evidence suggesting effectiveness of the drug has been obtained, and are intended to gather additional information to evaluate the overall benefit-risk relationship of the drug. | ClinicalTrials.gov |
| Phase 4 | Studies of approved drugs to delineate additional information including the drug's risks, benefits, and optimal use. | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Allocation type (interventional trials only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Not applicable | For a single-arm trial. | ClinicalTrials.gov |
| Randomised | Participants are assigned to intervention groups by chance. | ClinicalTrials.gov |
| Nonrandomised | Participants are expressly assigned to intervention groups through a non-random method, such as physician choice. | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Intervention model (interventional trials only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Single group assignment | Clinical trials with a single arm. | ClinicalTrials.gov |
| Parallel assignment | Participants are assigned to one of two or more groups in parallel for the duration of the study. | ClinicalTrials.gov |
| Crossover assignment | Participants receive one of two (or more) alternative interventions during the initial phase of the study and receive the other intervention during the second phase of the study. | ClinicalTrials.gov |
| Factorial assignment | Two or more interventions, each alone and in combination, are evaluated in parallel against a control group. | ClinicalTrials.gov |
| Sequential assignment | Groups of participants are assigned to receive interventions based on prior milestones being reached in the study, such as in some dose escalation and adaptive design studies. | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Primary purpose (interventional trials only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Treatment | One or more interventions are being evaluated for treating a disease, syndrome, or condition. | ClinicalTrials.gov |
| Prevention | One or more interventions are being assessed for preventing the development of a specific disease or health condition. | ClinicalTrials.gov |
| Diagnostic | One or more interventions are being evaluated for identifying a disease or health condition. | ClinicalTrials.gov |
| Supportive Care | One or more interventions are evaluated for maximizing comfort, minimizing side effects, or mitigating against a decline in the participant's health or function. | ClinicalTrials.gov |
| Screening | One or more interventions are assessed or examined for identifying a condition, or risk factors for a condition, in people who are not yet known to have the condition or risk factor. | ClinicalTrials.gov |
| Health Services Research | One or more interventions for evaluating the delivery, processes, management, organization, or financing of healthcare. | ClinicalTrials.gov |
| Basic Science | One or more interventions for examining the basic mechanism of action (for example, physiology or biomechanics of an intervention). | ClinicalTrials.gov |
| Device Feasibility | An intervention of a device product is being evaluated in a small clinical trial (generally fewer than 10 participants) to determine the feasibility of the product, or a clinical trial to test a prototype device for feasibility and not health outcomes. | ClinicalTrials.gov |
| Educational/Counselling / Training | An intervention involving psychosocial or educational input. | ClinicalTrials.gov |
| Other | None of the other options applies. | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Masking (interventional trials only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| None (Open Label) |  | ClinicalTrials.gov |
| Blinded (no details) | From a statement that says the study was blinded, but where the degree of blinding was not provided | ClinicalTrials.gov |
| Single |  | ClinicalTrials.gov |
| Double |  | ClinicalTrials.gov |
| Triple |  | ClinicalTrials.gov |
| Quadruple |  | ClinicalTrials.gov |
| Not applicable | Explicitly labelled as not applicable - usually because the study is non-interventional | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Observational model (observational studies only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Cohort | Group of individuals, initially defined and composed, with common characteristics (for example, condition, birth year), who are examined or traced over a given time period. | ClinicalTrials.gov |
| Case-Control | Group of individuals with specific characteristics (for example, conditions or exposures) compared to group(s) with different characteristics, but otherwise similar. | ClinicalTrials.gov |
| Case-Only | Single group of individuals with specific characteristics. | ClinicalTrials.gov |
| Case-Crossover | Characteristics of case immediately prior to disease onset (sometimes called the hazard period) compared to characteristics of same case at a prior time (that is, control period). | ClinicalTrials.gov |
| Ecologic or Community Study | Geographically defined populations, such as countries or regions within a country, compared on a variety of environmental (for example, air pollution intensity, hours of sunlight) and/or global measures not reducible to individual level characteristics (for example, healthcare system, laws or policies median income, average fat intake, disease rate). | ClinicalTrials.gov |
| Family-Based | Studies conducted among family members, such as genetic studies within families or twin studies and studies of family environment. | ClinicalTrials.gov |
| Defined population |  | ClinicalTrials.gov |
| Natural history |  | ClinicalTrials.gov |
| Other | Explain in Detailed Description. | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Time perspective (observational studies only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Retrospective | Look back using observations collected predominantly prior to subject selection and enrolment. | ClinicalTrials.gov |
| Prospective | Look forward using periodic observations collected predominantly following subject enrolment. | ClinicalTrials.gov |
| Cross-sectional | Observations or measurements made at a single point in time, usually at subject enrolment. | ClinicalTrials.gov |
| Retrospective/Prospective |  | ClinicalTrials.gov |
| Longitudinal |  | ClinicalTrials.gov |
| Other | Explain in Detailed Description. | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

#### **Biospecimens retained (observational studies only)**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| None Retained |  | ClinicalTrials.gov |
| Samples With DNA |  | ClinicalTrials.gov |
| Samples Without DNA |  | ClinicalTrials.gov |
| Not provided | No data was provided in the source record. | ECRIN |

## Study relationship types

Clinical research studies may relate to each other, forming clusters of inter-related research. The relationship types were all developed by ECRIN, and relate to relationships found particularly within clinical research. In almost all cases they come in related pairs, allowing both ‘views’ of a relationship.

| **name** | **description** | **source** |
| --- | --- | --- |
| Is a sub-study of | This study is a sub-study or sub-protocol undertaken at the same time as <the target study>. | ECRIN |
| Includes as a sub-study | Has <the target study> as a sub-study or sub-protocol, undertaken at the same time as this study. | ECRIN |
| Is in the same series as | This study is in a sequence that began with <the target study>. | ECRIN |
| Is the first of a sequence including | This study began a sequence that includes <the target study>. | ECRIN |
| Is a feasibility study for | This study is a feasibility or pilot study for <the target study>. | ECRIN |
| Is preceded by the feasibility study | This study was preceded by <the target study> as a feasibility or pilot study. | ECRIN |
| Is a later phase variant of | For trials using the phase I to IV classification, this study is a later phase continuation of <the target study>. | ECRIN |
| Is an earlier phase variant of | For trials using the phase I to IV classification, this study is an earlier phase precursor to <the target study>. | ECRIN |
| Is a continuation of | This study uses some or all of the same subject population as <the target study>. | ECRIN |
| Is continued by | This study has some or all of the same subject population targeted by <the target study>. | ECRIN |
| Is a repeat of | This study uses a different population but the same or similar protocol as <the target study>. | ECRIN |
| Is repeated by | This study is repeated by <the target study>, with the same or similar protocol but using a different population. | ECRIN |
| has an expanded access version | This study has <the target study> as an expanded access version, (for people who cannot enrol in the trial but who may benefit from the product under investigation). | ECRIN (ClinicalTrials.gov) |
| is an expanded access version of | This study is an expanded access version of <the target study>, catering for people who cannot enrol in the trial but who may benefit from the product under investigation. | ECRIN (ClinicalTrials.gov) |
| Includes target as one of a group of non-registered studies | This study includes <the target study>. That study is not registered independently, but instead shares this registry entry with one or more other non-registered studies. | ECRIN |
| Non registered but included within a registered study group | This study is registered as <the target study>, along with one or more other studies that share the same registry entry and id. | ECRIN |
| Has link listed in registry but nature of link unclear | This study is linked to <the target study> within the registry entry, but the nature of the linkage is not clear. | ECRIN |
| Includes target as one of a group of registered studies | This study includes <the target study>, which is registered elsewhere along with one or more other registered studies, forming a group that collectively equates to this study. | ECRIN |
| Registered and is included elsewhere in group | This study is also registered, along with one or more other studies that together form an equivalent group, as <the target study>. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Object classes

The object class represents the broad type of a data object (and is usually used in conjunction with the more detailed object type, to fully characterise the object). The options available are shown and are identical to the DataCite controlled list for ‘resourceTypeGeneral’. A dummy ‘Not yet known’ option was added, but in general the class should be known for all objects.

| **name** | **description** | **Source** |
| --- | --- | --- |
| Audiovisual | A series of visual representations imparting an impression of motion when shown in succession. May or may not include sound. May be used for films, video, etc. | DataCite |
| Collection | An aggregation of resources, which may encompass collections of one resource type as well as those of mixed types. A collection is described as a group; its parts may also be separately described, e.g. a collection of samples, or various files making up a report. | DataCite |
| Data Paper | A factual and objective publication with a focused intent to identify and describe specific data, sets of data, or data collections to facilitate discoverability. A data paper describes data provenance and methodologies used in the gathering, processing, organizing, and representing the data. | DataCite |
| Dataset | Data encoded in a defined structure. A data file or files. | DataCite |
| Event | A non-persistent, time based occurrence. Descriptive information and/or content that is the basis for discovery of the purpose, location, duration, and responsible agents associated with an event such as a webcast or convention. | DataCite |
| Image | A visual representation other than text, e.g. digitised or born digital images, drawings or photographs. | DataCite |
| Interactive Resource | A resource requiring interaction from the user to be understood, executed, or experienced, e.g. training modules, files that require use of a viewer, or query/response portals. | DataCite |
| Model | An abstract, conceptual, graphical, mathematical or visualization model that represents empirical objects, phenomena, or physical processes, e.g. modelled descriptions of different aspects of languages or a molecular biology reaction chain. | DataCite |
| Physical Object | An inanimate, three dimensional object or substance, e.g. artefacts, specimens. | DataCite |
| Service | An organized system of apparatus, appliances, staff, etc., for supplying some function(s) required by end users, e.g. a data management service, or long term preservation service. | DataCite |
| Software | A computer program in source code (text) or compiled form. Use this type for all software components supporting scholarly research. | DataCite |
| Sound | A resource primarily intended to be heard, e.g. an audio recording. | DataCite |
| Text | A resource consisting primarily of words for reading, includes grey literature, lab notes, accompanying materials as well as published articles. | DataCite |
| Workflow | A structured series of steps which can be executed to produce a final outcome, allowing users a means to specify and enact their work in a more reproducible manner. | DataCite |
| Other | Object class not listed elsewhere. | DataCite |
| Not yet known | Dummy value supplied by default on entity creation in the MDR system. | ECRIN |

In most cases, for clinical research data objects, the class will usually be either Text or Dataset. though other options could include Data Paper, Software, Service, Audiovisual, and Interactive Resource. It is not anticipated that there will be a need to increase the object classes available.

## Object types

The object type provides a more detailed categorisation of a data object, and is a concept taken from DataCite, where it is known as the ‘ResourceType’. Within DataCite the categories are not explicitly defined, but the recommendation is for any type to be combined with the overall class to form a binomial classification, e.g. Dataset / IPD final analysis dataset, or Text / Journal Article.  
  
In the ECRIN system this recommendation is made into a mandatory requirement, i.e. all object types must be associated with a ‘parent’ object class. Of the 117 object types currently defined, 70 are types of text objects, 24 are types of dataset objects, 3 are collection objects, 5 are types of software objects, and 15 are 'other' (i.e. not defined elsewhere) for each of the various object classes. The different types and sources of object types are described below.  
  
**CASRAI text types**

A suggestion in DataCite for Text objects is to use the CASRAI (Consortia Advancing Standards in Research Administration Information) output type list to provide the types (available at <https://dictionary.casrai.org/Output_Types>). This suggestion is followed in the ECRIN schema, with all 35 CASRAI categories being used in the system. They are listed below, with the details being taken directly from the CASRAI web site.

| **name** | **description** | **source** |
| --- | --- | --- |
| Journal Article | Articles in peer-reviewed publications that disseminate the results of original research and scholarship. | CASRAI |
| Book | Books written by a single author or collaboratively based on research or scholarly findings generally derived from peer reviewed funding. | CASRAI |
| Book Chapter | Texts written by a single author or collaboratively based on research or scholarly findings and expertise in a field. | CASRAI |
| Book Prospectus | Document that describes a forthcoming book based on research or scholarly findings. | CASRAI |
| Book Review | Critical review of works of fiction or non-fiction highlighting the contributions to an art, field or discipline. | CASRAI |
| Book Series | Set of related books written by a single author or collaboratively based on research or scholarly findings. | CASRAI |
| Conference Abstract | Texts of a specified length that states the issue to be discussed in a proposed conference paper. It serves as the basis for the acceptance of the paper at a conference. The abstract is published along with the paper. | CASRAI |
| Conference Paper | Papers written alone or collaboratively, presented at an academic conference, and published in the proceedings (not in scholarly journals). | CASRAI |
| Conference Poster | Posters displayed in a conference setting and conveying research highlights in an efficient manner by compelling graphics. They may be peer-reviewed prior to acceptance and be published in the proceedings. | CASRAI |
| Conference Program | Document giving details of papers to be presented at an academic conference, compiled from the accepted submissions. | CASRAI |
| Dictionary Entry | Entries of new words, new meanings of existing words, changes in spelling and hyphenation over a longer period of time, and grammatical changes. | CASRAI |
| Disclosure | Publications that establish inventions as prior art thereby preventing others from patenting the same invention or concept. | CASRAI |
| Dissertation | Treatise advancing an original point of view resulting from research: a requirement for a doctoral degree. | CASRAI |
| Edited Book | Books edited by a single author or collaboratively for the dissemination of research or scholarly findings that generally result from peer reviewed funding. | CASRAI |
| Encyclopaedia Entry | Authored entries in a reference work or a compendium focusing on a particular domain or on all branches of knowledge. | CASRAI |
| Funding Submission | Information about specific requests for funds submitted to potential funders of the activity. The standard allows details to be collected for multiple years. | CASRAI |
| Journal Issue | Periodical publications aimed at fostering intellectual debate and inquiry. Special journal issues are produced by editors with an established record of scholarship in the field and able to provide the direction of the theme. Journal issues bear a unique number of reference for publication. | CASRAI |
| License | Signed agreements to exploit a piece of IP such as a process, product, data, or software. | CASRAI |
| Magazine Article | Articles in thematic publications published at fixed intervals. | CASRAI |
| Manual | Course and assignment materials produced for teaching purposes. | CASRAI |
| Newsletter Article | Articles in publications aimed at researchers, decision-makers, professionals and the public that report on a research project or on the activities of a research chair or a research centre. The Newsletters promote research activities to the community and the university; mobilize knowledge to improve practice and inform policy, and provide relevant and accessible information to the broader public. | CASRAI |
| Newspaper Article | Articles in a daily, weekly or monthly publication reporting on news and social issues aimed at the public. May entail critical analysis based on expertise in the field. | CASRAI |
| Online Resource | Information accessible only on the web via traditional technical methods (i.e. hyperlinks). | CASRAI |
| Patent | A form of IP protection that defines the exclusive right by law for inventors and assignees to make use of and exploit their inventions, products or processes, for a limited period of time. | CASRAI |
| Registered Copyright | Registered ownership of rights under a system of laws for promoting both the creation of and access to artistic, literary, musical, dramatic and other creative works. | CASRAI |
| Report | Reports disseminating the outcomes and deliverables of a research contract. May entail a contribution to public policy. | CASRAI |
| Research Tool | Series of observations, measurements or facts identified from the research. They include bibliographies, indices and catalogues of research collections; concordances and dictionaries; materials that facilitate access to archival holdings or collections such as repository guides, inventories of a group of manuscripts or of a body of archives, inventories or documentary materials, thematic guides to archival materials, records surveys and special indices; scholarly editions; and data series. | CASRAI |
| Supervised Student Publication | Articles on research findings published jointly with or supervised by the thesis advisor. The findings relate to research undertaken by the student or the supervisor’s program of research. | CASRAI |
| Tenure-Promotion | A process of confidential deliberations by a committee, held in-camera, to decide tenure or promotion of an academic. | CASRAI |
| Test | Assessments that include tests designed for general university selection, selection into specific courses or other evaluation purposes. | CASRAI |
| Trademark | Marks such as a name, word, phrase, logo, symbol, design, image of a product or service that indicates the source and provides the right to control the use of the identifier. | CASRAI |
| Translation | Translations of books and articles that identify modifications to the original edition, such as a new or revised preface. | CASRAI |
| University Academic Unit | A primary academic organizational entity of a university that has appointed faculty. | CASRAI |
| Website | Stand-alone locations on the web where multiple types of information on a specific theme are available. May include interactive features for contributions from readers. | CASRAI |
| Working Paper | Preliminary versions of articles that have not undergone review but that may be shared for comment. | CASRAI |

**ECRIN text types**

The additional ECRIN defined text types added, all focused on clinical research, are:

| **name** | **description** | **source** |
| --- | --- | --- |
| Study Protocol | Structured document describing the study, its rationale, methodology, outcome measures etc. | ECRIN |
| Journal article abstract | A journal citation and abstract, usually on-line as a web page, e.g. within PubMed. | ECRIN |
| Trial Registry entry | Summary of the study and its aims, posted prospectively or retrospectively to a public registry. | ECRIN |
| Study Overview | A brief overview of the study, may be an abridged protocol, or as used within the study web site or other study documents. | ECRIN |
| Ethics submission | Documents provided to an ethics review board, often with the protocol, when seeking ethical approval for a study. | ECRIN |
| Ethics approval notification | Documents from an ethical review board confirming that ethical approval has been granted. | ECRIN |
| Regulatory authority submission | Documents provided to a regulatory authority, seeking approval to run a clinical trial. | ECRIN |
| Regulatory authority approval notification | Documents from a regulatory authority confirming that approval to run the trial has been granted. | ECRIN |
| Manual of Operations | Description of specific operations and workflow within the study. | ECRIN (BioLINCC) |
| Manual of Procedures | Description of specific procedures and techniques used within the study. | ECRIN (BioLINCC) |
| Informed consent forms | The form or forms given to participants to formally record their informed consent to study participation. | ECRIN |
| Patient information sheets | The information provided to study participants, especially as part of the consenting process. | ECRIN |
| Data Overview | A summary of the data, without the details of a data dictionary but indicating the nature of different tables, time points of data etc. | ECRIN (BioLINCC) |
| Database specification | Functional specification of the database including details of individual data items, types, ranges, etc. May also contain details of logic and consistency checks, and any derived values. | ECRIN |
| Data Dictionary | A detailed, item by item, description of the data points in the dataset, sufficient for accurate analysis of the data by others. | ECRIN (BioLINCC) |
| Data collection forms | Copies, in electronic and / or paper form, of the case report forms (CRFs and / or eCRFs) used for collecting data. | ECRIN |
| Annotated Data Collection Forms | Data Collection forms (CRFs or eCRFs) annotated to provide further details of each item (e.g. data type, allowable range). | ECRIN (BioLINCC) |
| Standard instruments | Standardised rating instruments, including questionnaires. | ECRIN (BioLINCC) |
| Statistical analysis plan | The details of the proposed analysis for the study, listing the individual descriptive statistics and tests of inference, and their parameters. | ECRIN |
| Analysis Notes | A summary of the analysis carried out and / or any caveats to be borne in mind when interpreting results. Less formal than a statistical analysis plan. | ECRIN (BioLINCC) |
| Protocol SAP | Study Protocol and Statistical Analysis Plan. | ECRIN (ClinicalTrials.gov) |
| Protocol ICF | Study Protocol and Informed Consent Form. | ECRIN (ClinicalTrials.gov) |
| Protocol SAP ICF | Study Protocol, Statistical Analysis Plan, and Informed Consent Form. | ECRIN (ClinicalTrials.gov) |
| Data management plan | A plan for, and record of, data management activities in the study, covering the whole data life cycle. | ECRIN |
| Definitions | A glossary list for use with other study documents. | ECRIN (BioLINCC) |
| Trial master file contents list | A listing of the documents expected within a trial master file, and their organisation, in electronic and / or paper form. | ECRIN |
| Data collection schedule | A document detailing the time points of data collection in a study (or 'visits'). | ECRIN (BioLINCC) |
| Data coding manual | A manual or guide that provides instructions on how to complete and / or interpret scores and codes within a study. | ECRIN (BioLINCC) |
| Data monitoring committee report | A report concerning the safety or efficacy of a study, from independent experts. Often containing recommendations about the continuation of the study. | ECRIN |
| Bibliography | A list of publications making up a bibliography relevant to the study, but not necessarily generated or triggered by the study. | ECRIN (BioLINCC) |
| Introduction to document set | A 'contents', 'readme' or similar document that describes the other documents available. | ECRIN (BioLINCC) |
| Unpublished Study Report | A report of a study, or part of a study, not formally published. May be an internal interim document within a long term study. | ECRIN (BioLINCC) |
| Clinical Study Report | Full end of study report with detailed efficacy and safety results. | ECRIN |
| CSR Summary | Summary of the Clinical Study Report. | ECRIN (ClinicalTrials.gov) |
| Redacted Clinical Study Report | End of study report with some data withheld, usually because of commercial sensitivity. | ECRIN |
| List of web links | A web page that includes a list of links to different items, e.g. individual CRFs. | ECRIN (BioLINCC) |
| Trial registry results summary | Summary of study results, as displayed in a trial registry. | ECRIN |
| Literature Review | Publications referenced within the literature review undertaken prior to the study. | ECRIN (BioLINCC) |
| Publication List | List of publications related to the study. | ECRIN (BioLINCC) |
| Investigational Product Information | Summary of information about a medicinal product. May be a package insert or investigator's brochure. | ECRIN (ClinicalTrials.gov) |
| General background to research topic | Supporting document summarising relevant research and / or research programs, or aspects of condition pathology, epidemiology, etc. | ECRIN (ClinicalTrials.gov) |
| Redacted Protocol | A redacted version of the study protocol. | ECRIN (ClinicalTrials.gov) |
| Redacted SAP | A redacted version of the statistical analysis plan. | ECRIN (ClinicalTrials.gov) |

**ECRIN dataset Types**

In addition, there were 23 clinical research specific dataset objects defined by ECRIN, listed below (IPD = Individual Participant Data).

| **name** | **description** | **source** |
| --- | --- | --- |
| Individual Participant Data | A dataset simply called Individual Participant Data, or its equivalent, with no further qualification or description. | ECRIN |
| IPD final analysis dataset (full study population) | Full final dataset supporting all analyses carried out on the study data. | ECRIN |
| IPD final analysis dataset (supporting specific paper) | Dataset supporting the analyses and conclusions of a specific single paper. | ECRIN |
| IPD final analysis dataset (sub-population) | Dataset with the data from a sub-population of the complete study (e.g. the control arm, or a particular age group). | ECRIN |
| IPD final analysis dataset (sub-study) | Dataset with the data of a sub-study, supporting the analysis only of that sub-study - may also involve a sub-population of the whole. | ECRIN |
| IPD interim analysis dataset (sub-population) | Dataset with the data from a sub-population of the complete study, from an earlier time point than the primary analysis. | ECRIN |
| IPD interim analysis dataset (sub-study) | Dataset with the data of a sub-study, supporting the analysis only of that sub-study, from an earlier time point than the primary analysis. | ECRIN |
| IPD long term follow up analysis dataset | Supplementary dataset with additional data from long term follow up, data collected after primary analysis. | ECRIN |
| IPD interim analysis dataset | Dataset covering the whole study but from an earlier time point than the primary analysis. | ECRIN |
| IPD safety analysis dataset | Dataset of individual data supporting comprehensive safety analysis. | ECRIN |
| IPD PK analysis dataset | Dataset of pharmaco-kinetic data. | ECRIN |
| IPD PD analysis dataset | Dataset of pharmaco-dynamic data. | ECRIN |
| IPD quality of life analysis dataset | Dataset supporting analysis of quality of life measures within the study. | ECRIN |
| IPD analysis dataset, other | Analysis dataset not listed in other options. | ECRIN |
| IPD analysis dataset metadata definition | A dataset that is the metadata for an analysis dataset. | ECRIN |
| IPD CDMS format dataset | A dataset in the format used by the data collection system (CDMS). May be a database file. | ECRIN |
| IPD CDMS format dataset metadata definition | The metadata definition for the study's data collection system - equivalent to the functional specification of the system but as a dataset rather than a document. | ECRIN |
| IPD transport format dataset | A dataset in a format designed specifically for transfer between systems. | ECRIN |
| IPD transport format dataset metadata definition | Metadata describing the dataset when in a transport format. | ECRIN |
| Aggregated result dataset | Dataset with aggregated / summary results and statistics from the study. | ECRIN |
| Aggregated result dataset, efficacy measures | Dataset with aggregated results and statistics focusing on efficacy measures. | ECRIN |
| Aggregated result dataset, safety measures | Dataset with aggregated results and statistics focusing on safety measures. | ECRIN |
| Aggregated result dataset, other | Dataset with aggregated results and statistics, not listed elsewhere. | ECRIN |
| Aggregated result dataset metadata definition | Metadata definition of an aggregated result set. | ECRIN |

**Collection types**

The collection types defined so far are:

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Grouped journal articles | A collection of journal articles on the same topic or study (should also be recorded separately). | ECRIN |
| Grouped analysis datasets | A collection of analysis datasets on the same topic or study (should also be recorded separately). | ECRIN |
| Grouped aggregated result datasets | A collection of aggregated result datasets on the same topic or study (should also be recorded separately). | ECRIN |

**Software types**

Whilst the 5 software types currently defined are:

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Script(s) used in analysis | Statistical software scripts used in study analysis. | ECRIN |
| CDMS | Clinical Database Management System, or component within such a system, as used for clinical data management | ECRIN |
| Trial Management System | System used to support study administration. | ECRIN |
| eTMF | Electronic trial management system. | ECRIN |
| Data Extraction System | Used for extracting data from remote resources using XML or web scraping. | ECRIN |

**'Other' types**

14 records provide a single ‘other’ type, one for each object class. This is necessary because a type must belong to a specific class – in the ECRIN schema a generic ‘other’ is not possible.

If additional specific types are required for the text, dataset and software classes they should generally be created rather than subsumed under the ‘other’ type, but ‘other’ can be useful as a home for relatively rare types of object or those that are difficult to classify. There is also a single ‘unknown’ (type and class) entry, to avoid possible referential integrity issues within systems at entity creation.

## Object filter types

Object types are grouped under the 'filter types' listed below, to bring the list of options in the UI (e.g. when filtering by object type) down to a manageable number.

The 'T numbers' in the descriptions relate to the type ids (see object types table). Each type has a reference to the filter group to which it belongs (as well as to its overall object class).

|  |  |  |
| --- | --- | --- |
| **filter as** | **description** | **source** |
| Trial registry entry | T13 - Summary of the study and its aims, posted prospectively or retrospectively to a public registry. | ECRIN |
| Registry results summary | T28 - Summary of study results, as displayed in a trial registry. | ECRIN |
| Journal article | T12 - Articles in peer-reviewed publications that disseminate the results of original research and scholarship; T100 - Journal article abstract; T117 - Special journal issues; T135 - Working Paper / Pre-print; T152 - Grouped journal articles. | ECRIN |
| Study protocol | T11 - Protocol: Structured document describing the study, its rationale, methodology, outcome measures etc; T42 - Redacted protocol; T74 - Protocol with SAP; T75 - Protocol with informed consent forms; T76 - Protocol with SAP and ICFs. | ECRIN |
| Study overview | T38 - A brief overview of the study, may be an abridged protocol, or as used within the study web site or other study documents. | ECRIN |
| Patient consent / information forms | T18 - Informed consent forms, the form or forms given to participants to formally record their informed consent to study participation; T19 - Patient information sheets, the information provided to study participants, especially as part of the consenting process. | ECRIN |
| Data collection forms | T21 - Data collection forms, Copies, in electronic and / or paper form, of the case report forms (CRFs and / or eCRFs) used for collecting data. T30 - Annotated copies of CRFs / eCRFs; T40 - Standard instruments: standardised rating instruments, including questionnaires. | ECRIN |
| Manual of procedures | T35 - Manual of Operations, T36 - Manual of Procedures. Description of specific operations, workflow, procedures and techniques within the study. | ECRIN |
| Statistical analysis plan | T22 - Statistical analysis plan: the details of the proposed analysis for the study, listing the individual descriptive statistics and tests of inference, and their parameters; T29 - Analysis notes: a summary of the analysis carried out and / or any caveats to be borne in mind when interpreting results; T43 - a redacted SAP. | ECRIN |
| Clinical study report | T26 - Clinical study report - a full end of study report with detailed efficacy and safety results; T27 - a redacted clinical study report; T79 - a summary CSR; T85 - Unpublished Study Report: A report of a study, or part of a study, not formally published. May be an internal interim document within a long term study. | ECRIN |
| Data description | T20 - Database specification: Functional specification of the database including details of individual data items, types, ranges, logic checks, etc.; T31 - Data dictionary: A detailed, item by item, description of the data points in the dataset; T32 - Data Overview: A summary of the data indicating the nature of different tables, time points of data etc; T81 - Data collection schedule; T82 - Data coding manual. | ECRIN |
| Individual participant data | T80 - Individual participant data: A dataset simply called Individual Participant Data, or its equivalent, with no further qualification or description. Any of T51 - T68, being different specific types of IPD, e.g. relating to sub-populations or different time points, or metadata associated with IPD. Also T153 - Grouped analysis datasets. | ECRIN |
| Aggregated data | T69 - Aggregated result dataset: Dataset with aggregated / summary results and statistics from the study. Any of T70 - T73, being different specific types of aggregate data or metadata associated with it. Also T154 - Grouped aggregate datasets. | ECRIN |
| Other study resource | T14 - Ethics submission; T15 - Ethics approval notification; T16 - Regulatory authority submission; T17 - Regulatory authority approval notification; T23 - Data management plan; T24 - Trial master file contents list; T25 - Data monitoring committee report; T33 - Definitions of terms; T34 - Literature Review; T39 - Publication List; T77 - Investigational Product Information; T78 - General background to research topic; T83 - Bibliography; T84 - Introduction to document set; T86 - List of web links. | ECRIN |
| Conference material | T106 - Conference Abstract; T107 - Conference Paper; T108 - Conference Poster: T109 - Conference Program. | ECRIN |
| Other article | T119 - Magazine Article; T121 - Newsletter Article; T122 - Newspaper Article. | ECRIN |
| Book or chapter | T101 - Book; T102 - Book chapter: T103 - Book Prospectus; T104 - Book Review; T105 - Book Series; T113 - Edited Book. | ECRIN |
| Other information resource | T112 - Dissertation; T120 - Manual (for education / training purposes); T123 - Online Resource; T126 - Report; T127 - Research Tool; T128 - Supervised Student Publication. | ECRIN |
| Website | T134 - Stand-alone locations on the web where multiple types of information on a specific theme are available. May include interactive features for contributions from readers. | ECRIN |
| Software | T166 - Script(s) used in analysis; T167 - CDMS (Clinical Data Management System); T168 - Trial Management System; T169 - eTMF; T170 - Data Extraction System. | ECRIN |
| Other | All types in each object class labelled as "Other" (T37, T151, T155 - T165, T171). All remaining CASRAI defined text types: T110 - Dictionary Entry; T111 - Disclosure; T114 -Encyclopaedia Entry; T115 - Funding Submission; T118 - License; T124 - Patent; T125 - Registered Copyright; T129 - Tenure-Promotion; T130 - Test; T131 - Trademark; T132 - Translation; T133 - University Academic Unit. | ECRIN |

## Object access types

Data objects have an associated access type, one of the categories listed below. All of the options were developed by ECRIN, as the issue of access is rarely explicitly considered in other scientific domains (where open public access to data is assumed).

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Public on-screen access | Completely open access on the web but content not available in any other format. | ECRIN |
| Public on-screen access and download | Completely open access, material viewable and also directly available as a file download. | ECRIN |
| Public on-screen and API access | Completely open access, material viewable and also available through an API. | ECRIN |
| Public download (self-attestation required) | Public and downloadable |(only) once the user identifies and / or describes themselves. | ECRIN |
| Public on-screen access (self-attestation required) | Public and viewable (only) once the user identifies and / or describes themselves. | ECRIN |
| Restricted download | The user is an authenticated member of a defined group and can download the material (includes pay-walled journal articles). | ECRIN |
| Restricted on-screen access | The user is an authenticated member of a defined group but can only view the material. Analysis tools may be available. | ECRIN |
| Case by case download | Based on a review of an individual request, often also requiring supporting documentation. | ECRIN |
| Case by case on-screen access | Based on a review of an individual request, often also requiring supporting documentation. Analysis tools may be available. | ECRIN |
| Non public access - no details | The site asserts that the resource is not publicly available but provides no further details. | ECRIN |
| Other | None of the listed options. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Object instance types

In most cases the instance of a data object that is being referenced in the MDR system will be the entire object, as described by its various attributes. In some cases, however, the instance is not the full object but a description or summary of it.  
  
This can be the case when a public summary or abstract of a data object exists, but the full object is not publicly available. For example, an abstract of a journal article will normally be public, even though the article itself may be behind a pay wall. Rather than treat the abstract as a completely different data object (which would be one approach, but not very useful for most users) it is easier to categorise the abstract as a separate instance of ‘article abstract’ type, and describe it using the same attributes (title, creators, topics etc.) as the article itself.

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Full Resource | The entire resource as described (e.g. the file(s), document). | ECRIN |
| Summary description | A summary description of the resource, e.g. as provided by a data repository. | ECRIN |
| Article abstract | For journal articles, bibliographic details normally including an abstract, e.g. a PubMed entry. | ECRIN |
| Summary version | A summary of the resource itself (not a description), e,g, a protocol summary. | ECRIN |
| Redacted version | A redacted version of the resource, e.g. a redacted CSR. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Object relationship types

Relationships between data objects is a part of the DataCite scheme and is also included in the ECRIN schema. The types available are listed below. In almost all cases they come as corresponding pairs, allowing both ‘views’ or perspectives on a relationship between data objects to be stored.

| **name** | **description** | **source** |
| --- | --- | --- |
| Is cited by | Indicates that B includes A in a citation. | DataCite |
| Cites | Indicates that A includes B in a citation. | DataCite |
| Is supplement to | Indicates that A is a supplement to B. | DataCite |
| Is supplemented by | Indicates that B is a supplement to A. | DataCite |
| Is continued by | Indicates A is continued by the work B. | DataCite |
| Continues | Indicates A is a continuation of the work B. | DataCite |
| Is described by | Indicates A is described by B. | DataCite |
| Describes | Indicates A describes B. | DataCite |
| Has metadata | Indicates resource A has additional metadata B. | DataCite |
| Is metadata for | Indicates additional metadata A for a resource B. | DataCite |
| Has version | Indicates A has a version B. | DataCite |
| Is version of | Indicates A is a version of B. | DataCite |
| Is new version of | Indicates A is a new edition of B, where the new edition has been modified or updated. | DataCite |
| Is previous version of | Indicates A is a previous edition of B. | DataCite |
| Is part of | Indicates A is a portion of B; may be used for elements of a series. | DataCite |
| Has part | Indicates A includes the part B. | DataCite |
| Is referenced by | Indicates A is used as a source of information by B. | DataCite |
| References | Indicates B is used as a source of information for A. | DataCite |
| Is documented by | Indicates B is documentation about/ explaining A; e.g. points to software documentation. | DataCite |
| Documents | Indicates A is documentation about B; e.g. points to software documentation. | DataCite |
| Is compiled by | Indicates B is used to compile or create A. | DataCite |
| Compiles | Indicates B is the result of a compile or creation event using A. | DataCite |
| Is variant form of | Indicates A is a variant or different form of B. | DataCite |
| Is original form of | Indicates A is the original form of B. |  |
| Is identical to | Indicates that A is identical to B, for use when there is a need to register two separate instances of the same resource. | DataCite |
| Is reviewed by | Indicates that A is reviewed by B. | DataCite |
| Reviews | Indicates that A is a review of B. | DataCite |
| Is derived from | Indicates B is a source upon which A is based. | DataCite |
| Is source of | Indicates A is a source upon which B is based. | DataCite |
| Is required by | Indicates A is required by B (may be used to indicate software dependencies). | DataCite |
| Requires | Indicates A requires B (may be used to indicate software dependencies). | DataCite |
| Obsoletes | Indicates A replaces B. | DataCite |
| Is obsoleted by | Indicates A is replaced by B. | DataCite |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Contribution types

Contributions to the creation and management of data objects and studies can come from both individuals and organisations in a variety of ways, and this table lists the contribution types that can be referenced.

| **name** | **description** | **source** |
| --- | --- | --- |
| Creator | The main researchers involved in producing the data, or the authors of the publication, in priority order. To supply multiple creators, repeat this property. May be a corporate/institutional or personal name. | DataCite |
| Contact Person | Person with knowledge of how to access, troubleshoot, or otherwise field issues related to the resource. May also be “Point of Contact” in organisation that controls access to the resource. | DataCite |
| Data Collector | Person/institution responsible for finding, gathering/collecting data under the guidelines of the author(s) or Principal Investigator (PI). | DataCite |
| Data Curator | Person tasked with reviewing, enhancing, cleaning, or standardizing metadata and the associated data submitted for storage, use, and maintenance within a data centre or repository.  The Data Curator’s role encompasses quality assurance focused on content and metadata, e.g. checking whether the submitted dataset is complete, with all files and components as described by submitter, whether the metadata is standardized to appropriate systems and schema, whether specialized metadata is needed to add value and ensure access across disciplines, and determining how the metadata might map to search engines, database products, and automated feeds. | DataCite |
| Data Manager | Person (or organisation with a staff of data managers, such as a data centre) responsible for maintaining the finished resource. The work done by this person or organisation ensures that the resource is periodically “refreshed” in terms of software/hardware support, is kept available or is protected from unauthorized access, is stored in accordance with industry standards, and is handled in accordance with the records management requirements applicable to it. | DataCite |
| Distributor | Institution tasked with responsibility to generate/disseminate copies of the resource in either electronic or print form. Works stored in more than one archive/repository may credit each as a distributor. | DataCite |
| Editor | A person who oversees the details related to the publication format of the resource. If the Editor is to be credited in place of multiple creators, the Editor’s name may be supplied as Creator, with “(Ed.)” appended to the name. | DataCite |
| Hosting Institution | Typically, the organisation allowing the resource to be available on the internet through the provision of its hardware/software/operating support. May also be used for an organisation that stores the data offline. Often a data centre (if that data centre is not the “publisher” of the resource). | DataCite |
| Producer | Typically a person or organisation responsible for the artistry and form of a media product. In the data industry, this may be a company “producing” DVDs that package data for future dissemination by a distributor. | DataCite |
| Project Leader | Person officially designated as head of project team or subproject team instrumental in the work necessary to development of the resource. The Project Leader is not “removed” from the work that resulted in the resource; he or she remains intimately involved throughout the life of the particular project team. | DataCite |
| Project Manager | Person officially designated as manager of a project. Project may consist of one or many project teams and sub-teams. The manager of a project normally has more administrative responsibility than actual work involvement. | DataCite |
| Project Member | Person on the membership list of a designated project/project team. This vocabulary may or may not indicate the quality, quantity, or substance of the person’s involvement. | DataCite |
| Registration Agency | Institution/organisation officially appointed by a Registration Authority to handle specific tasks within a defined area of responsibility, e.g. DataCite is a Registration Agency for the International DOI Foundation (IDF). | DataCite |
| Registration Authority | A standards-setting body from which Registration Agencies obtain official recognition and guidance, e.g. the IDF serves as the Registration Authority for the International Standards Organisation (ISO) in the area/domain of Digital Object Identifiers. | DataCite |
| Related Person | A person without a specifically defined role in the development of the resource, but who is someone the author wishes to recognize. This person could be an author’s intellectual mentor, a person providing intellectual leadership in the discipline or subject domain, etc. | DataCite |
| Researcher | A person involved in analyzing data or the results of an experiment or formal study. May indicate an intern or assistant to one of the authors who helped with research but who was not so “key” as to be listed as an author. Should be a person, not an institution. | DataCite |
| Research Group | Typically refers to a group of individuals with a lab, department, or division; the group has a particular, defined focus of activity. May operate at a narrower level of scope; may or may not hold less administrative responsibility than a project team. | DataCite |
| Rights Holder | Person or institution owning or managing property rights, including intellectual property rights over the resource. | DataCite |
| Sponsor | Person or organisation that issued a contract or under the auspices of which a work has been written, printed, published, developed, etc. Includes organisations that provide in-kind support, through donation, provision of people or a facility or instrumentation necessary for the development of the resource, etc. | DataCite |
| Supervisor | Designated administrator over one or more groups/teams working to produce a resource or over one or more steps of a development process. | DataCite |
| Work Package Leader | A Work Package is a recognized data product, not all of which is included in publication. The package, instead, may include notes, discarded documents, etc. The Work Package Leader is responsible for ensuring the comprehensive contents, versioning, and availability of the Work Package during the development of the resource. | DataCite |
| Study Lead | The individual who, if not the sponsor themselves, leads and co-ordinates the scientific and clinical activity within a clinical study, including co-ordinating the work of principal investigators at clinical sites. May be known as the Co-ordinating Investigator, the Study Chair, Study Director or similar terms. | ECRIN |
| CT Site Principal investigator | The individual responsible for the safe conduct of a clinical trial at a particular clinical site. | ECRIN |
| Clinical Study Manager | An individual responsible for the operational management of a clinical study. Similar to a Project Manager but a study manager is heavily involved in the management of data and data collection. | ECRIN |
| Trial Sponsor | The organisation or individual that has the formal, legal role of a clinical trial sponsor, and is legally responsible for all aspects of a clinical trial. | ECRIN |
| Sponsor contact | An individual representing the sponsor and acting as an initial contact point. | ECRIN |
| Public contact | An individual designated as dealing with non-scientific queries from the public or press. | ECRIN |
| Recruitment contact | An individual designated as providing periodic updates on recruitment information or status, at all sites, usually for monitoring purposes. | ECRIN |
| Study Funder | An organisation providing some or all of the additional funds required for the study. | ECRIN |
| Funder contact | An individual representing the funder and acting as an initial contact point. | ECRIN |
| Independent monitoring committee member | A member of a safety monitoring committee for a clinical trial, independent of the researchers and research activity. | ECRIN |
| Medicinal product supplier | Organisation that provides one or more of the medicines investigated in a clinical study. | ECRIN |
| Medical device supplier | Organisation that provides one or more of the medical devices in a clinical study. | ECRIN |
| Logistics support organisation | Organisation that provides logistical input, e.g. provides a drug distribution service. | ECRIN |
| Scientific support organisation | Organisation that provides scientific support, e.g. a national or international research network. | ECRIN |
| Central Laboratory | Organisation that provides a central specialist laboratory testing facility. | ECRIN |
| Central Imaging facility | Organisation that provides a central specialist imaging or scanning facility. | ECRIN |
| Clinical organisation | Organisation, usually a primary or secondary health care organisation, that manages one or more of the sites where a clinical study takes place. | ECRIN |
| Clinical site | Organisation or location, usually in a primary or secondary health care organisation, that is one of the sites where a clinical study takes place. | ECRIN |
| Collaborating organisation | May be listed as a secondary sponsor, an organisation other than the lead sponsor involved in supporting a study. | ECRIN |
| Sponsor-investigator | An individual with the role of sponsor as well as being the co-ordinating investigator for the study. | ECRIN |
| Results contact | The individual, occasionally organisation, to be contacted for further information on the study results. | ECRIN |
| Research Group Member | From PubMed, an 'investigator' is an individual (e.g., collaborator or investigator) who is not an author of a paper but is listed as a member of a collective/corporate group that is an author of the paper. | PubMed |
| Other | Any person or institution making a significant contribution to the development and/or maintenance of the resource, but whose contribution does not “fit” other controlled vocabulary for contributor type. | DataCite |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Dataset consent types

The categories based on the Data Use Ontology (DUO) classification of biomedical consent types - see <https://github.com/EBISPOT/DUO>, which has been approved by the Global Alliance for Genomics and Health (GA4GH). The main categories are:

|  |  |  |  |
| --- | --- | --- | --- |
| **name** | **description** | **source** | |
| Not known | No clear information available about consent for secondary use, or if any exists. | ECRIN | |
| No explicit consent | No specific consent was given for the sharing of data or its re-use beyond the study in which it was originally collected. | ECRIN | |
| No restriction | No explicit restriction in the consent documents, though a broad consent to re-use is assumed, or this category is meaningless. | DUO | |
| General research use | Consent indicates that use is allowed for general research use for any research purpose. | DUO | |
| Health/medical/biomedical research | Consent indicates that use is allowed for health/medical/biomedical purposes; does not include the study of population origins or ancestry, or the development of methods / algorithms (e.g. for ML) | DUO | |
| Disease-specific research | Consent indicates that use, for health / medical / biomedical research is allowed provided it is related to a specified disease (area). The disease (area) must be named or coded in the associated comments field. | | DUO |

## Dataset de-identification levels

These categories indicates the level of de-identification that has been applied. The possible values are:

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Not known | No clear information available about the de-identification, if any, applied to the data. | ECRIN |
| No de-identification | Confirmed that no de-identification measures have been applied to the data set. | ECRIN |
| De-identification applied | Some de-identification measures have been applied. Details should be described in comments and / or indicated in the linked boolean fields, or in separate documents. | ECRIN |
| De-identification applied, primary outcomes re-assessed | Some de-identification measures have been applied and are described. In addition the data has been re-analysed against the primary outcomes and the results described. | ECRIN |

## Dataset record key types

These categories indicate the type of record keys within the dataset, in particular whether the dataset is anonymised, pseudonymised, or neither. Note that the categorisation applied in any particular case (if one is supplied) is based upon the description provided by the data controller / manager. The classification is therefore based on the data controller's understanding of the relevant terms.

No attempt is made to apply a standardised criteria, because the meaning of the words used ('pseudonymised'. 'anonymised', etc.) may vary between different legal jurisdictions, over time, and in different usage contexts. The categorisation should therefore be read as only a **very approximate** guide to any legal requirements associated with the data.

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Not known | No clear information available about the record keys in use. | ECRIN |
| Anonymised | Data controller or manager describes dataset as ‘anonymised’, in their interpretation of the term. | ECRIN |
| Pseudonymised | Data controller or manager describes dataset as ‘pseudonymised’, in their interpretation of the term. | ECRIN |
| Identifiable | Data controller or manager describes dataset as ‘identifiable’, in their interpretation of the term. | ECRIN |

## Date types

| **name** | **description** | **source** |
| --- | --- | --- |
| Accepted | The date that the publisher accepted the resource into their system. To indicate the start of an embargo period, use Submitted or Accepted, as appropriate. | DataCite |
| Available | The date the resource is made publicly available. May be a range. To indicate the end of an embargo period, use Available. | DataCite |
| Controlled access in force | The date the resource is made available under controlled access of some form. May be a range. To indicate the end of such a period, use Available. | ECRIN |
| Copyrighted | The specific, documented date at which the resource receives a copyrighted status, if applicable. | DataCite |
| Collected | The date or date range in which the resource content was collected. To indicate precise or particular timeframes in which research was conducted. | DataCite |
| Created | The date the resource itself was put together; this could be a date range or a single date for a final component, e.g., the finalised file with all of the data. Recommended for discovery. | DataCite |
| Issued | The date that the resource is published or distributed e.g. to a data centre. | DataCite |
| Submitted | The date the creator submits the resource to the publisher. This could be different from Accepted if the publisher then applies a selection process. Recommended for discovery. | DataCite |
| Updated | The date of the last update to the resource, when the resource is being added to. May be a range. | DataCite |
| Valid | The date or date range during which the dataset or resource is accurate. | DataCite |
| PubMed citation created | The date in the Pubmed XML, in the element 'DateCreated'. | PubMed |
| PubMed citation revised | The date in the Pubmed XML, in the element 'DateRevised'. | PubMed |
| PubMed citation completed | The date in the Pubmed XML, in the element 'DateCompleted'. | PubMed |
| Epublish | Date of electronic publication. | PubMed |
| Ppublish | Date of print publication. | PubMed |
| Revised | Date an article was revised in publication by the authors. | PubMed |
| Ahead of print publication | Date an article was published electronically, ahead of print publication (PubMed date category). | PubMed |
| Retracted | Date the publisher retracted an article or a resource. | PubMed |
| Added to eCollection | Date an article was included in an electronic collection (similar to an issue). | PubMed |
| Added to PMC | Date article was added to PMC. | PubMed |
| Added to Pubmed | Date the citation was added to PubMed, unless the citation is added to PubMed more than twelve months since the date of publication. In that case, the PubMed date is set to the date of publication. | PubMed |
| Added to Medline | Date the citation completed Medline processing. Up until the citation has been indexed for Medline, the Medline date is the same as the Entrez date. | PubMed |
| PMC embargo release | Date a full-text article was released from embargo in PubMed Central (PMC). | PubMed |
| Added to Entrez | Date when PubMed entry entered into the e-utils Entrez system. | PubMed |
| Other | Date type not defined elsewhere. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Description types

| **name** | **description** | **source** |
| --- | --- | --- |
| Abstract | A complete abstract or summary from a document. | DataCite |
| Abstract Section | A labelled section in a structured abstract. | PubMed |
| External Abstract | Abstract not written by the authors (may be a translation). | PubMed |
| Methods | A description of how the data or other object type was constructed. | DataCite |
| Series Information | Information about a repeating series, such as volume, issue, number. | DataCite |
| Table of Contents | A listing of the Table of Contents. | DataCite |
| Technical Info | Detailed information that may be associated with design, implementation, operation, use, and/or maintenance of a process or system. Includes population information for a clinical research dataset. | DataCite |
| Journal Source String | The bibliographic reference to a journal article – the string that normally follows authors and title. | PubMed |
| Data availability description | A description of the location of a Dryad dataset. | PubMed |
| Other | A description not falling into any of the other categories. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Identifier types

The identifiers fall into three groups: those that are applicable to studies, those that are applicable to data objects, and those that are applicable to both, and these groups are reflected in the three different lists.

**Study identifiers**

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| Trial Registry ID | Org id should identify the registry. | ECRIN |
| Regulatory Body ID | Number generated by a national authority when a study applies for authorisation / ethical approval, e.g. IRAS number in the UK, ANSM number in France. | ECRIN |
| Ethics Review ID | Org id should identify the review body. | ECRIN |
| Funder's ID | May be a grant id, org id should identify the funder. | ECRIN |
| Sponsor's ID | Org id should identify the sponsor. | ECRIN |
| NIH CTRP ID | ID generated by the US NCI for supported cancer trials (in the Clinical Trial Reporting Program). | NIH |
| DAIDS ID | ID generated for the protocol by the US NIH DAIDs for supported trials (allergy and infectious diseases). | NIH |
| NHLBI ID | Accession number, from NHLBI study database. | NIH |

**Data Object identifiers**

| **name** | **description** | **source** |
| --- | --- | --- |
| Repository accession number | Org id should identify the repository manager. | ECRIN |
| PMID | ID assigned by PubMed. | PubMed |
| PMCID | Pub med central ID, of the manuscript itself. | PubMed |
| NIH Manuscript ID | A Manuscript ID, an identifier assigned to an author manuscript submitted to the NIH Manuscript Submission System. | PubMed |
| Medline UID | Medline identifier. | PubMed |
| bioRxiv ID | Pre-prints, ID assigned by Cold Spring Harbour Laboratory. | ECRIN |
| arXiv ID | Pre-prints, ID assigned by Cornell university. | DataCite |
| psyXiv ID | Pre-prints, ID assigned by the Centre for Open Science. | ECRIN |
| socXiv ID | Pre-prints, ID assigned by the Centre for Open Science. | ECRIN |
| Handle ID | ID assigned by a naming authority (handle system is a superset of DOIs). | DataCite |
| ISBN | International Standard Book Number, assigned by the publisher. | DataCite |
| ISTC | International Standard Text code ID, assigned by the international ISTC agency. | DataCite |
| ISAN | International Standard Audiovisual Number, assigned by the ISAN international agency. | ECRIN |
| LSID | Life Science Identifier, a URN like specification with various issuing authorities. | DataCite |
| Other Bibliographic ID | Org id should identify the system manager. | ECRIN |
| NRCBL | KIE Reference Library (bioethics library) shelving location. | PubMed |
| Publisher article ID | Internal reference of article publisher for journal article. | PubMed |
| PMC Publisher ID | Publisher Id supplied to PubMed Central. | PubMed |
| PM Publisher ID | Publisher Id supplied to PubMed. | PubMed |
| Serial Item and Contribution Identifier | The Serial Item and Contribution Identifier (SICI), a code used to uniquely identify specific volumes, articles or other identifiable parts of a serial. | PubMed |

Note that some of these identifiers were added to accommodate the ids referenced in PubMed data. The exact relationship between the three ‘publisher IDs’ needs to be clarified.

**Study or Data Object identifiers**

| **name** | **description** | **source** |
| --- | --- | --- |
| URL | Resource locator for a web based resource. | DataCite |
| PURL | Persistent URL that redirects if necessary. | DataCite |
| URN | Uniform Resource Name (a URI using the URN schema) that is location independent. | DataCite |
| ARK | Archival Resource Key, a URL that provides additional metadata. | DataCite |
| Other | None of the listed identifier types. | ECRIN |
| Type not provided | Missing type data in data source. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Language codes

The table below lists the languages, their codes and their source that are currently in the system. (*\* = an official ISO 369-1 code, as an ISO two letter code does not exist for tis language)*

| **369-1** | **369-2** | **Language** | **source** |
| --- | --- | --- | --- |
| af | afr | Afrikaans | web |
| am | amh | Amharic | PubMed |
| ar | ara | Arabic | web |
| az | aze | Azerbaijani | web |
| be | bel | Belarusian | web |
| bg | bul | Bulgarian | web |
| bn | ben | Bengali | web |
| bo | tib | Tibetan | web |
| br | bre | Breton | web |
| bs | bos | Bosnian | web |
| ca | cat | Catalan | web |
| ce | che | Chechen | web |
| co | cos | Corsican | web |
| cs | cze | Czech | web |
| cy | wel | Welsh | web |
| da | dan | Danish | web |
| de | ger | German | web |
| el | gre | Greek | web |
| en | eng | English | web |
| eo | epo | Esperanto | PubMed |
| es | spa | Spanish | web |
| et | est | Estonian | web |
| eu | baq | Basque | web |
| fa | per | Persian | web |
| fi | fin | Finnish | web |
| fr | fre | French | web |
| ga | gle | Irish Gaelic | web |
| gd | gla | Scottish Gaelic | web |
| gl | glg | Galician | web |
| gu | guj | Gujarati | web |
| ha | hau | Hausa | web |
| he | heb | Hebrew | web |
| hi | hin | Hindi | web |
| hr | hrv | Croatian | web |
| hu | hun | Hungarian | web |
| hy | arm | Armenian | web |
| id | ind | Indonesian | web |
| is | ice | Icelandic | web |
| it | ita | Italian | web |
| iu | iku | Inuktitut | web |
| ja | jpn | Japanese | web |
| jv | jav | Javanese | web |
| ka | geo | Georgian | web |
| kk | kaz | Kazakh | web |
| kl | kal | Greenlandic, Kalaallisut | web |
| km | khm | Central Khmer | web |
| ko | kor | Korean | web |
| ks | kas | Kashmiri | web |
| ku | kur | Kurdish | web |
| la | lat | Latin | ECRIN |
| lb | ltz | Luxembourgish | web |
| lo | lao | Lao | web |
| lt | lit | Lithuanian | web |
| lv | lav | Latvian | web |
| mi | mao | Maori | web |
| mk | mac | Macedonian | web |
| ml | mal | Malayalam | PubMed |
| mn | mon | Mongolian | web |
| ms | may | Malay | web |
| mt | mlt | Maltese | web |
| mu\* | mul | Multiple languages | PubMed |
| my | bur | Burmese | web |
| ne | nep | Nepali | web |
| nl | dut | Dutch | web |
| no | nor | Norwegian | web |
| os | oss | Ossetian | web |
| pa | pan | Punjabi | web |
| pl | pol | Polish | web |
| ps | pus | Pashto | PubMed |
| pt | por | Portuguese | web |
| qu | que | Quechua | web |
| rm | roh | Romansh | web |
| ro | rum | Romanian, Moldavian | web |
| ru | rus | Russian | web |
| rw | kin | Kinyarwanda | PubMed |
| se | sme | Northern Sami | web |
| si | sin | Sinhalese | web |
| sk | slo | Slovak | web |
| sl | slv | Slovenian | web |
| sm | smo | Samoan | web |
| sn | sna | Shona | web |
| so | som | Somali | web |
| sq | alb | Albanian | web |
| sr | srp | Serbian | web |
| sv | swe | Swedish | web |
| sw | swa | Swahili | web |
| ta | tam | Tamil | web |
| te | tel | Telugu | web |
| tg | tgk | Tajik | web |
| th | tha | Thai | web |
| tk | tuk | Turkmen | web |
| to | ton | Tongan | web |
| tr | tur | Turkish | web |
| tt | tat | Tatar | web |
| ty | tah | Tahitian | web |
| uk | ukr | Ukrainian | web |
| un\* | und | Undetermined | PubMed |
| ur | urd | Urdu | web |
| uz | uzb | Uzbek | web |
| vi | vie | Vietnamese | web |
| xh | xho | Xhosa | web |
| yo | yor | Yoruba | web |
| zh | chi | Chinese | web |
| zu | zul | Zulu | web |

## Resource types

| **name** | **description** | **source** |
| --- | --- | --- |
| PDF |  | ECRIN |
| Web text | Web page or text, no download or API options, | ECRIN (PubMed) |
| Web text with download | Web text, with download option (e.g. PubMed Central entry). | ECRIN (PubMed) |
| Web text with XML via API | Web page or text, content as XML via API. | ECRIN (ClinicalTrials.gov) |
| Web text with JSON via API | Web page or text, content as JSON via API. | ECRIN (ClinicalTrials.gov) |
| Web text with XML or JSON via API | Web page or text, content as XML or JSON via API (e.g. from Clinical trials.gov, PubMed). | ECRIN (ClinicalTrials.gov) |
| Comma separated values | Usually .csv files. | ECRIN |
| Tab separated values | May be .tsv or .txt files. | ECRIN |
| Plain text file | e.g. from text editor, or data with delimiter unknown. | ECRIN |
| Rich text file | .rtf files. | ECRIN |
| Word doc | .doc or .docx files. | ECRIN |
| Other WP document | e.g. .odt (libra office), .gdoc (google docs) files. | ECRIN |
| Excel Spreadsheet(s) | .xsl or .xslx. files. | ECRIN |
| Other spreadsheet(s) | e.g. .ods (libra office), .gsheet (google docs) files. | ECRIN |
| PowerPoint | .ppt or .pptx files. | ECRIN |
| Other presentation | e.g. .odp (libra office), .gslides (google docs) files). | ECRIN |
| ODM XML document | Data or metadata following CDISC ODM schema. | ECRIN |
| SDTM XML document | Data using the CDISC SDTM schema. | ECRIN |
| Define.XML document | Metadata following the CDISC Define.xml schema. | ECRIN |
| ADaM XML document | Analysis data following thbe CDISC ADaM schema. | ECRIN |
| XML document | XML schema not listed elsewhere. | ECRIN |
| SAS Transport file | .xpt files. | ECRIN |
| R workspace file | .rdata or .rda files. | ECRIN |
| SSPS data file | .sav file. | ECRIN |
| Stata data file | .dta file. | ECRIN |
| Statistical program data file | Stats data file in format not listed elsewhere. | ECRIN |
| Database file | A complete DB file that could be remounted directly into a DB system. | ECRIN |
| Graphic image | e.g. .png, .jpeg, .svg files. | ECRIN |
| Media file | e.g. mp3, .mp4, .mpeg files. | ECRIN |
| Other | File format not listed elsewhere. | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |

## Time Units

|  |  |  |
| --- | --- | --- |
| **name** | **description** | **source** |
| seconds |  | ECRIN |
| minutes |  | ECRIN |
| hours |  | ECRIN |
| days |  | ECRIN |
| weeks |  | ECRIN |
| months |  | ECRIN |
| years |  | ECRIN |
| not provided |  | ECRIN |

## Title types

| **name** | **applies\_to** | **description** | **source** |
| --- | --- | --- | --- |
| Trial registry title | Study | Full scientific title, as quoted in a trial registry. | ECRIN |
| Protocol title | Study | Full scientific title, from a punlished protocol. | ECRIN |
| Other scientific title | Study | Full scientific title, not from the protocol or registry entry. | ECRIN |
| Acronym or Abbreviation | Study | As provided by study sponsors. | ECRIN |
| Public Title | Study | In contrast to the full scientific title, usually from a trial registry. | ECRIN |
| Subtitle | All | A subtitle provided by object creators or study sponsors. | DataCite |
| Translated Title | All | Used in conjunction with language code to indicate language translated into. | DataCite |
| Journal article title | Data Object | Full journal title, as listed in citation. | ECRIN |
| Unique data object title | Data Object | Use if the given name for the data object can be guaranteed to be unique. | ECRIN |
| Study short name :: object name | Data Object | Constructed using study short name to prefix a non unique name. If obvious from context study name can be omitted. | ECRIN |
| Study short name :: object type | Data Object | Constructed using study short name to prefix object's type. If obvious from context study name can be omitted. | ECRIN |
| Study scientific name :: object name | Data Object | Constructed using study full name to prefix a non unique name. If obvious from context study name can be omitted. | ECRIN |
| Study scientific name :: object type | Data Object | Constructed using study full name to prefix object's type. If obvious from context study name can be omitted. | ECRIN |
| Study registry ID :: object name | Data Object | Constructed using registry id to prefix a non unique name. If obvious from context study ID can be omitted. | ECRIN |
| Study registry ID :: object type | Data Object | Constructed using registry id to prefix object's type. If obvious from context study ID can be omitted. | ECRIN |
| Other Alternative Title | All | Any alternative title not described elsewhere. | DataCite |
| Not yet known | All | Dummy value supplied by default on entity creation. | ECRIN |

The context will often make it unnecessary to include the prefix. For example the display of linked data objects under a study heading would allow the study name prefix to be dropped.

## Topic types

In many of the source systems some of the topics / keywords are categorised according to the type of data that they represent. This categorisation can be useful in searching and filtering scenarios and so is included in the ECRIN schema.

| **name** | **description** | **type** | **source** |
| --- | --- | --- | --- |
| Keywords | Topic that was not categorised or does not fit into one of the categories listed. Often written by study or object creators. | Free-text or controlled vocabulary | ClinicalTrials.gov |
| chemical / agent | One or more chemicals or biological agents, relevant to the study, including as interventions under test. | Free-text or controlled vocabulary | ClinicalTrials.gov |
| Condition | Illness or condition that is being targeted within study. | Free-text or controlled vocabulary | ClinicalTrials.gov |
| Design | Aspect of study design methodology. | Free-text or controlled vocabulary | PubMed |
| Outcome | Outcome measure or outcome produced within the study. | Free-text or controlled vocabulary | PubMed |
| Geographic | A geographical entity that was the particular focus or limit of the study. | Free-text or controlled vocabulary | PubMed |
| Organism | Organism, e.g. particular bacterium, that was targeted during the study. | Free-text or controlled vocabulary | PubMed |
| treatment protocol | The name of a particular treatment regime / protocol, e.g. a chemotherapy regime. | Free-text | PubMed |
| subject characteristics | Descriptive term pertaining to the subject group of the study. | Free-text | PubMed |
| Not yet known | Dummy value supplied by default on entity creation | not applicable | ECRIN |

## Topic vocabularies

| **name** | **description and / or url** | **source** |
| --- | --- | --- |
| Generated by authors | Free text, not intentionally part of any controlled vocabulary | ECRIN |
| ICD 10 | <https://icd.who.int/browse10/2016/en> | ECRIN |
| ICD 11 | <https://icd.who.int/browse11/l-m/en> | ECRIN |
| MESH | <https://www.nlm.nih.gov/mesh/meshhome.html> | ECRIN |
| SnoMed CT | <http://www.snomed.org/> | ECRIN |
| MedDRA | <https://www.meddra.org/> | ECRIN |
| NCI thesaurus | <https://ncit.nci.nih.gov/ncitbrowser/> | ECRIN |
| Cochrane PICO terminology | <https://linkeddata.cochrane.org/pico-ontology> | ECRIN |
| CDISC controlled terminology | <https://www.cdisc.org/standards/terminology> | ECRIN |
| LOINC | <https://loinc.org/> | ECRIN |
| ATC drug classification | <https://www.whocc.no/atc/structure_and_principles/> | ECRIN |
| WHO Drug | <https://www.who-umc.org/whodrug/whodrug-portfolio/> | ECRIN |
| IUPAC chemical names | and biochemical names. <https://iupac.org/what-we-do/nomenclature/> | ECRIN |
| InChI chemical identifier | <https://iupac.org/who-we-are/divisions/division-details/inchi/> | ECRIN |
| Enzyme Commission numbers / names | <https://www.qmul.ac.uk/sbcs/iubmb/> | ECRIN |
| HGNC human genome codes / names | and related proteins. <https://www.genenames.org/> | ECRIN |
| Taxonomic names | e.g. Linnaean binominals. <http://www.iczn.org/iczn/index.jsp> | ECRIN |
| Not yet known | Dummy value supplied by default on entity creation. | ECRIN |