PCORI Funding Announcement: Improving Infrastructure for Conducting Patient-Centered Outcomes Research

National Patient-Centered Clinical Research Network (PCORnet) Glossary of Terms

**Common Data Model (CDM)** – A common data model (CDM) standardizes the definition, content, and format of data across networks to enable a single standardized view that can be used for querying. For example, the common data model might specify that SEX is the data element name that refers to self-reported or clinician observed patient sex and the allowable values are “M,” “F,” “A,” and “U,” and that all missing values be coded as “U.”

**Common Data Element (CDE)** – Common data elements (CDEs) are standardized data elements that are shared across networks. The data elements in the PCORnet CDM are CDEs. For information about other CDEs, see the [NIH CDE Resource Portal](https://cdmr.gov/cde)

**Computable Phenotype** – A computable phenotype is an electronic algorithm used to identify certain conditions or characteristics. Within PCORnet, we will use computable phenotypes to query the PCORnet CDM to identify cohorts of individuals.

**Data Domains** – In healthcare research, the term “data domain” is used to refer to different categories of healthcare data in a data model (some examples are procedures, laboratory test results, demographics, and vital signs).

**Data Element** – A data element is the information that describes a piece of data (i.e., the metadata). Attributes of a data element may include the name, definition, provenance, and value set. For example, in PCORnet CDM v1.0, the data element for “discharge disposition” has the following attributes:

- Name = Discharge Disposition
- Definition: Vital status at discharge
- Provenance: Not specified in the CDM because it can vary (could be electronic health records, billing records, or healthcare claims)
- Value set = A (Discharged alive); E (Expired); NI (No information); UN (Unknown); OT (Other)

**DataMart** – Within PCORnet, a DataMart refers to a specific data resource that can be uniquely defined and queried using the PCORnet DRN Query Tool. Networks will create their DataMart(s) through an (Extract, Transform, and Load) ETL of source data. PCORnet DataMarts are DataMarts that adhere to the PCORnet CDM.
**DataMart Client** – The DataMart Client is the software application that supports the PCORnet Distributed Research Network (DRN) Query Tool system. The DataMart Client is used by the analysts and other authorized users who administer the DataMarts. See [PopMedNet DataMart Client](#) for more information.

**Data Harmonization/Standardization** – These terms are sometimes used interchangeably to refer to the process of transforming heterogeneous data sources into a common structure and format, such as a common data model.

**Data Provenance** – Data provenance describes a datum’s origins and how it has been derived, manipulated, combined, transformed, moved, and updated over time. For example, the provenance of a patient problem list might trace it from the point of entry into an electronic health record system through its inclusion in a clinical data warehouse and into a PCORnet DataMart. The [PCORnet Extract, Transform, and Load (ETL) Annotated Data Dictionary](#) is the mechanism for PCORnet networks to communicate data provenance related to the PCORnet Common Data Model. This information will be used to help the PCORnet Coordinating Center better understand the transformation process, appropriate uses of the PCORnet data, and the comparability of data sources.

**Data Transformation** – Data transformation is the process of converting a set of data values from the data format of a source data system into another data format, such as a CDM. For example, a simple transformation of data values for the variable SEX might be to transform source values of male to M, female to F, unknown to U, and ambiguous to A. Similarly, a simple transformation of data names might be to transform source data names of “BIRTH_DT,” “DOB,” and “B_DATE” to “BIRTH_DATE.”

**Data Warehouse/Clinical Data Warehouse /Enterprise Data Warehouse** – A data warehouse typically houses standardized, structured, integrated healthcare data that are extracted from various operational systems or data sources. Data warehouses may include electronic health record systems, billing systems, claims and authorization systems (for insurers and payers), clinical registries, clinical trial databases, and patient health surveys. A data warehouse is structured to facilitate use (i.e., to make the data more useable than source data). For example, hospital bills are often adjusted before being finalized. The hospital’s billing system would contain each of those bills, but the data warehouse would include only the final bill.

**Database** – A database is an application that stores electronic data and allows retrieval of that data. The most common type of database is a relational database. The same structure is repeated for every row (record) in a table, and the tables are related to one another in a defined manner. For example, a simple relational database might have a demographics and phone number table. The demographics table has five fields (patient ID, first name, last name, date of birth, and sex) and two records (John Doe and Sally Smith). The phone number table has four fields and three records (two for John Doe and one for Sally Smith). The tables can be joined together (connected) by assigning an ID.

**Distributed Research Network (DRN)** – A distributed research network (DRN), otherwise known as a distributed health data network, is an approach to multi-site research that allows secure analysis of separate data resources held by partners behind their firewalls. In a DRN there is no central data
Each data resource is held locally and consists of data collected, captured, or otherwise obtained by the local health system.

**Electronic Health Record (EHR)** – An electronic health record (EHR) is a repository of electronic information about an individual’s health status and health care. EHRs contain much of the same information that is found in a patient’s (paper) medical chart, but because the records are digitized, the data can be viewed, and providers (e.g., primary care physicians and specialists) can capture far more extensive information. EHRs may contain administrative and billing data, patient demographics, progress notes, vital signs, medical histories, diagnoses, medications, immunization records, allergies, radiology images, laboratory and other test results, and much more.

**ETL (Extract, Transform, and Load)** – ETL is a process in which programmers extract data from one or more data sources, transform the data to fit certain requirements or specifications, and then load the data into a desired location. In the context of PCORnet, programmers at the various clinical data research networks (CDRNs) and patient-powered research networks (PPRNs) will extract the data needed to populate the PCORnet Common Data Model from the data sources that house the necessary information, transform their data to fit into the common data model, and then load that transformed data into a defined location.

**Local Data Partner** – A local data partner is an individual site or organization that contributes data for use by a CDRN or PPRN.

**HMORN (HMO Research Network) VDW (Virtual Data Warehouse)** – HMORN is a network of research centers based in multiple healthcare systems. The VDW is the common data model used by the HMORN research network. It contains data on enrollment, vital signs, pharmacy, and other standardized data elements.

**I2b2 (Informatics for Integrating Biology and the Bedside)** – I2b2 is an NIH-funded National Center for Biomedical Computing based at Partners HealthCare System. The I2B2 Center has developed a software program that organizes and transforms clinical data into tables. The central table contains many different “facts” or observations on a patient, such as patient and provider numbers, the concept that was observed (e.g., diagnosis, medication, procedure, or laboratory test), the value of the concept (e.g., diagnosis code, procedure code, laboratory test result), start and end dates, and other elements.

**Metadata** – Metadata is structured information that describes, explains, locates, and makes it easier to retrieve and use an information resource. It can be thought of as the data equivalent of words in a dictionary (words that describe words). The PCORNet DRN Query Tool will include many different types of metadata, including metadata about the organization, PCORnet DataMarts, and other data resources, such as registries or other research data sets. Examples of organizational metadata include the type of data collected, data models used, and willingness to participate in research. See [PopMedNet Metadata](https://www.popmednet.org) for more information.

**Mini-Sentinel** – The Mini-Sentinel Network is a distributed research network sponsored by the US Food and Drug Administration (FDA) to help monitor the safety of FDA-regulated medical products, such as
medications and vaccines. See Mini-Sentinel for more information about the Mini-Sentinel database and common data model.

**Natural Language Processing (NLP)** – Natural language processing (NLP) software translates unstructured language into structured, and therefore analyzable, data. NLP is often used for such text data as clinical notes. For example, NLP software might transform a clinical note in an EHR of “8-year-old boy presents with concussion” into the following data elements: “male,” “pediatric,” “age 8,” and “concussion.” NLP is a tool that is part of a larger field known as machine learning, where computers are “taught” to read large quantities of text and look for identifiable patterns and structures.

**OMOP (Observational Medical Outcomes Partnership)** – The OMOP program was created to develop methods for analyzing data drawn from varied data sources, such as healthcare claims and EHRs. The program developed a common data model.

**Patient-Generated Data (PGD)** – Patient-generated data (PGD) is data that is created, gathered, or inferred by the patient or his or her proxy to help address a health concern. PGD may include demographics, patient identifiers, health history, symptoms, biometrics, treatment history, lifestyle choices, vital signs, patient-reported outcomes (PROs), and other health-related data. Patients or their proxies are responsible for capturing or recording these data and directing the sharing or distribution of it to healthcare providers and other stakeholders. PGD may be recorded directly or derived from paper medical charts, EHRs, personal health records, personal health devices, or other sources.

**Patient-Reported Outcomes (PROs)** – Patient-reported outcomes (PROs) are defined by the FDA as a report of the status of a patient’s health condition by the patient or his or her proxy without interpretation by a clinician or anyone else. PCORnet’s Patient-Reported Outcomes Task Force has incorporated proxy reports under the umbrella of PROs only if the proxy is the sole source of patient-clinician communication (e.g., cognitive impairment, pediatrics, etc.). PROs may address any aspect of a patient’s health or healthcare experiences, including mental health, physical health, social health, experiences of care, patient-reported behaviors, and treatment effects. They are often measured through standardized, validated instruments.

**Patient Portal** – A patient portal is the portion of an EHR system that is accessible to patients through a Web-based application. Although patient portals vary in their capabilities and functions, many offer ways to track lab test results, communicate with healthcare providers, and make appointments.

**Protected Health Information (PHI)** – PHI (protected health information) is information about patients that is protected from inappropriate disclosure under the privacy and security mandates of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and subsequent related legislation. A list of the 18 personal health information identifiers can be found [here](#).

**PCORnet DRN Query Tool** – The PCORnet DRN Query Tool is the system used to create, operate, and govern the PCORnet DRN. The Query Tool includes a Web portal and a software application (the DataMart Client), which are powered by PopMedNet. Authorized users will be able to query PCORnet data based on chosen criteria via the PCORnet Query Tool.
**PCORnet External Partner** – A PCORnet partner is any other outside participant or funder of a PCORnet study. All external partners will be expected to abide by PCORnet policies and principles as they are developed.

**PCORnet Research Network Partner** – A PCORnet Research Network Partner is a clinical data research network or patient-powered research network that has a contractual arrangement with PCORI to abide by PCORnet policies and principles, standardize their data to the PCORnet CDM, and participate in queries on a voluntary basis.

**PopMedNet™** – PopMedNet is an open-source software application that enables simple creation, operation, and governance of distributed health data networks. See [PopMedNet](https://example.com) for more information.

**Query** – Queries are the primary mechanism for requesting information from a database. Queries can be thought of as questions presented to the database in a predefined query format, written in a language such as Structured Query Language (SQL) or Statistical Analysis System (SAS). For example, to count the number of unique patients in a clinical data warehouse, a programmer might write the following SQL query: “Select count unique (patient_ID) from cdw.enrollment.” In the context of PCORnet, a query is a question or request initiated from the DRN Query Tool that is distributed to selected sites or networks.

**Research Network Data Partner** – A research network data partner is an individual site or organization that contributes data for use by a CDRN or PPRN.

**Source Data** – Source data is the raw data residing in operational healthcare systems, such as electronic health record systems, claims systems, patient registries, and databases containing survey data.