APPENDIX 3: GLOSSARY OF TERMS

This glossary defines key terms used throughout the NEST Mark™: Real-World Data Source Quality Tool. The definitions are intended to promote consistent interpretation of questions and responses across data source holders, NEST Mark tool reviewers, and regulatory stakeholders. Citations, where available, are provided as numbered references at the end of this appendix.

A

AI / ML / LLM (Artificial Intelligence / Machine Learning / Large Language Models)

A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. (1)

ALCOA++ Principles

A framework for data integrity indicating that data are *Attributable, Legible, Contemporaneous, Original, Accurate*. The ++ refers to the additional attributes of *Complete, Consistent, Enduring, Available when needed, and Traceable*. (2)

Audit Trail

A chronological record of system activities that is sufficient to enable the reconstruction and examination of the sequence of events and activities surrounding or leading to an operation, procedure, or event in a security-relevant transaction from inception to results. (1)

Audit (Internal, External, Health Authority Inspections)

Independent review and examination of records and activities to assess the adequacy of system controls, to ensure compliance with established policies and operational procedures. (1)

Automated Data Streams

Data flows that are transferred, refreshed, or ingested electronically through predefined system processes with limited or no manual intervention.

B

Biobank

A repository that collects, processes, stores, and distributes biological specimens and associated data for research purposes. (3)

Bias

Systematic error in the design, conduct, analysis, interpretation, publication, or review of a study and its data that results in a mistaken estimate of a treatment's effect on disease. This systematic error may result from flaws in the method of selecting study participants, in the procedures for gathering or analyzing data, and/or in the decision of how and whether to publish the results. (4)

C

CAPA (Corrective and Preventive Action)

A structured approach to identifying, correcting, and preventing recurrence of quality issues or non-conformances.

CDISC (Clinical Data Interchange Standards Consortium)

An organization that develops standards to support the acquisition, exchange, submission, and analysis of clinical research data. (5)

Claims Data

Data that arise from a person’s use of the health care system and reimbursement of health care providers for that care. (4)

Clinical Outcome Assessments (COAs)

Assessment of a clinical outcome can be made through report by a clinician, a patient, a non-clinician observer or through a performance-based assessment that describes or reflects how a patient feels, functions, or survives. (6)

Clinician-Reported Outcome

A measurement based on a report that comes from a trained health-care professional after observation of a patient’s health condition. (7)

Closed Claims

Claims that have completed processing and adjudication, with a final determination made (paid, partially paid, or denied). They are considered complete and stable for the period covered. (8)

Code Mappings

Documented correspondences between one coding system and another, or between local source codes and a target terminology.

Common Data Model (CDM)

A standardized data structure that enables transformation and analysis across disparate sources.

Conformance

Extent to which data adhere to specific standards, e.g., data structures, formats.

Continuity of Care / Continuum of Care (see also Observability)

Extent of an individual’s pertinent health data, which is captured across settings / environments of care that is represented in the RWD source. (4)

D

Data Accrual

The process by which data are collected and processed in a consistent and methodical manner. (4)

Data Cleaning

Process of identifying and correcting data that are inaccurate, missing, or incomplete. (9)

Data Completeness

The extent to which expected data elements are present and populated.

Data Consistency

The degree to which data are comparable across sites, systems, or time.

Data Curation

Ongoing processing and maintenance of data throughout its lifecycle to ensure long-term accessibility, sharing, and preservation. (9)

Data De-identification

Process of removing personally identifiable information such as names, social security numbers, and street addresses from records or a dataset. (9)

Data Dictionary

Document that outlines the structure, content, and meaning of a given variable. (9)

Data Domain

A thematic grouping of related data elements.

Data Governance

A set of processes that ensure that data assets are formally managed throughout the enterprise. A data governance model establishes authority, management, and decision-making parameters related to data produced or managed by the enterprise. (9)

Data Holder / Data Source Holder

The organization responsible for stewardship of the real-world data source.

Data Imputation

Process of replacing missing or incomplete data values with estimated values using defined statistical or rule-based methods.

Data Integrity

Guarding against improper modification or destruction of data including assurance of the origin and authenticity of data. (1)

Data Integration

The process of combining data from multiple systems.

Data Interoperability

Ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways. (9)

Data Latency (Time Lag)

The duration between data capture and availability for research.

Data Lineage (see also Data Provenance)

Documented trail that accounts for the origin of a piece of data and where it has moved from to where it is presently. (9)

Data Linkage (see Linkage)

Process of matching and combining data from multiple databases (1)

Data Management Plan (DMP)

A written document outlining how health care data are collected, stored, secured, shared, and preserved throughout the lifecycle.

Data Model

Data model is a conceptual model that defines how data are organized, stored, and connected.

Data Partner

An organization that creates, owns, manages, aggregates, or provides health care data that can be used for research purposes, including entities such as health systems, data aggregators, academic institutions, and health care registries.

Data Privacy

Protection of personal information under applicable laws.

Data Provenance (see also Data Lineage)

Documented trail that accounts for the origin of a piece of data and where it has moved from to where it is presently (9)

Data Re-identification

Processes by which de-identified data is matched with other data to identify the individual to whom the data belongs.

Data Representativeness

The extent to which the data source or study sample reflects the target or intended-use population to which the findings are meant to generalize.

Data Retention

Policies and practices governing how long data and related documentation are kept, in what form, and under what disposal or archival rules.

Data Schema

A schema is the formal implementation of a data model, defining the exact structure such as tables, fields, and data types.

Data Source

A structured or semi-structured system from which RWD are obtained.

Data Suppression

Intentional withholding, masking, or aggregation of sensitive values or cells in outputs to reduce disclosure risk.

Data Traceability

The ability to track data from origin through processing and release.

Deterministic Matching

A linkage approach in which records are matched using an exact match on a unique identifier or a set of common identifiers.

Device-Generated Data

Data produced directly by medical devices.

Device Identifier (DI)

The fixed portion of a Unique Device Identifier.

E

Electronic Health Record (EHR)

Record of health-related information on an individual that conforms to nationally recognized/utilized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff. (4)

Encounter

A recorded interaction of any type between a patient and the health care system as defined by the data source.

ETL (Extract, Transform, Load)

Data integration process that combines, cleans and organizes data from multiple sources into a single, consistent data set for storage in a data warehouse or other target system. (10)

External Researcher

An individual or group outside the data partner or primary data-holding organization that accesses or uses one or more data sources to conduct RWD studies. May also be called data users or end users.

F

False Linkage Rates

Proportion of different individuals that are erroneously linked among all linked individuals.

Fit for Purpose

The suitability of a data source, data capture process, or dataset for its intended regulatory or research use, considering whether it is adequate to address the specific research question.

G

Generalizability

The extent to which the study sample is adequately representative of the target population of interest. (4)

L

Large Language Model (LLM)

A type of AI model trained on large text datasets to learn the relationships between words in natural language.

Linkage

Process of matching and combining data from multiple databases.

Link Accuracy

Proportion of correct links (true positive links divided by total links conducted).

Linkage Matching Rate

Proportion of matches linked (true positive links divided by total matches conducted).

Longitudinality

The extent to which data support observation over time

M

Machine Learning (ML)

The development and use of computer systems that adapt and learn from data with the goal of improving accuracy. (1)

Metadata

Information describing the characteristics of a database. Examples include structural metadata describing data structures (e.g., data format, syntax, and semantics) and descriptive metadata describing data contents (e.g., information security labels). (1)

Missed Linkage Rates

Proportion of records belonging to the same individual that fail to be linked among linkable individuals.

N

Natural Language Processing (NLP)

Technology that allows computers to interpret, manipulate, and comprehend human language.

NEST (National Evaluation System for health Technology)

An initiative of the Medical Device Innovation Consortium focused on advancing the quality, generation, and use of real-world evidence for regulatory decision-making for medical devices through diverse, multi-stakeholder collaborations across the medical device ecosystem. (11)

NEST Mark™: Real-World Data Source Evaluation

A study question agnostic, expert-developed, and FDA guidance-informed evaluation of a real-world data source’s foundational capabilities (governance, provenance, quality management, privacy and security, and transparency) to determine whether a data source can reliably generate high quality real-world data that may support real-world evidence for medical device regulatory submissions (12)

O

Observability

Extent of the individual’s pertinent health data which is captured across settings / environments of care that is represented in the RWD source. (4)

Open Claims

Claims that have been submitted but are not yet fully processed or adjudicated. They may include preliminary or partial information and can change as additional services, corrections, or payment decisions are recorded. (8)

P

Patient Reported Outcome

A type of Clinical Outcome Assessments (COA) based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else. (13)

Personally Identifiable Information (PII)

Any data that can identify or locate a specific individual.

Point of Care

The clinical setting where care is delivered.

Population Representativeness

The degree to which a data source reflects a target population.

Privacy-Preserving Data Access

Methods that allow for the authorized analysis, sharing, or processing of data while ensuring that sensitive individual information (PII) remains secure.

Privacy Preserving Transformations

Data techniques applied to reduce re-identification risk.

Probabilistic Matching

A linkage approach using less restrictive steps than deterministic linkage, allowing records to be linked based on exact matches on some fields together with partial or approximate agreement across others.

Q

Quality Assurance

Planned, systematic activities designed to provide confidence that data processes and outputs meet predefined quality standards**.**

Quality Control

Operational checks performed on data or processes to detect errors, deviations, or anomalies and verify that quality requirements are being met.

Quality Management System (QMS)

An organized framework of policies and procedures to meet quality objectives.

R

Real-World Data (RWD)

Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. (4)

Real-World Evidence (RWE)

Clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from analysis of RWD. (4)

Re-identification

Process of restoring or determining the identity of an individual from de-identified data by linking it back to identifying information.

Research-Ready Data Source

A processed and quality-assured version of a data source suitable for research.

Role-Based Access Control (RBAC)

A security model in which access to data systems and resources is granted based on a user's assigned role, rather than on individual user identity. (1)

S

Security Controls

Actions, devices, procedures, techniques, or other measures that reduce the vulnerability of an information system. (1)

Standard Operating Procedure (SOP)

Established or prescribed methods to be followed routinely for the performance of designated operations or in designated situations. (14)

Structured Data

Data recorded in predefined fields, formats, or coded elements that are readily computable and queryable, such as diagnosis codes, procedure codes, medication fields, or lab result fields.

T

Target Population

The specific group of individuals which a study, survey, or intervention aims to investigate or apply its findings.

Tokenization

A method replacing identifiers with tokens to enable linkage while protecting privacy.

Transformation

Conversion of data into standardized or analytical formats.

U

Unique Device Identifier (UDI)

A code that uniquely identifies a medical device. (15)

Unstructured Data

Data not stored in standardized computable fields. Examples include narrative notes, scanned documents, or free text, that often require abstraction or NLP before analysis. (4)

V

Validation

Confirmation that systems or processes meet predefined requirements.

Version Control / Versioning

Processes used to manage and document changes over time.

References

1) National Institute of Standards and Technology (NIST). Computer Security Resource Center. Glossary. Available at: https://csrc.nist.gov/glossary. Accessed April 30, 2026.

2) European Medicines Agency (EMA). Guideline on computerised systems and electronic data in clinical trials. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf Accessed April 30, 2026.

3) National Institutes of Health (NIH). Biorepositories and Biospecimen Research Branch. Available at: https://dctd.cancer.gov/programs/cdp/organization/bbrb. Accessed April 28, 2026.

4) U.S Food and Drug Administration (FDA). Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. 2025. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices. Accessed April 30, 2026.

5) Clinical Data Interchange Standards Consortium (CDISC). Standards Catalog. Available at: https://www.cdisc.org/. Accessed April 28, 2026.

6) FDA. Clinical Outcome Assessment (COA) Compendium. Available at: https://www.fda.gov/drugs/development-resources/clinical-outcome-assessment-compendium. Accessed April 28, 2026.

7) FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-. Glossary. 2016 Jan 28 [Updated 2025 Jan 16]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK338448/ . Co-published by National Institutes of Health (US), Bethesda (MD).

8) International Society of Pharmacoeconomics and Outcomes Research. ISPOR. Fit-for-Purpose Real-World Data: An Integral Component of Evidence Planning. Available at: https://www.ispor.org/publications/journals/value-outcomes-spotlight/vos-archives/issue/view/real-world-evidence-in-healthcare-decisions/fit-for-purpose-real-world-data--an-integral-component-of-evidence-planning. Accessed April 30, 2026.

9) National Library of Medicine (NNLM). Data Glossary. Available at: https://www.nnlm.gov/resources/data/data-glossary. Accessed April 30, 2026.

10) IBM. What is ETL (Extract, Transform, Load)? Available at: https://www.ibm.com/think/topics/etl Accessed April 30, 2026.

11) FDA. National Evaluation System for Health Technology (NEST) Program Documentation. Available at: https://www.fda.gov/about-fda/cdrh-reports/national-evaluation-system-health-technology-nest Accessed April 28, 2026.

12) NEST Coordinating Center. NEST Mark™ Program Materials. Available at: https://nestcc.org/nest-mark/ Accessed April 28, 2026.

13) BEST. Glossary - BEST (Biomarkers, EndpointS, and other Tools) Resource - NCBI Bookshelf. Available at: https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-CEndpointS. Accessed April 30, 2026.

14) Merriam-Webster. Available at: https://www.merriam-webster.com/dictionary/standard%20operating%20procedure. Accessed April 29, 2026.

15) FDA. UDI Basics. Available at: https://www.fda.gov/medical-devices/unique-device-identification-system-udi-system/udi-basics. Accessed April 30, 2026.