RDCA-DAP Glossary of Terms

Administrator

A person who is responsible for the upkeep, configuration, and reliable operation of a data collection (registry or clinical trial database). There are varying levels of administrator roles: system administrator, data engineer, data curator. [1]

Anonymized Data

Previously identifiable data (indirectly or individually identifiable) that have been de-identified and for which a code or other link no longer exists. An investigator has NO means for linking anonymized data back to a specific subject. [2]

See also: "de-identified data"

Assent

A process used when patients are below the age of consent for the patient to actively show willingness to participate in the research and understanding about the research to the degree they are capable. [3]

Biomarker

A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. [4]

Clinical outcome assessment

A measure that describes or reflects how a patient feels, functions, or survives. [5]

Clinical Trial Simulation Tool

A computer program, based on mathematical models of disease progression built from existing data, that allows users to test different trial designs in silico to determine the more efficient trial design for a proposed trial.

Data Contributor

A data contributor (also known as data custodian) willing and able to share data with RDCA-DAP. The contributor retains ownership of the data and tells RDCA-DAP how the data may be shared or used in the RDCA-DAP platform.

Data contribution agreement

A legal document signed by a data custodian (also known as Data Contributor) for a specific dataset and by C-Path that defines how the data will be used within RDCA-DAP. The contributor states that the data was collected and shared ethically and C-Path agrees to keep the data secure and share the data only as agreed within the document.

Data Curation

The organization and integration of data collected from various sources. It may involve annotation, publication or presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation. [6]

Data Custodian

The person or entity that has collected data in a registry, study, clinic or other process and is legally able to share data with RDCA-DAP. The data custodian is responsible for ethical collection and sharing of data, using appropriate consent documents and ethics approvals for the study. (Also known as Data Contributor)

Data Engineer

A person who sets up and maintains the data infrastructures that support information systems and applications. Data engineers are responsible for building and maintaining pipelines that feed data to data scientists. [7]

Data governance

The process of creating and maintaining mechanisms for responsibly acquiring, storing, safeguarding, and using data in a way that demonstrates good stewardship. [8]

Data Integration

Combining data from different sources and providing users with a unified view of them. [9]

Data Lake

A system or repository of data stored in its natural/raw format, usually object blobs or files that can include structured data from relational databases, semi-structured data (CSV, logs, XML, JSON), unstructured data and binary data. [10]

Data Silo

A data store or repository that is isolated from other data sources due to lack of access or shared standards, metadata, and formats.

Data standard

The rules by which data are described and recorded. [11]

Common Data Model

Common Data Models are used to integrate data that come from multiple different sources in a standardized format using a commonly defined structure and relationships between the data. An example of a common data model is the Study Data Tabulation Model. [12]

Data Use Committee

RDCA-DAP has established a data use committee that reviews research applications from users who wish to access and use data from the platform. This committee consists of representatives from NORD, C-Path, the rare disease community and academia. The committee will review all ethical research requests that can be completed by the proposed user with available data.
Aka: data standards and monitoring board, data access committee

Database

A structured set of data held in a computer or cloud environment, especially one that is accessible in various ways. Database structures can be as simple as a spreadsheet or as complex as a complex relational or graph model. [1]

Datamart

Subset of data extracted from all the data within RDCA-DAP to be used for a specific analysis.

De-identified data

Also known as: anonymized data, pseudonomyzed data: A record in which identifying information is removed so that the data cannot be traced back to an individual.

Under the HIPPA Privacy Rule, data are de-identified if either:
1. an experienced expert determines that the risk that certain information could be used to identify an individual is "very small" and documents and justifies the determination, or

2. the data do not include any of the 18 identifiers (of the individual or his/her relatives, household members, or employers) which could be used alone or in combination with other information to identify the subject. Note that even if these identifiers are

removed, the Privacy Rule states that information will be considered identifiable if the covered entity knows that the identity of the person may still be determined. [13]

Under GDPR all direct and indirect identifiers must be removed from the data. [14]

See also: "anonymized data", pseudonomyzed data"

Federated data

A virtual database or data system that aggregates data that are stored in multiple physical locations by providing a shared data model and access method.

IRB

Institutional Review Board (IRB)/Independent Ethics Committee (IEC) – An independent body constituted of medical, scientific, and nonscientific members whose responsibility it is to ensure the protection of the rights, safety, and well-being of human subjects involved in a trial or other study by, among other things, reviewing, approving, and providing continuing review of protocols and amendments, and of the methods and material to be used for obtaining and documenting informed consent of the trial participant. [15]

Informed consent

A process by which a participant or legal guardian voluntarily confirms his or her willingness to participate in a particular trial, after having been informed of all aspects of the trial that are relevant to the participant's decision to take part in the clinical trial. Informed consent is usually documented by means of a written, signed, and dated informed consent form, which has been approved by an IRB/IEC. [15]

Individually identifiable data

Any information that includes personal identifiers (18 HIPAA Identifiers or any subset of health information that identifies the individual or can reasonably be used to identify the individual). [16]

Indirectly identifiable

Indirect identification means you cannot identify an individual through the information you are processing alone, but you may be able to by using other information you hold or information you can reasonably access from another source. A third party using your data and combining it with information they can reasonably access to identify an individual is another form of indirect identification. To determine what data may be considered identifiable, please see de-identified. [17]

Medical Product Development Tool

Methods, materials, or measurements used to assess the effectiveness, safety, or performance of a medical product. In a regulatory context, examples of MPDTs are clinical outcome assessments, assessments of biomarkers, and non-clinical assessment methods or models. [18]

Metadata

Data that provides information about other data. Includes descriptive metadata, structural metadata, and administrative metadata.

Natural History Study

A study that collects information about the natural history of a disease (I.e. disease course) in the absence of an intervention, from the disease's onset until either its resolution or the individual's death. [19]

Ontology

A representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many, or all.
More simply, an ontology is a way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject. [20]

Study Participant/ Subject

Human subject means a living individual about whom an investigator (whether professional or student) conducting research:

i.   Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or

ii.   Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens. [21]

Patient-reported outcome instrument

Any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. [21]

P.I. - Principal Investigator

The person who is responsible for the scientific and technical direction of the entire clinical study or other data collection. [22]

Pseudonymized data

Previously identifiable data (indirectly or individually identifiable) that have been de-identified and for which a code or other link still exists but is kept separately from the data. An investigator can only link pseudonymized anonymized data back to a specific subject by going back to the original source of data. [23]

Query

A request for information from a database. Queries can be conducted in the database by selecting parameters from a pre-determined menu and specifying certain fields and values that define that query to produce tailored results. [1]

Real-World Data

Real-world *data* are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources (for example: electronic health records, claims and billing activities, product and disease registries, patient-generated data including in home-use settings, data gathered from other sources that can inform on health status, such as mobile devices). [24]

Registry

A registry is simply a database that collects and stores specified types of information that are usually related in some way. In the context of the therapy development pathway, a registry usually collects information about patients who have a specific disease or condition and may be referred to as a patient registry. However, other registries may seek participants who are healthy and are interested in volunteering for phase 1 clinical trials. Registries may contain information that is reported by patients, by clinicians or researchers, or a combination. The goals of registries vary as does the information being collected. [25]

Registry Platform

An existing IT platform designed to host and run a registry in a consistent way across multiple disease areas or types.

Reporter

In research studies designed to collect patient experience data, the reporter is the individual, group of individuals, or entity providing patient experience data. Reporters may be patients, parents, sexual/romantic partners, caregivers, physicians, or other healthcare professionals. Selection of an appropriate reporter in a given research study will depend on the definition of the target patient population of interest. If a patient in the target population can be reasonably expected to reliably self-report, then one would expect the patient herself/himself to be the reporter in that research study. [26]

Respondent

The individual who is entering data into a registry system. For patient-reported registries this may be the individual themselves or a legally authorized representative. For clinical registries this may be a doctor, other health professional or a member of the clinical staff. (Also known as Reporter)

Sponsor

The organization or individual that sponsors or funds a clinical trial or study including physicians, foundations, medical institutions, voluntary groups, and pharmaceutical companies, as well as Federal agencies such as NIH, FDA, the Department of Defense, and the Department of Veterans Affairs. [27]

Data user

A person (from academia, industry, patient group or other researcher) who wishes to access data within RDCA-DAP to answer specific research questions related to rare diseases. The data user must request access to the data of interest using a standardized research request, be approved for access and sign terms and conditions for use of the data.

Data Use Agreement

A legal document signed by the data user prior to gaining access to patient-level data in RDCA-DAP. Explaining the conditions requiring ethical use of data, protection of data, and acknowledgement of the source of the data etc.