NCATS

COLLABORATE. INNOVATE. ACCELERATE.

National Clinical Cohort Collaborative

Past, present, and future

Hythem Sidky, PhD

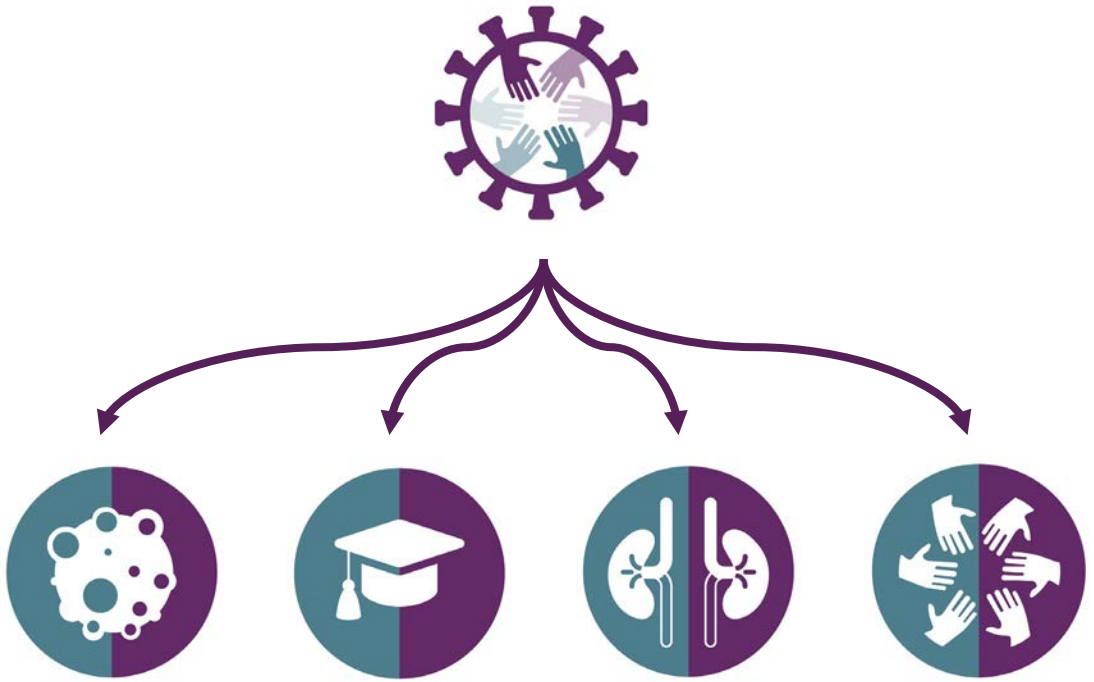
National Center for Advancing Translational Sciences

Why N3C?

- Urgent need for observational data at scale.
- In the US, there is no centralized health care, and therefore no centralized health care data.
- Data from a single person is spread across multiple providers across time and geography.

Why Expand?

- COVID-19 served a powerful proof-of-concept, demonstrating the value of large-scale clinical data harmonization and collaboration.
- Significant investments and teamwork have created a robust platform and governance model that can be expanded for broader use.
- Beyond COVID-19, chronic diseases, health disparities, and complex conditions remain critical public health priorities.
- Broader utilization of N3C infrastructure provides valuable opportunities for training researchers, clinicians, and students in real-world data analysis, informatics, and collaborative science.



N3C Past and Present

N3C has grown from a COVID-19 response into a national translational research infrastructure, combining harmonized EHR data, scalable governance, and team science to accelerate discovery across diseases and institutions.

2020

N3C is Launched!

In response to the COVID-19 pandemic, the National COVID Cohort Collaborative was formed to create the largest publicly available, harmonized EHR data set in U.S. history.

Today

N3C's impact

5000+ citations, H-index 33, 1589 authors. N3C enabled transformative research and care guidelines, disease definitions, and predictive models for outcomes across comorbidities.

2022–2023

Phase 1 Clinical Pilot

N3C successfully expanded beyond COVID-19, piloting clinical tenants for Alzheimer's, COPD, and Renal disease across 12 institutions.

2024–2027

Phase 2 Clinical Pilot

Building on Phase 1, Phase 2 scales with enhanced PPRL, data integration (e.g., CMS, SEER), and supports new tenants like cancer and renal.

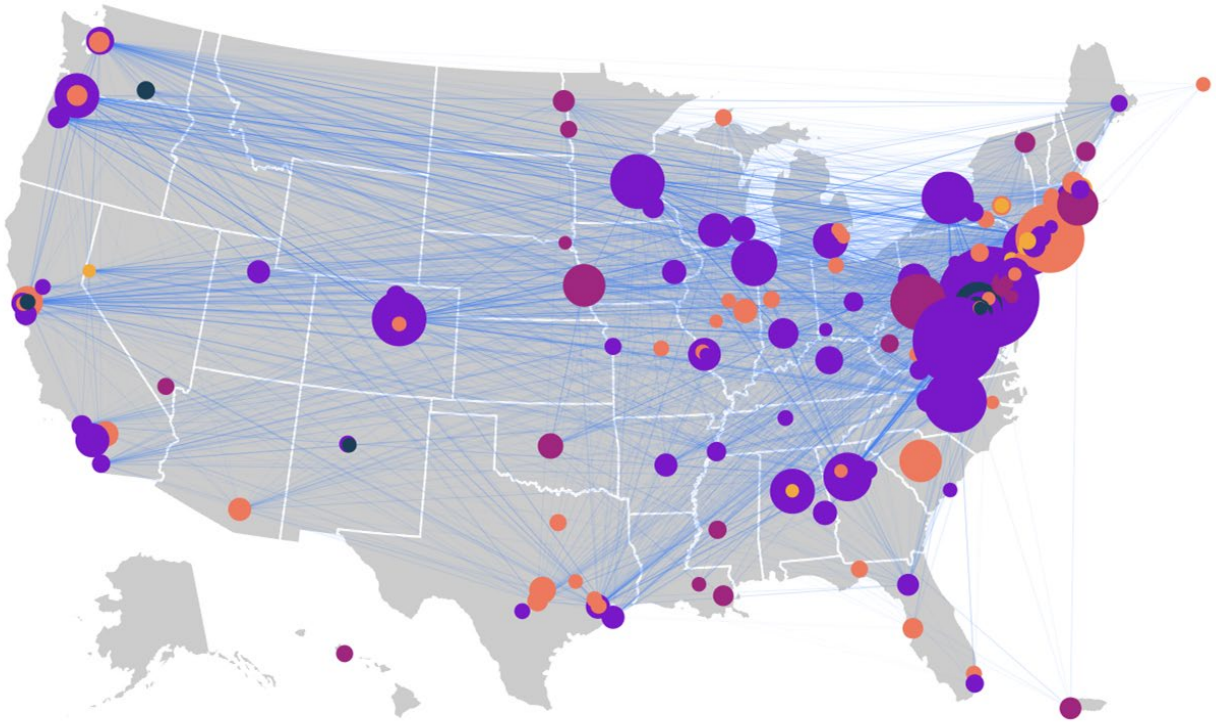
Continued Community Engagement

96 Data Contributors signed the original COVID Data Transfer Agreement

76 Data Contributors signed the COVID Data Transfer Agreement Extensions

12 institutions participated in the Phase I Clinical Pilot

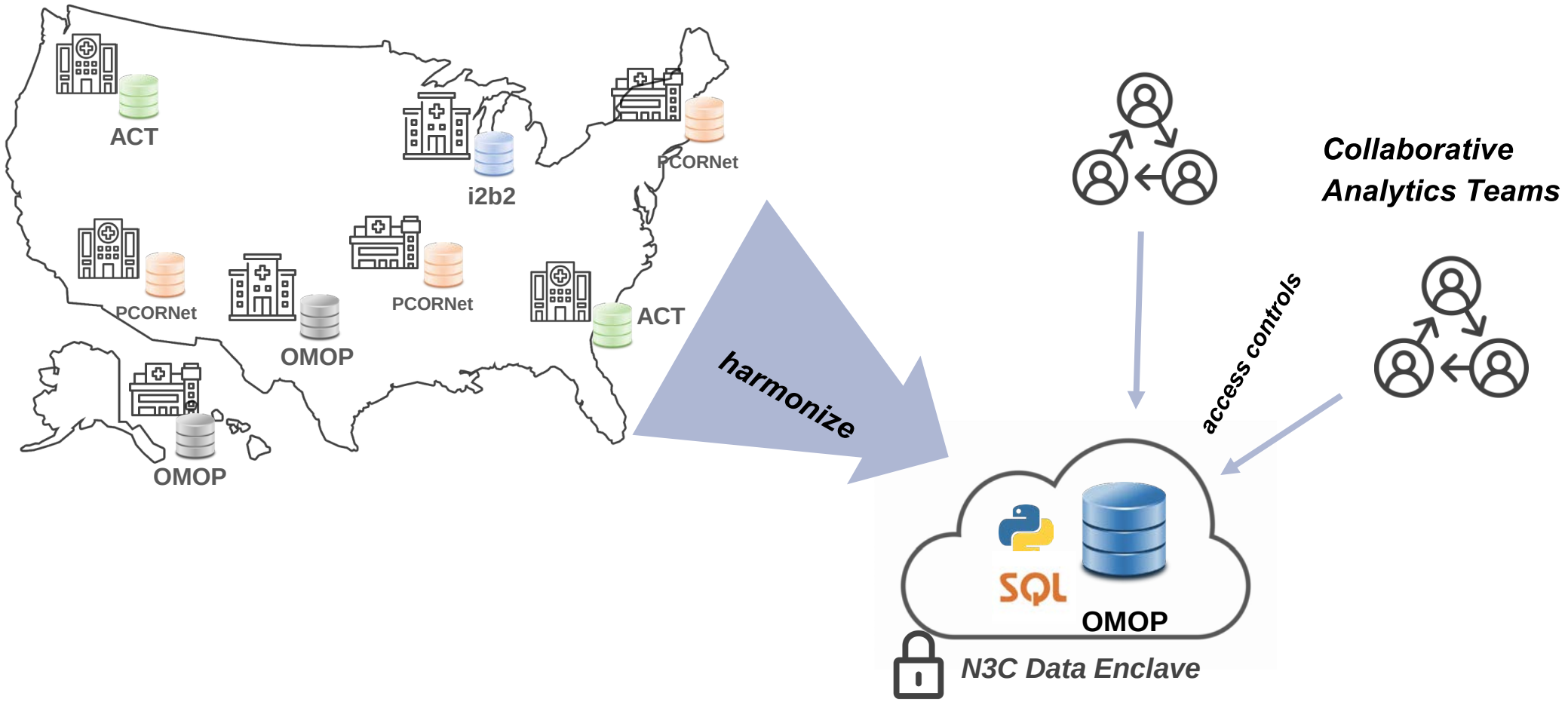
18 institutions are participating in the Phase II Clinical Pilot



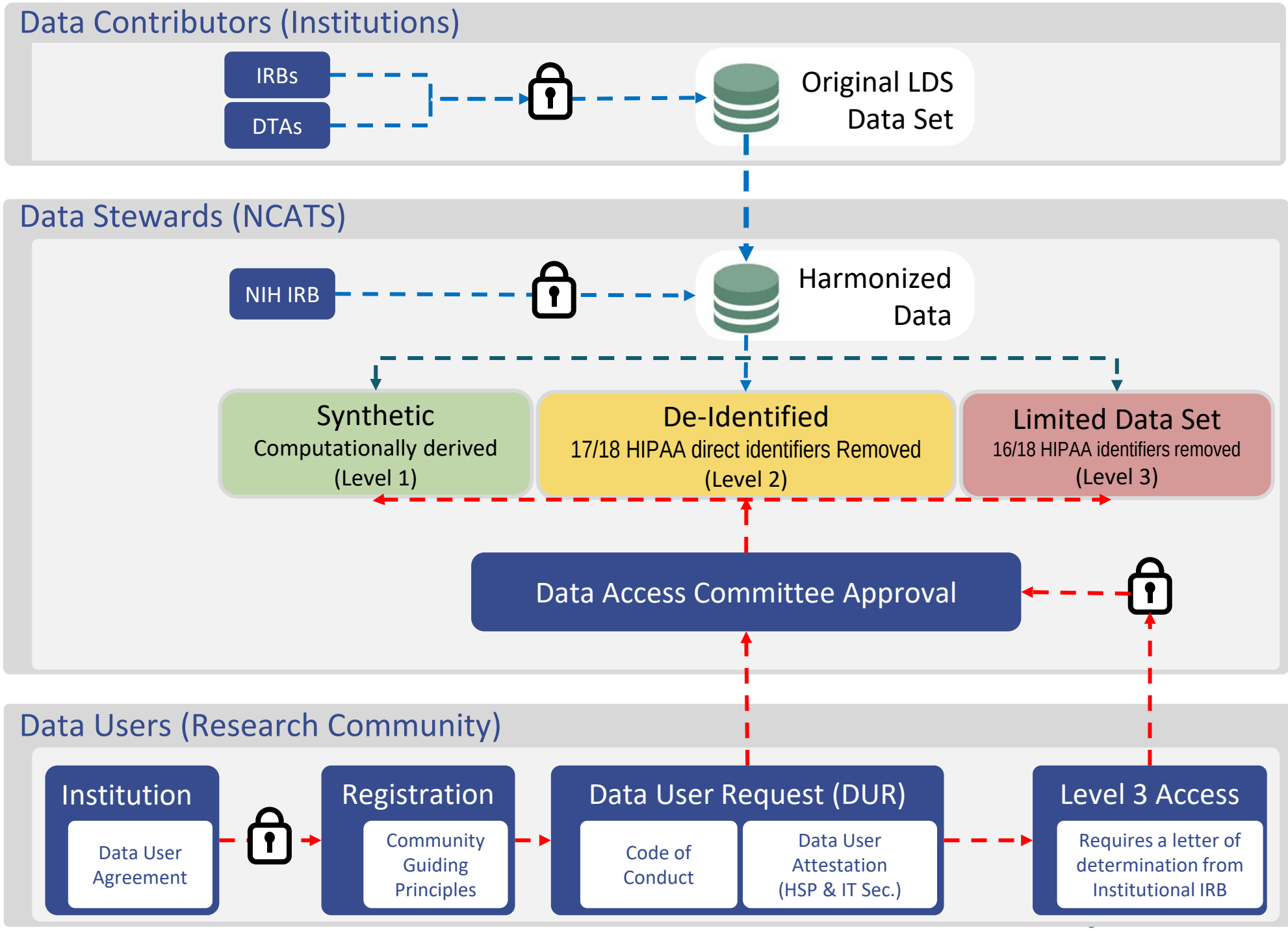
National
Clinical
Cohort
Collaborative

How it works

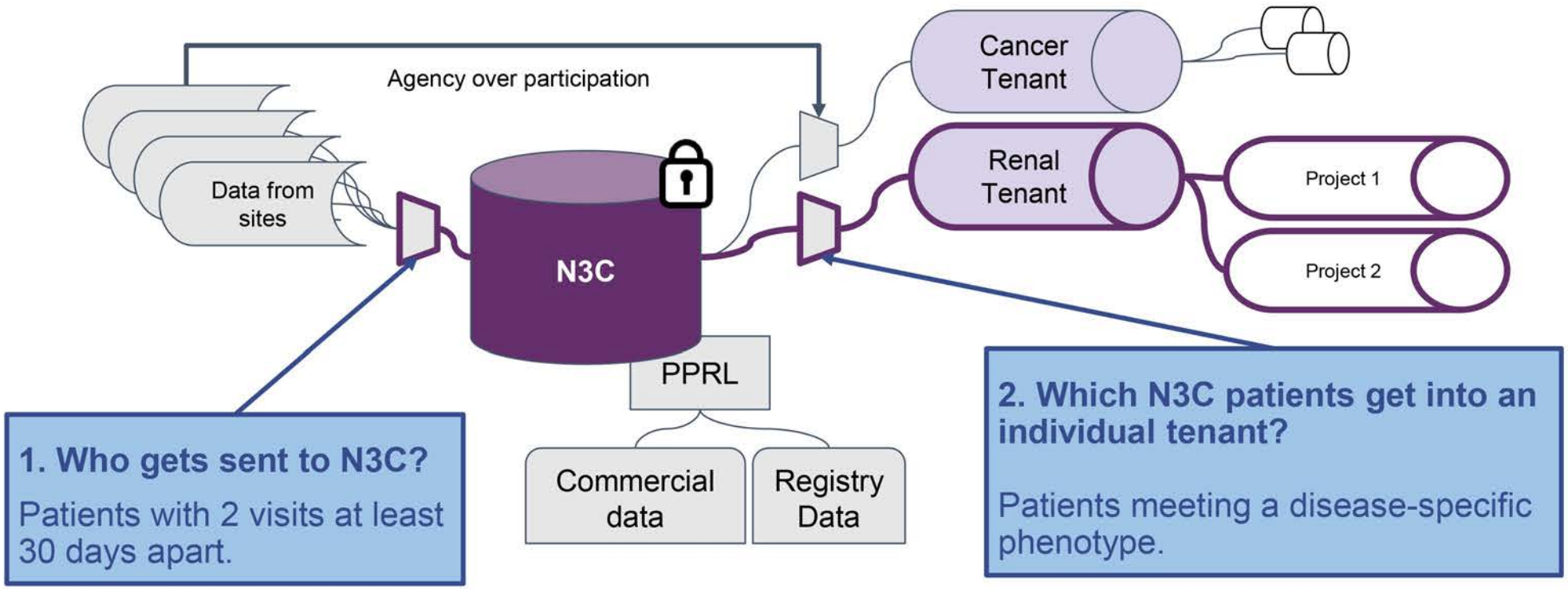
N3C: High Level



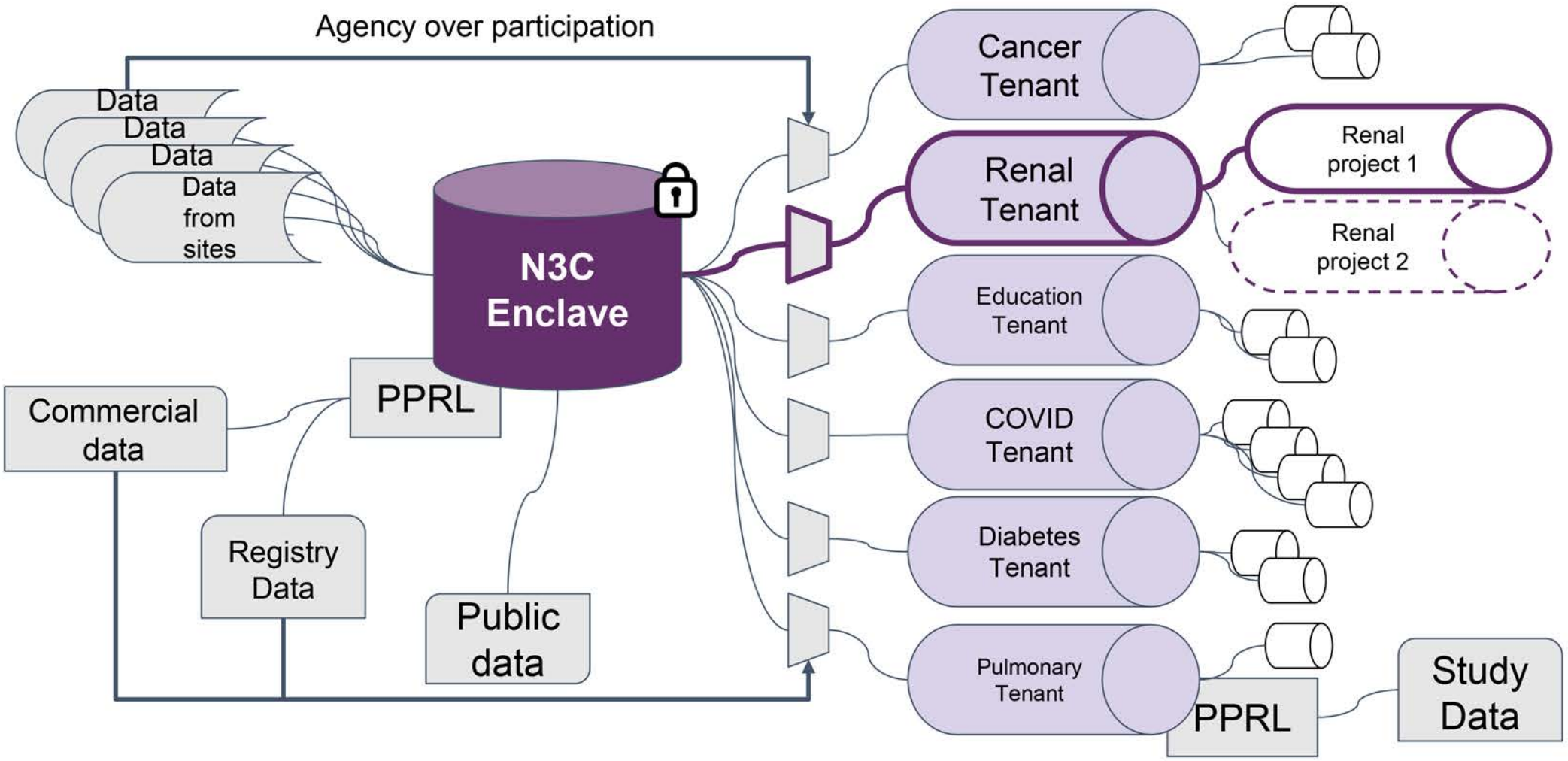
N3C: Data Governance and Access



N3C Phase 2 Clinical Pilot Model



N3C - Envisioned Future



National Center for Advancing Translational Sciences

Objectives of the N3C Clinical Pilots

- N3C Clinical pilots were meant to help NCATS more accurately understand the financial, infrastructure, and community resources needed to develop and maintain future tenants.
- Pilots will facilitate refining operations, governance and technical architecture.
- Establish partnerships with CMS, HRSA, NCI, NIDDK.
- Next-generation health care interoperability is being developed (HL7 FHIR US Core).
- New capabilities will expand the space of scientific questions that can be asked and answered.

Feedback We have Received

- Streamlined and robust agreement tracking
- More granular institutional user management
- More granular institutional control of data
- Time-limited investigator access to institutional data
- Interest in disease areas beyond the pilots
- Maximize use of linked data sets
- A lot more!

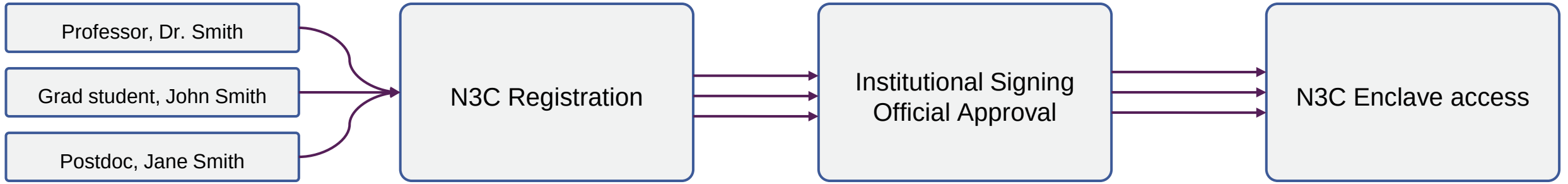
National
Clinical
Cohort
Collaborative

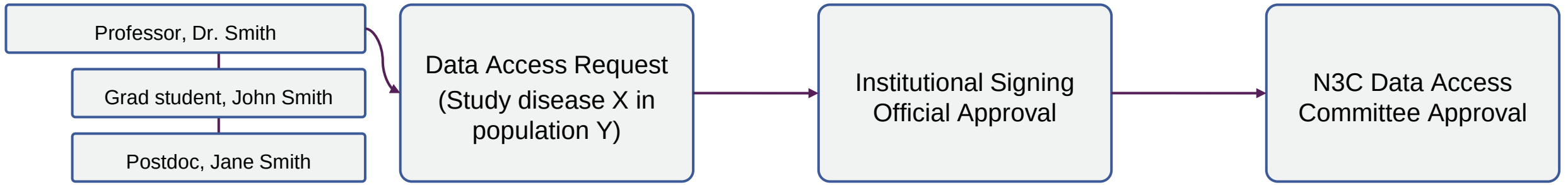
N3C: Data Access Requirements

New Data Access Requirements

- N3C is a Controlled Access Data Repository (CADR)
- Data Use Requests (DURs) are now Data Access Requests (DARs)
- Institutional Signing Officials (ISOs) will manage their users and their DARs
- DARs can be submitted by a Data Access Requester who:
  - Is a permanent employee at their institution, holding a position equivalent to at least a tenure-track professor or senior researcher. This does not include lab technicians, trainees, postdoctoral researchers, and graduate students.
  - Has oversight responsibility for all additional individuals named on the DAR who will be granted access to the data.
  - Will ensure that all data usage fully aligns with the terms of the institution's Data Use Agreement (DUA) and institutional policies governing data access and use.
- Individuals who do not meet these criteria (e.g., trainees or graduate students) may participate in projects but must be included under the supervision of an eligible Data Access Requester.

CADR Requirements in Action





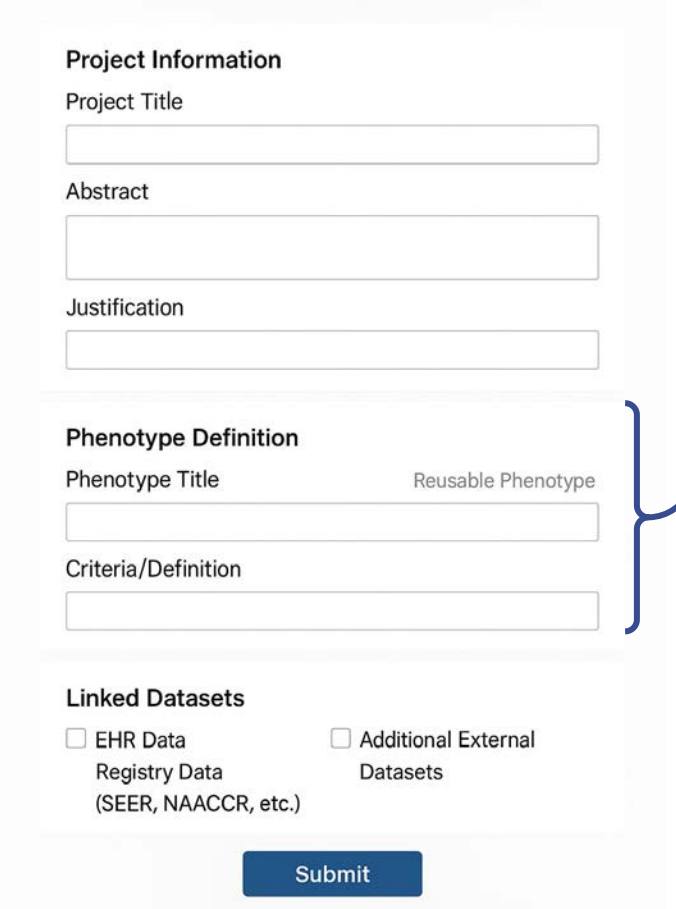
National
Clinical
Cohort
Collaborative

N3C: The Next Phase

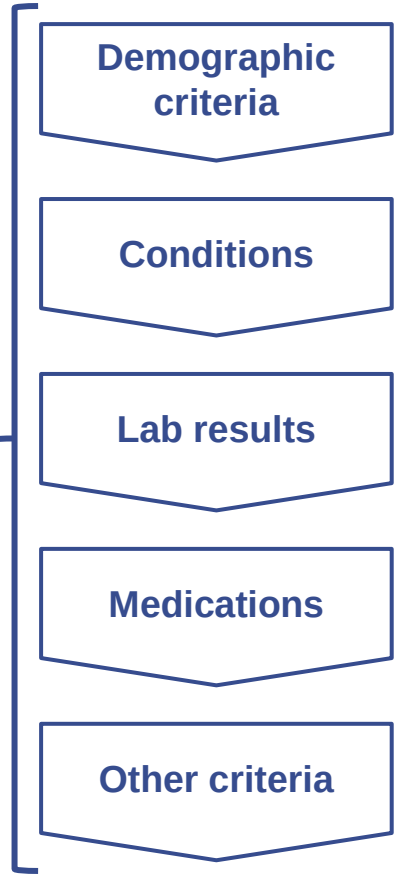
Scaling Across Clinical Domains

Challenge: Each new disease area requires significant time and effort to design cohorts, onboard data, and build dedicated environments.

Data Access Request



Phenotype builder



The dynamic model shifts the paradigm

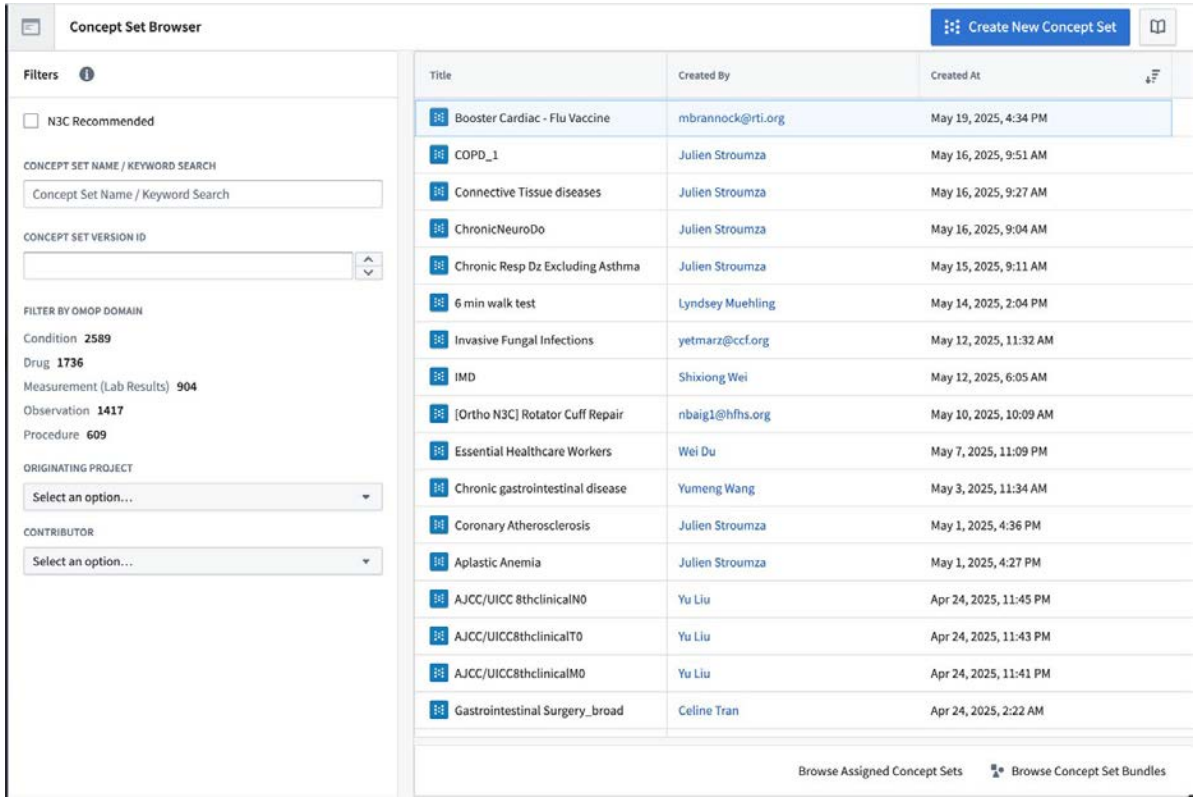
By embedding phenotype logic directly into each Data Access Request, we allow data environments to be created automatically and tailored to the research question, not a pre-built domain.

This enables rapid, reproducible research without duplicating infrastructure, and keeps institutional control and data governance fully intact.

Data access aligned precisely with scientific intent.

Phenotype Development as a Community

Challenge: Phenotype development is challenging. Development is often ad hoc, inconsistent, and siloed, making it difficult to reproduce results or build upon previous work.



A Community-Driven, Transparent Phenotype Ecosystem

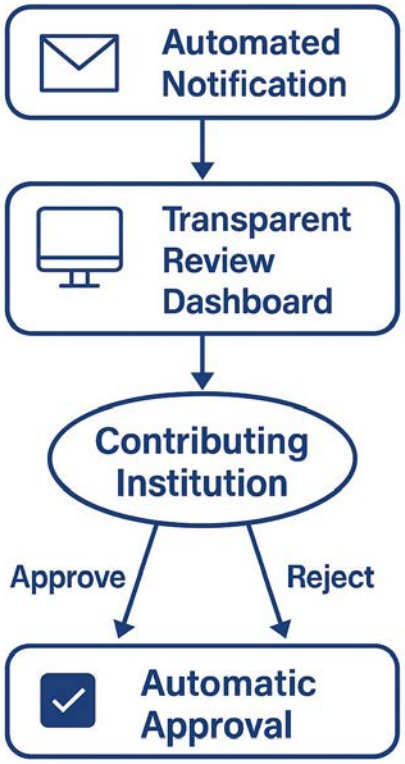
All phenotypes are versioned and made available for reuse by others in the enclave.

A dedicated “phenotype sandbox” allows researchers to build, test and refine phenotypes safely. No data contributor approval required.

Enables researchers to build on one another’s work, similar to N3C’s concept sets, promoting consistency, reproducibility and shared standards.

Balancing Scale with Institutional Control

Challenge: As the number of research projects and data-contributing institutions grows, maintaining review and approval workflows for each project becomes resource-intensive and slow.



Opt-Out Model for DARs Including Contributing Institution

Dynamic Opt-Out Model: Scalable Oversight

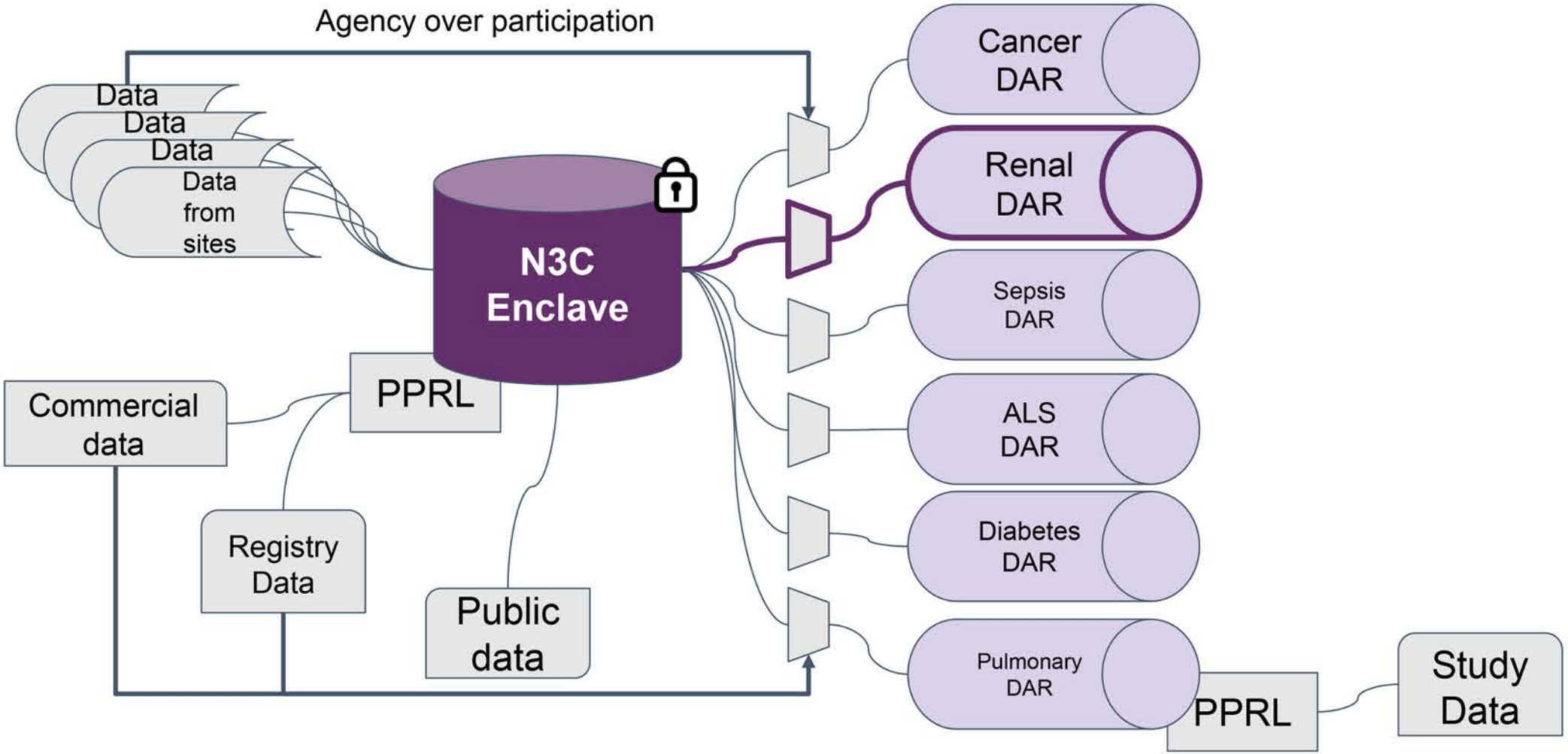
- Institutions are automatically alerted when a new Data Access Request (DAR) includes their data.
- Designated institutional representatives review incoming DARs in a centralized interface.
- Institutions can explicitly approve, reject or take no action. If no objection is raised within a defined window, the request is automatically approved.
- Institutions retain per-DAR control, ensuring data is only used in projects they implicitly or explicitly support.

A Refined Set of Agreements

- Data Contributor will have control over which Data Access Requests have access to their Data.
- Data Access Requests will have a 1 year term and be eligible for 2 renewals (3 years total).
- Same HIPAA Limited Data Set.
- Data linkage through Privacy-Preserving Record Linkage (PPRL) included.
- Disease agnostic phenotype.

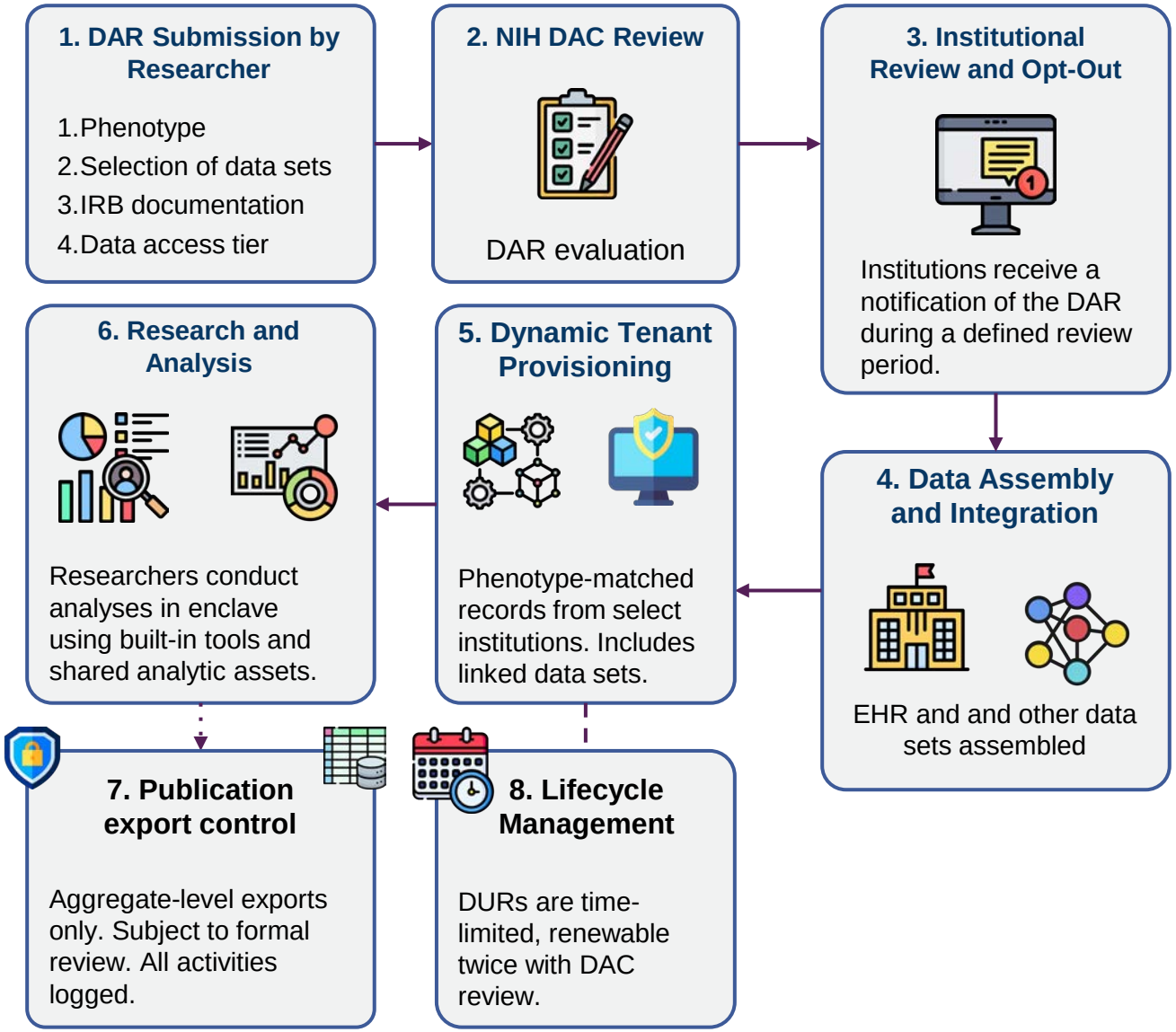
The new agreements safeguard institutional control while enabling scalable, secure data sharing. They lay the foundation for a more flexible, future-ready research ecosystem.

N3C: Dynamic Tenants



National Center for Advancing Translational Sciences

Data Access Life Cycle



1. **DUR Submission**
   Researchers submit a Data Access Request (DAR) with a computable phenotype, IRB documentation, access tier, and selected data sets.
2. **NIH DAC Review**
   The Data Access Committee reviews the DAR for probability of technical success, policy alignment, and privacy.
3. **Institutional Review**
   Institutions whose data match the phenotype can review and opt out. Lack of response is considered approval.
4. **Data Assembly & Integration**
   The platform assembles the relevant data set(s).
5. **Dynamic Tenant Provisioning**
   A secure workspace is provisioned with approved data, linked external data sets. Access is restricted to authorized users.
6. **Secure Research**
   Analysis occurs within a FedRAMP-authorized enclave.
7. **Publication Data Export & Oversight**
   Only aggregate results are exportable, following privacy review..
8. **Lifecycle Management**
   DURs are time-limited (1 year) and renewable (twice). Phenotypes may be reused across studies.

National Center for Advancing Translational Sciences

National
Clinical
Cohort
Collaborative

Why Participate?

Imagine if...

You want to study sepsis with 5 other institutions

Without N3C

Kickoff

Administrative Setup
Identify partners, Initiate data-sharing agreement negotiations, Start individual IRB processes

Month 4

Legal Negotiations
Finalize multiple institution-specific agreements, address privacy/security/legal compliance across multiple institutions.

Month 7

Technical Infrastructure Setup
Build/procure secure collaborative platform, establish compute environment (high cost).

Month 10

Data Harmonization
Independently clean and harmonize data at each institution. Link external data sets (high cost).

Month 13

Project Start
Data ready for analysis after significant delays.

With N3C

Kickoff

Setup
Submit a DAR. Single Master Data Transfer Agreement (already executed). Rapid institutional opt-out review process

Week 3

Data & Workspace Provisioning
Creation of secure dynamic tenant. Data already harmonized to OMOP. Immediate linkage to external data (CMS, mortality data, and more).

Week 5

Project Start
Begin collaborative analysis in secure, scalable environment.

National Clinical Cohort Collaborative

NIH National Center for Advancing Translational Sciences

Why participate in N3C?

- **Minimal Data Contribution Burden**
  - NCATS provides resources, guidance, and support.
  - One-time, simplified institutional agreement.
- **Reduced Fragmentation**
  - Centralized, scalable platform prevents consortium duplication.
  - Rapidly launch multi-institutional studies.
- **Advanced Data Cleaning & Harmonization**
  - N3C handles data cleaning, standardization, and returns enhanced data sets to contributors.
- **Data Quality Reporting**
  - Contributors receive detailed quality scorecards to guide internal improvements.
- **Data Linkages**
  - Immediate access to difficult-to-obtain data sets.
- **Lower Cost, Greater Efficiency**
  - Dramatically reduces setup time, administrative burden, and operational costs.

National
Clinical
Cohort
Collaborative

Next Steps

What to Expect

- Website updates coming soon. N3C moving to https://n3c.ncats.nih.gov/
- New agreements under final review. Expect them for execution shortly.
- Interested in helping us stand up this new model? We're looking for early adopters!
- N3C is working with other NIH resources like Data COUNTS and NIA LINKAGE.
- Regular Community Forum relaunch coming soon.
- Watch for Google Forms to collect thoughts, feedback on relevant notices (e.g. Protecting Human Genomic Data), and help us rejuvenate and organize Domain Teams.

General inquiries:
NCATS_N3C@nih.gov

Direct contact:
hythem.sidky@nih.gov

National Center for Advancing Translational Sciences

National
Clinical
Cohort
Collaborative

Questions?