National Institutes of Health
Division of Program Coordination, Planning, and Strategic Initiatives

NIH Real World Data Network

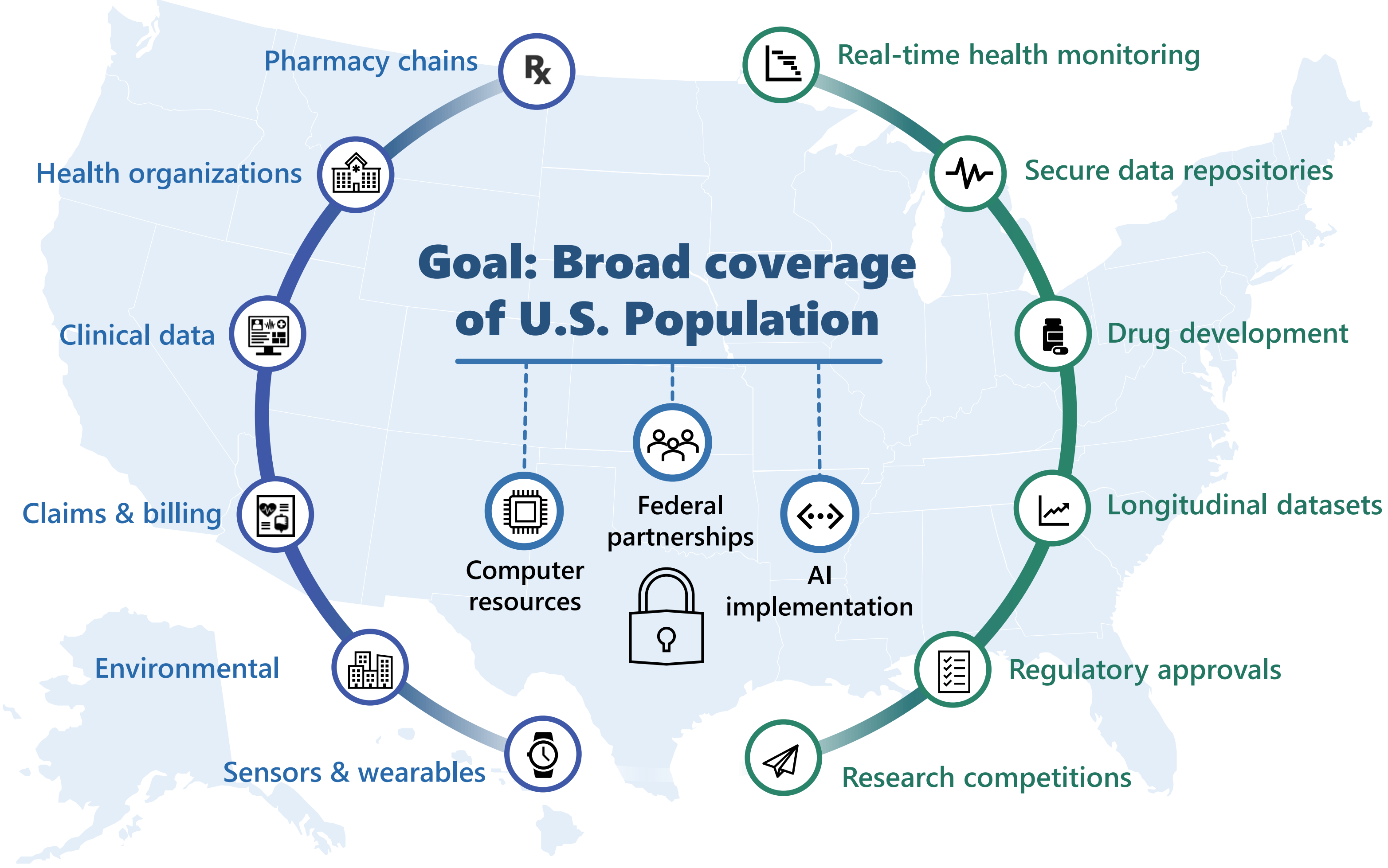
NIH Council of Councils
September 12, 2025

Nicole C. Kleinstreuer, Ph.D.
Acting NIH Deputy Director for Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health

Real World Data Network (RWDN)

The overarching goal of the **Real World Data Network** (RWDN) is to create a **robust** and **secure** national data infrastructure for accelerating cutting-edge research to address public health concerns.

RWDN Data Sources & Opportunities



Data Sources

- Pharmacy chains
- Health organizations
- Clinical data
- Claims & billing
- Environmental
- Sensors & wearables

Goal: Broad coverage of U.S. Population

- Computer resources
- Federal partnerships
- AI implementation

Opportunities

- Real-time health monitoring
- Secure data repositories
- Drug development
- Longitudinal datasets
- Regulatory approvals
- Research competitions

National Institutes of Health
Division of Program Coordination, Planning, and Strategic Initiatives

NIH Real World Data Network

Private and Public Data. Protection of Patient Confidentiality. Researcher Access.

Partnerships

- Build partnerships across HHS and with external stakeholders
- Expand data sources for real-world evidence
- Study chronic disease, neurodevelopmental disorders (e.g. autism)

Development

- Leverage existing NIH programs (e.g. N3C) to fast-track progress
- Uphold maximal data security standards
- Protect patient confidentiality

Timeline

- Define go/no-go milestones and progress checkpoints
- Pilot the Network within six months of project launch

Partnership with CMS

NIH, CMS Partner to Advance Understanding of Autism Through Secure Access to Select Medicare and Medicaid Data

May 7, 2025 — The partnership will focus first on enabling research around the root causes of autism spectrum disorder.

- NIH signed a data use agreement under CMS’ Research Data Disclosure Program focused on Medicare and Medicaid enrollees with a diagnosis of ASD.
- Teams at CMS and NIH will establish a secure tech-enabled mechanism leveraging N3C to enhance data sharing with timely, privacy and security compliant data exchange.
- Autism pilot research program will inform continued development of an NIH Network to be used by researchers in understanding healthcare utilization, chronic disease etiology and treatment, and the economic burden of chronic conditions.

Prior Experience for the Real World Data Network

- The **National Clinical Cohort Collaborative (N3C)** is a nationwide resource of real-world health data stewarded by NCATS in collaboration with 75+ institutions that researchers are using to speed medical research.
- The **LINKAGE** program links NIA-funded study data with existing datasets from Centers for Medicare & Medicaid Services (CMS) and other sources via a cloud-based environment to support data accessibility and sharing.
- **All of Us** is a national resource of hundreds of thousands of patients who have consented for genomic testing and linkage to their health records and clinical data.
- **These NIH RWD enclaves have the highest level of data security and all access to these data are strictly controlled.**

National Clinical Cohort Collaborative (N3C)

National
Clinical
Cohort
Collaborative

- N3C is a national resource of real-world data that researchers are using to speed medical research.
- N3C systematically collects data derived from electronic health records from different institutions and harmonizes these data into data enclaves. The enclaves offer spaces for collaborative research. In addition, the enclaves can connect to other data sets, which creates a more complete picture of health outcomes.
- **Implements a comprehensive, multi-layered approach to address the security of the data enclaves and protect patient privacy.**

NIH National Center for Advancing Translational Sciences

National Clinical Cohort Collaborative (N3C)

National
Clinical
Cohort
Collaborative

- Initially created in response to the COVID-19 pandemic, NCATS is piloting the use of the original N3C infrastructure for additional health conditions.

Clinical
Cohort
Cancer
Tenant

N3C Cancer

Linking data from NCI's Surveillance, Epidemiology, and End Results (SEER) program with N3C EHR data to capture the full trajectory of cancer patients' care, characterize variations in cancer treatments, and ascertain impact on patient outcomes.

Clinical
Cohort
Renal
Tenant

N3C Renal

Inter-HHS agency data-sharing enclave to better understand and conceptualize the patient journey through chronic kidney disease to dialysis, transplant, and beyond.

Streamlining Efficient RWD Collection: DataCOUNTS

- Partnership with healthcare organizations and Institutions to embed advanced technologies for data collection.
- Provides the ability to capture standard and non-standard data from medical encounters.
- Allows for lineage and provenance of data to ensure reproducibility and transparency
- Lowers barriers and cost for data collection and formatting
- **Only de-identified data is provided for RWD Network with full audits**

Nebraska Medicine

SANFORD HEALTH

TGH Tampa General Hospital

UH University Hospitals Cleveland Medical Center

Baylor College of Medicine

EMORY HEALTHCARE

Cincinnati Children's

University of Kentucky

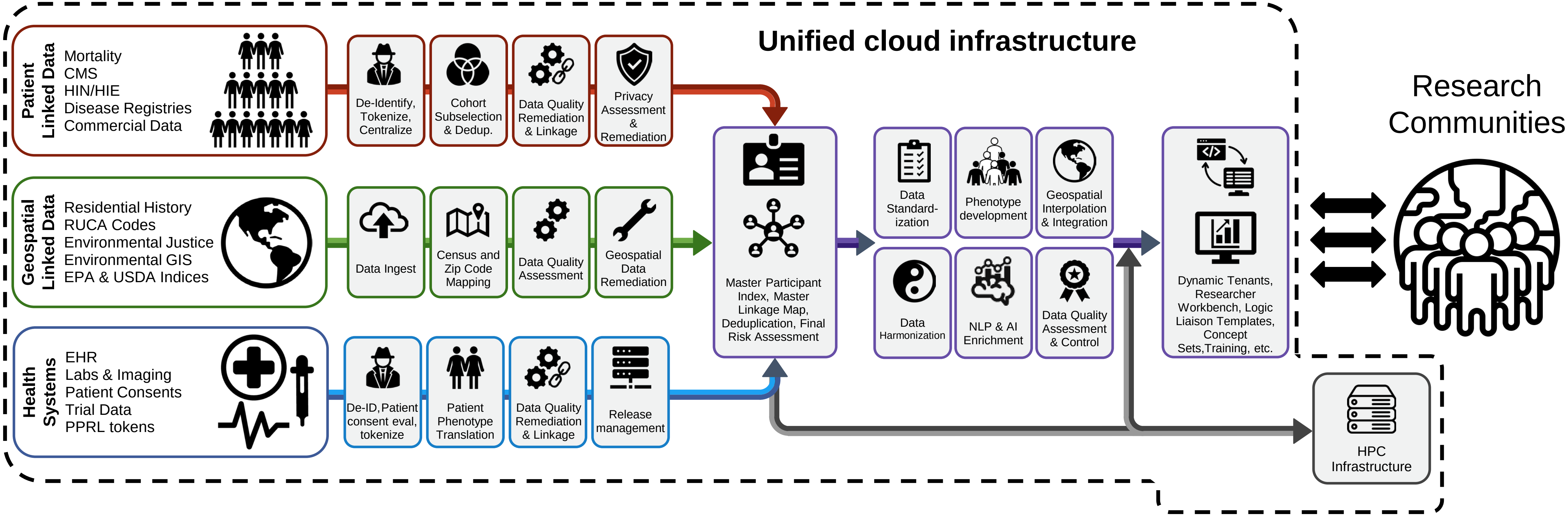
THE UNIVERSITY OF TEXAS MD Anderson ~~Cancer~~ Center Making Cancer History

The UNIVERSITY of OKLAHOMA HEALTH SCIENCES

UCI Health

NIH National Institutes of Health Division of Program Coordination, Planning, and Strategic Initiatives

NIH Real-World Data Network (vision)



Unified cloud infrastructure

Patient Linked Data
Mortality
CMS
HIN/HIE
Disease Registries
Commercial Data

De-Identify, Tokenize, Centralize
Cohort Subselection & Dedup.
Data Quality Remediation & Linkage
Privacy Assessment & Remediation

Geospatial Linked Data
Residential History
RUCA Codes
Environmental Justice
Environmental GIS
EPA & USDA Indices

Data Ingest
Census and Zip Code Mapping
Data Quality Assessment
Geospatial Data Remediation

Health Systems
EHR
Labs & Imaging
Patient Consents
Trial Data
PPRL tokens

De-ID, Patient consent eval, tokenize
Patient Phenotype Translation
Data Quality Remediation & Linkage
Release management

Master Participant Index, Master Linkage Map, Deduplication, Final Risk Assessment

Data Standard-ization
Phenotype development
Geospatial Interpolation & Integration
Data Harmonization
NLP & AI Enrichment
Data Quality Assessment & Control

Dynamic Tenants, Researcher Workbench, Logic Liaison Templates, Concept Sets, Training, etc.

HPC Infrastructure

Research Communities

- Single integrated Network merging capabilities of N3C, DataCOUNTSand other NIH resources. **Direct connection across HHS, e.g. ARPA-H Data Commons, FDA Sentinel and integration across HHS OpenData Landscape.**
- Eliminates redundancies from separate data collection, linkage, and analytics infrastructures.
- Dramatically reduces administrative overhead (single set of agreements, streamlined governance).
- High-Performance Computing (HPC) integration via Interconnection Security Agreement (ISA).
- Direct access to advanced computational resources (high-throughput analytics, ML/AI model training, petabyte-scale storage and computation).

NIH National Center for Advancing Translational Sciences