


## TECHNICAL REPORT

# Cross-Network Directory Service: Infrastructure to enable collaborations across distributed research networks

Jessica M. Malenfant<sup>1</sup>  | Jenny Hochstadt<sup>1</sup> | Bridget Nolan<sup>1</sup> | Kimberly Barrett<sup>1</sup> | Dean Corriveau<sup>2</sup> | Daniel Dee<sup>2</sup> | Marcelline Harris<sup>3</sup> | Chayim Herzig-Marx<sup>1</sup> | Vinit P. Nair<sup>4</sup> | Zachary Wyner<sup>1</sup> | Jeffrey S. Brown<sup>1</sup>

<sup>1</sup>Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, Massachusetts

<sup>2</sup>Avacoda LLC, Amherst, Massachusetts

<sup>3</sup>School of Nursing, University of Michigan, Ann Arbor, Michigan

<sup>4</sup>Government & Academic Research, Humana, Inc., Comprehensive Health Insights, Louisville, Kentucky

## Correspondence

Jessica Malenfant, Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401 East, Boston, MA 02215.

Email: [jessica\\_malenfant@harvardpilgrim.org](mailto:jessica_malenfant@harvardpilgrim.org)

## Funding information

US Food and Drug Administration through the Department of Health and Human Services (HHS), as agents for the Patient-Centered Outcomes Research Trust Fund, Grant/Award Number: HHSF2232014000301/HHSF22301006T

## Abstract

**Introduction:** Existing large-scale distributed health data networks are disconnected even as they address related questions of healthcare research and public policy. This paper describes the design and implementation of a fully functional prototype open-source tool, the Cross-Network Directory Service (CNDS), which addresses much of what keeps distributed networks disconnected from each other.

**Methods:** The set of services needed to implement a Cross-Directory Service was identified through engagement with stakeholders and workgroup members. CNDS was implemented using PCORnet and Sentinel network instances and tested by participating data partners.

**Results:** Web services that enable the four major functional features of the service (registration, discovery, communication, and governance) were developed and placed into an open-source repository. The services include a robust metadata model that is extensible to accommodate a virtually unlimited inventory of metadata fields, without requiring any further software development. The user interfaces are programmatically generated based on the contents of the metadata model.

**Conclusion:** The CNDS pilot project gathered functional requirements from stakeholders and collaborating partners to build a software application to enable cross-network data and resource sharing. The two partners—one from Sentinel and one from PCORnet—tested the software. They successfully entered metadata about their organizations and data sources and then used the Discovery and Communication functionality to find data sources of interest and send a cross-network query. The CNDS software can help integrate disparate health data networks by providing a mechanism for data partners to participate in multiple networks, share resources, and seamlessly send queries across those networks.

## KEYWORDS

cross-network communication, cross-network discovery, data network infrastructure, distributed health data networks, network ecosystem, network interoperability

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. Learning Health Systems published by Wiley Periodicals, Inc. on behalf of the University of Michigan

## 1 | INTRODUCTION

The growing adoption of distributed health data networks to facilitate large-scale evidence generation studies, as well as other public health activities, provides an opportunity to leverage those investments to create a national resource that enables viable learning health systems (LHS) that continuously drive data into knowledge and knowledge into practice.<sup>1-3</sup> A digital infrastructure is recognized as a core component for LHS success, including infrastructure that enables the work of distributed health data networks. The U.S. health care system, along with health care systems across the globe, are characterized by data siloes defined by local health system structures and payment systems. The U.S. health care system has siloes defined by factors such as health insurer, provider, and public health agencies. Systems outside the United States have similar silo characteristics, with additional siloes related to age group, geography, and type of care (eg, medication dispensing). Although each system is unique, the challenges associated with siloed data are consistent across the globe.

Existing large-scale distributed health data networks include the Centers for Disease Control and Prevention's (CDC) Vaccine Safety Datalink, the U.S. Food and Drug Administration's (FDA) Sentinel System, the Health Care Systems Research Network (HCSRN), the NIH Health Care Systems Research Collaboratory, the Patient-Centered Outcomes Research Institute's (PCORI), National Patient-Centered Clinical Research Network (PCORnet), and the Observational Health Data Sciences and Informatics (OHDSI) program. These networks enable collaborators to maintain physical and operational control of their data while making multidatabase analysis more secure and feasible.<sup>4-7</sup> Together, the individual investments in each of these networks can be leveraged to expand overall capabilities across funding agencies and the broader public health community, improve opportunities to generate shareable knowledge, and provide extensible infrastructure for the development of LHS.<sup>8-13</sup>

Broadly, the goal of these networks is to create multisite multiuse network structures and governance to facilitate implementation of studies using real-world data to generate real world evidence. Each network uses a common data model (CDM) approach to standardize data and has built analytic tools to facilitate use of the data. Although the networks share many similarities in data sources, data models, and approach to distributed analytics with standardized toolkits, each network has unique features related to governance, available data, data curation approaches, and restrictions on use that make it difficult to easily navigate the ecosystem. Although the networks have demonstrated the substantial benefits realized from establishing distributed networks, the networks have not yet been able to meet a longer-term goal of efficiently leveraging the entirety of the health data network ecosystem to support more robust generation of real-world evidence. To date, each network and the individual sites within remain largely siloed and disconnected. Five important limitations contribute to keeping these networks disconnected and impede collaboration across networks:

1. Networks have different governance policies and different requirements for participation.
2. There is no mechanism for broadcasting research capabilities—the types of data available and the research and clinical expertise of their staffs—in a way that facilitates discovering common research interests and gives network participants control over who sees what.
3. Between networks, there is no secure and reliable means of making data requests and tracking response activity.
4. There are no operational standards or metrics for describing data at a level that enables researchers to judge fitness-for-use of others' data sources.
5. There is no reliable mechanism for sending queries that will execute correctly across networks with different CDMs.

The Cross-Network Directory Service (CNDS) was developed to address these limitations by creating the infrastructure and technical substrate to enable *cross-network* collaboration. CNDS is intended to help foster collaboration by helping researchers ask questions such as “Does anyone have the data I need to implement my comparative effectiveness study in COPD?”; “Which sites have biorepository data linked to administrative claims data?”; and “Who has patient-reported data on depression?”

From discussions with stakeholders and a review of existing meta-data curation projects, we prioritized the creation of a system with maximum flexibility and extensibility to adapt to changes in metadata requirements rather than trying to define all the specific data elements and variables that should be included in a network. CNDS is meant to serve as the underlying infrastructure to connect people, organizations, networks, and systems. This paper describes the design and implementation of CNDS prototype—an open-source tool that was designed to overcome much of what keeps distributed networks from collaborating with each other.

## 2 | METHODS

This project built and pilot tested the CNDS across two existing networks: FDA's Sentinel and PCORnet. Both networks use the open source platform, PopMedNet (PMN), to facilitate the implementation and operation of distributed health data networks.<sup>14</sup> This section describes the implementation details.

The overall scope of the project included the following:

- Design and develop web services that communicate with PMN to sync the metadata and related information about people, organizations, and data sources between the networks
- Implement a general-purpose data model flexible enough to capture nearly any metadata element desired
- Develop functionality to distribute requests across multiple PMN networks
- Demonstrate the ability to register and discover data sources external to a network and communicate with (ie, send a request) these data sources via PMN requests

## 2.1 | System design and requirements gathering

The initial system design work was drafted by the study team composed of representatives from the Harvard Pilgrim Health Care Institute, the FDA, Humana Comprehensive Health Insights, Inc (a Sentinel data partner), the Department of Learning Health Sciences, University of Michigan (a PCORnet data partner), and Avacoda LLC (the software developer). Early in the project, the study team met with a broad group of stakeholders to discuss the project goals and solicit feedback, which informed the system design and beta version of the CNDS software. The stakeholder group included representatives from a range of organizations, including academic institutions, health systems, health services researchers, contract research organizations, and the pharmaceutical industry.

Four major CNDS functional features were identified: governance, metadata capture, querying, and communication. There was a clear need identified to establish a CNDS governance mechanism to account for privacy, confidentiality, data sharing, and proprietary information policies of existing distributed networks and the individual participating organizations. Stakeholders also informed the metadata component of CNDS and illustrated the complexities of how to define and curate information about people, organizations, and data sources. System requirements for enabling the ability to discover potential collaborators and querying across networks were initially discussed at these stakeholder and study team meetings.

## 2.2 | System description

A key design decision was to ensure that the architecture would be flexible and extensible. Given that CNDS would be used to connect distinct health data networks, we elected to implement it with PMN software application as our base technology.<sup>14</sup> PMN supports distributed within-network querying for Sentinel, PCORnet, MDPHnet, HCSRN, the HCSRN Cancer Research Network, the Biologics and Biosimilars Collective Intelligence Consortium (BBCIC), the Reagan-Udall Foundations Innovation in Medical Evidence and Development Surveillance (IMEDS), and the NIH Health Care Systems Research Collaboratory Distributed Research Network, among others.

PMN provides capabilities for creating and managing distributed networks, including capturing information about participating organizations, users, queryable data sources, and registries. Additionally, PMN provides the functionality for creating, distributing, and responding to queries and provides an extensive suite of access controls that can be configured at the network, project, and user levels. These access controls grant the ability to determine at a very granular level what users can and cannot do within PMN.<sup>15</sup>

## 2.3 | Governance

The need for granular software-enabled governance and administration via visibility rules and access controls was a key need identified by stakeholders and the study team. The CNDS design enables visibility rules entered in metadata (via the Registration function) and enforced when users search for organizations or data sources via the

Discovery function and when they attempt to send data requests through the Communication function. These rules identify who is authorized to see each organization and data source metadata element based on information about the requesting party and how widely the information owner has indicated willingness to share. Visibility can be imagined as a set of widening circles—each subsequent layer permits more users to view the metadata. Information owners can tag metadata elements as being visible to:

- No one (ie, just myself and the system administrators)
- Registrants in my PMN-based network
- Registrants in any PMN network
- All CNDS registrants

The PMN access controls are available to allow CNDS to control every aspect of use of the application, for example, adding, editing, deleting, and viewing users, organizations, and DataMarts; responding to, rejecting, and uploading results; managing security; and running audit reports. Additional access controls implemented in CNDS govern actions such as who can manage metadata, send a cross-network request, or set visibility. Table 1 provides a list of the CNDS access controls as they relate to discovery, registration, and administration.

## 2.4 | Registration

Registration enables users to request an account; enter and edit metadata and information about themselves, their organizations, and their data resources; and determine what data others can see via the visibility settings. The user-entered information creates the metadata database and directory described below. Though referred to as Registration, users can update their information at any time, not just during the initial setup process. The ability to register in CNDS independent of network affiliation extends distributed networks beyond their boundaries.

## 2.5 | Discovery

Discovery enables users to explore the metadata database, via a user interface dynamically generated from the data model, to find new data sources and potential collaborators. Users search based a set of criteria that matches the metadata information filled in by the organizations and data source owners. The result set returned from a search is constrained by visibility levels set by the metadata owners.

## 2.6 | Communication

Communication enables users to send and receive data requests both within and across networks. PMN provides functionality for creating, distributing, and responding to data queries within a single PMN distributed network. There are multiple “query request types” available in PMN for users to send “questions” to data sources, such as a simple point-and-click query interface and secure file distribution where an

**TABLE 1** Cross-Network Directory Service (CNDS) access controls

| Access Control                   | Description   |
|----------------------------------|---|
| Discovery                        |   |
| Search CNDS                      | Governs whether the user sees the “Search” menu item used to access CNDS search and therefore whether the user can access CNDS search functionality. No additional levels of governance are applied for accessing search. Users without this permission cannot see the “Search” option in the CNDS menu.        |
| Communication                    |   |
| Create CNDS request              | Governs the ability to create a request that will be sent to DataMarts in and out of network. Users who have this permission can create a request from the results of a Discovery search. Existing PMN permissions govern all other request creation functionality (e.g., edit, copy, and distribute requests). |
| Map request type                 | Governs the ability to associate a request type in one network with a request type in another network. Users without this permission cannot see the “Manage Request Type Mappings” option in the CNDS menu.   |
| Administration                   |   |
| Manage metadata                  | Governs the ability to perform all functions related to metadata management including adding, editing, deleting domains, and assigning domains to organization and/or data sources. Users without this permission cannot see the “Manage Metadata” option in the CNDS menu.                                     |
| Manage CNDS Access & Permissions | Governs the ability to set CNDS permissions for security groups and assign users to CNDS security groups. Users without this permission cannot see the “permissions” option in the CNDS menu.   |
| Create CNDS security group       | Governs the ability to create a CNDS security group   |
| Edit CNDS security group         | Governs the ability to edit the description/name of a CNDS security group. (note: It does not govern the ability to assign permissions to the security group. This is covered by the access control “Manage CNDS Access & Permissions”).  |
| Delete CNDS security group       | Governs the ability to delete a CNDS security group. Deleting is performed by clicking “remove” in associated row of the security group table. Deleting will remove the group from the CNDS database and all profiles to which it is assigned.  |

investigator sends an analytic program (eg, SAS) to sites where they run the program locally and return the aggregate results.<sup>16-18</sup>

CNDS extends these capabilities across networks by mapping the request types used by multiple networks to enable each network to process these external requests. CNDS users can send and receive requests, regardless of network affiliation, according to the governance rules of the recipients. Because of different CDMs used by different distributed data networks (DDNs), not all data requests are appropriate to send; CNDS anticipates this by enabling the configuration of appropriate request types.

In PMN, request types are defined to express questions investigators wish to ask. Questions are sent to selected DataMarts via a chosen request type (eg, file distribution). Request types are subject to local governance controls and security policies at both the network and project levels. A project is an entity within a network that allows for users and DataMarts to be grouped according to investigator questions, request types, security policies, and governance. For example, a group within a network that is working on obesity research can be set up as a “project” that includes a subset of the larger network’s DataMarts and request types. One DataMart can be a part of multiple projects.

Traditionally, in PMN, the combination of a project, request type, and DataMart is defined as a route. Requests can only be sent via routes to DataMarts within the same project. CNDS expands this by enabling questions to be sent to DataMarts across projects and networks (ie, “external” routes). To accomplish this, a CNDS system administrator creates mappings that define allowed external routes. An external route is defined as a combination of a network, project, request type, and DataMart.

Since a request type in one network is defined independently from a request type in another, CNDS depends on the CNDS administrator to correctly identify the external route that can service a request type created in the network initiating the request. Discovery may return DataMarts that have and are willing to share the data of interest, but the necessary route must be in place for CNDS Communication to handle the request.

## 2.7 | Data model

As noted above, CNDS rests on a flexible metadata model designed to accommodate an unlimited number of metadata elements. Each metadata element can apply to one or more system entities, and each element is of one metadata type, as described below. Users with sufficient rights can determine what metadata and information are available to be captured on users, organizations, and data sources. These administrative users can add, edit, or delete metadata elements and value sets. Notably, the CNDS metadata data model enables changes to metadata elements without software redesign or programming.

A flexible data model was developed to store the information entered. CNDS is not meant to re-create other professional networking platforms or registries, but to set a foundation upon which future integrations with such systems is possible via application programming interfaces (APIs).<sup>19-21</sup>

The following system entities, which exist in PMN, also exist in CNDS:

- Users: Investigators, data source owners, and researchers are examples of users; in the prototype CNDS, all users must be part of an existing PMN-based network.
- Organizations: Health plans, integrated healthcare delivery networks, and other institutions are examples of organizations.
- Data sources: Queryable data marts, registries, and clinical research databases are examples of data sources.

Metadata fields in CNDS can be associated with the organizations and/or data source system entities. For example:

- “Willingness to accept data requests” could be associated with data sources, but not organizations.
- “Clinical Trial Expertise” could be associated with users and organizations, but not data sources.
- “Data Models” could be associated with both data sources and organizations.

## 2.8 | Metadata types

The available metadata data types are container, text, whole number, TRUE|FALSE, reference, and Boolean group. References can be single or multiselect. Most of the data types are conventional and self-explanatory except container and Boolean group, which can both contain other data types within them and thereby allow for the creation of hierarchy among metadata elements. Container has no intrinsic value while Boolean group does (ie, TRUE|FALSE). This functionality allows data elements to be organized in a searchable hierarchically. An example of a hierarchy of metadata elements is:

Types of Encounters: Inpatient encounters: Inpatient diagnosis codes: Inpatient diagnosis code types.

CNDS enables data partners to describe the types of data and information they collect and the systems they use within their organizations.

## 2.9 | Metadata management

The metadata model was designed to be extensible and flexible, with a goal of simplifying additions to the model. For any new attribute or metadata element about a user, organization, or data source, one would navigate to the CNDS metadata management function to create a new Domain. As illustrated in Figure 1, Domain defines the individual metadata entries as well as the associated hierarchy. For example, as shown in Figure 2, the Types of Data Collected is the highest level, modeled as a Group, with children for Inpatient Encounters and Demographics, both of which have their own set of children attributes. These attributes would be associated with an EntityType (ie, the user, organization, and/or data source) in the DomainUse section of the model.

## 2.10 | Web services architecture

CNDS is designed as a web service with the metadata database described above and invoked using an API that enables communication between web applications. Implementing CNDS using API calls between PMN and CNDS makes CNDS feel like part of PMN while insulating PMN and CNDS from each other and enables changes to either system without affecting the other. Figure 3 is a high-level depiction of the CNDS architecture. What is important to understand is that CNDS is a collection of web services, that is, a collection of functions or utilities that can be invoked from any distributed network. As web services, CNDS does not offer an out-of-the-box user interface. Instead, each distributed network's user interface must be adapted to take advantage of CNDS services, which we demonstrated with two PMN instances.

## 2.11 | Request workflow

As illustrated in Figure 4, once a user discovers a data source of interest, and that data source is willing to accept out-of-network requests, the investigator can then distribute a PMN request to the data source. The request is routed via the CNDS web services from Network 1 to Network 2. PMN is configured so that requests cross-network requests are captured in an “inbox” or PMN project separate from the core network section of the app.

# 3 | RESULTS

We implemented the CNDS design described in Section 2 as an extension to the PMN software application. As part of the implementation and testing, we created demonstration versions of PMN for Sentinel and PCORnet with the new CNDS interfaces and functionality. The workgroups then populated user, organization, and data source information in the CNDS database using the PMN-like interface. The pilot CNDS implementation is currently hosted in a test environment. Two mock websites representing the Sentinel and the PCORnet networks participating in CNDS represent how CNDS would work in production.

## 3.1 | User interface

Because the CNDS metadata model is highly extensible, the user interface that displays metadata must be similarly flexible and extensible. This requirement pertains both to the interface through which metadata values are entered and updated and to the interface for exploring metadata, ie, Discovery.

This project adapted the demonstration instances of the PCORnet and Sentinel PMN; we created new user interfaces for metadata management modules and “profile pages”—screens on which users can update information about themselves, their organizations, and their data sources. These screens basically re-engineered the existing PMN profile interfaces to be dynamically generated by the CNDS data model. Similar interfaces were also created to capture the visibility governance related to who can see users' information. Profile pages

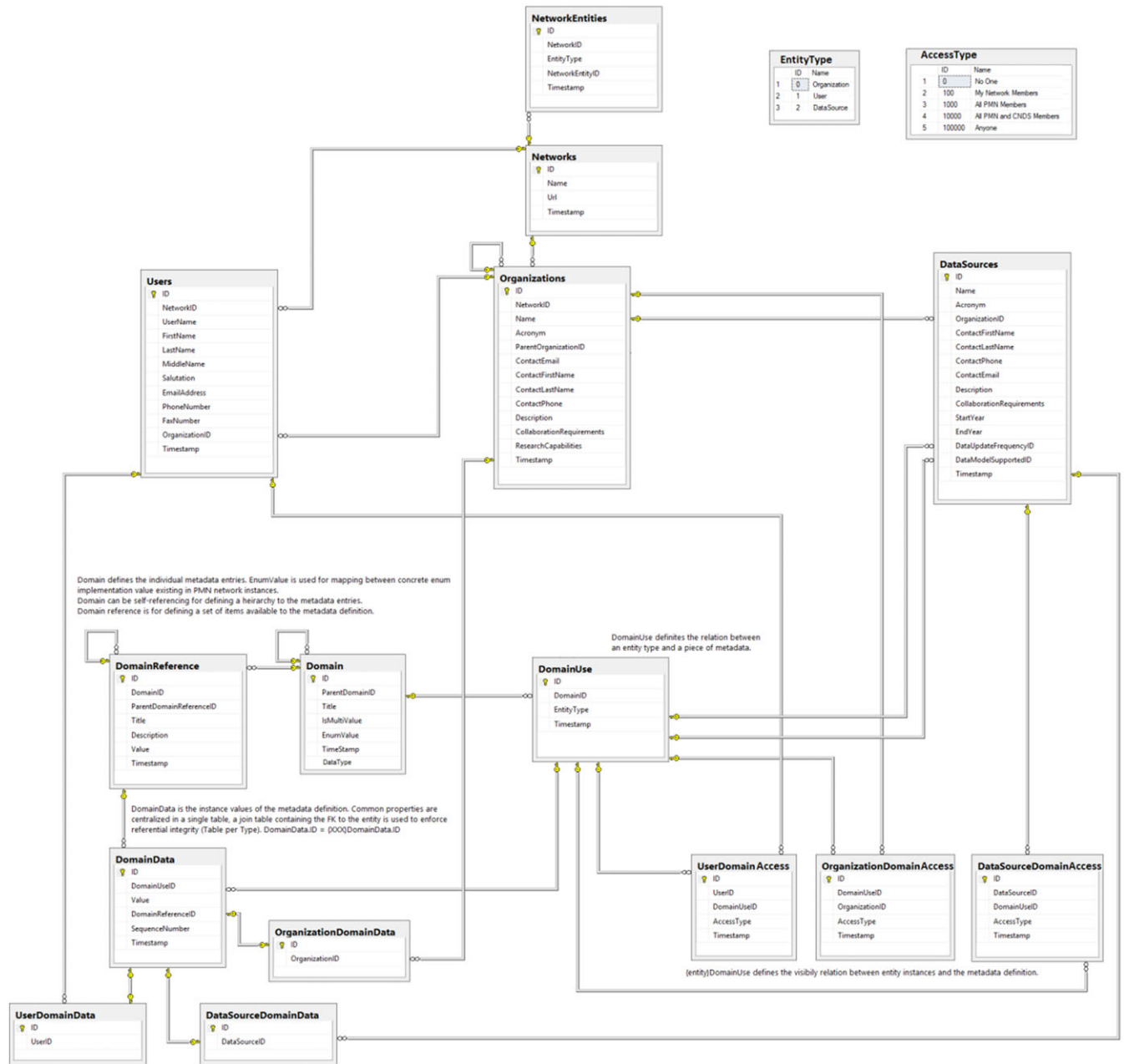


FIGURE 1 Metadata physical data model

are dynamically generated each time such a page is accessed or refreshed, according to the most current metadata values. Similarly, Discovery was developed as additional tools in these PMN instances, is also flexible, and includes an automated dynamic data-driven user interface. In this way, the application does not require reprogramming as the metadata catalog and standards change.

Figure 5 illustrates the Discovery functionality. In this example, a Sentinel user searched for data sources that collect biorepository information. The search returns one fictitious PCORnet data source and associated contact information.

The Sentinel user could “discover” the PCORnet data source because, in registration, the data source administrator had indicated both that the data source includes biorepository information and the “governance” is that this fact can be visible outside the PCORnet network (Figure 6 and Figure 7).

### 3.2 | Beta testing

In the first round of beta testing, the data partners registered and entered their metadata. This experience presented a variety of important topics related to metadata definitions and standards; what information to collect; data provenance and stewardship; and overall workflow. In the second and final round, the data partners successfully completed a round trip through Discovery and Communication. This means that each successfully (1) discovered data the other did have and was willing to share out of network, (2) sent the other partner a data request, and (3) received a response to the request. Both partners received automatic notifications of each of these events. Importantly, data partners were not able to discover data that the other partner did not indicate it had or had indicated it did not choose to make visible outside its own network.

▼ **Types of Data Collected (Group)** ⋮

**Title\***

**Data Type**

▼ **Inpatient Encounters (Group)** ⋮

**Title\***

**Data Type**

New Child Domain

▼ **Child Domains**

▼ **Diagnosis Codes (Reference)** ⋮

**Title\***

**Data Type**

**References**  Supports Multi-Value Selection

⊕ Add new record

| Title  | Description | Value |   |
|--------|-------------|-------|---|
| ICD-9  |             |       | <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">🔄 Edit</span> <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">✕ Remove</span> |
| ICD-10 |             |       | <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">🔄 Edit</span> <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">✕ Remove</span> |
| SNOMED |             |       | <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">🔄 Edit</span> <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">✕ Remove</span> |

▼ **Procedure Codes (Reference)** ⋮

**Title\***

**Data Type**

**References**  Supports Multi-Value Selection

⊕ Add new record

| Title  | Description | Value |   |
|--------|-------------|-------|---|
| ICD-9  |             |       | <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">🔄 Edit</span> <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">✕ Remove</span> |
| ICD-10 |             |       | <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">🔄 Edit</span> <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">✕ Remove</span> |
| SNOMED |             |       | <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">🔄 Edit</span> <span style="border: 1px solid #ccc; padding: 2px 5px; border-radius: 3px;">✕ Remove</span> |

▼ **Demographics (Group)** ⋮

**Title\***

**Data Type**

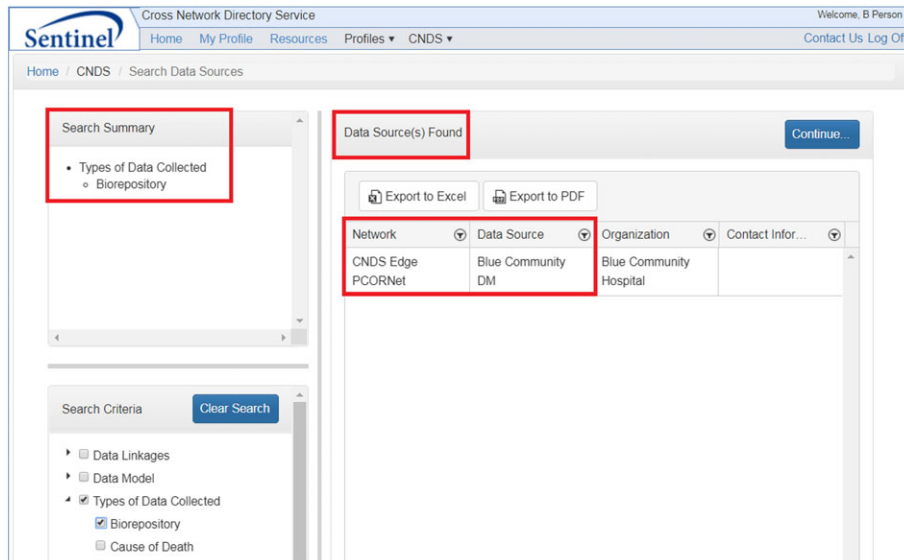
New Child Domain

**Child Domains**

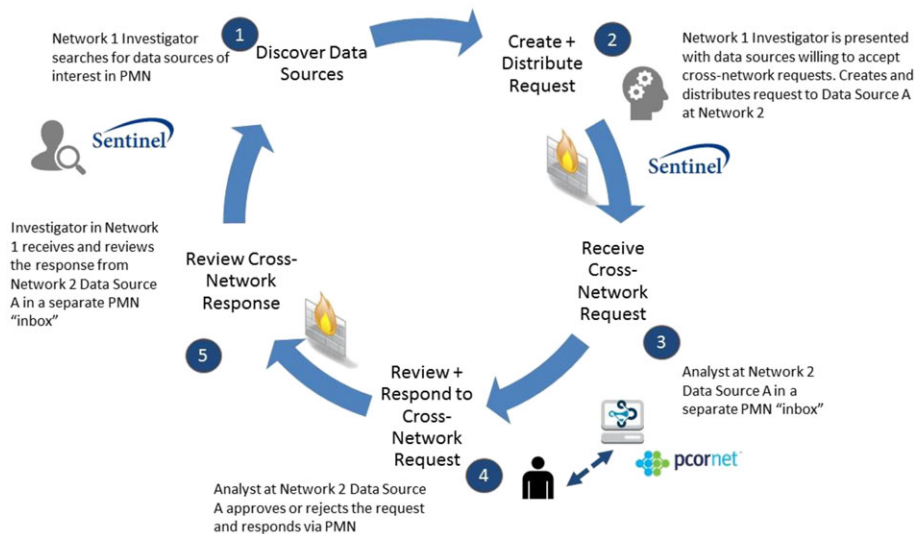
- ▶ Age (True | False)
⋮
- ▶ Sex (True | False)
⋮

📄 Save
✕ Cancel

**FIGURE 2** Metadata management. This figure illustrates the Cross-Network Directory Service (CNDS) metadata management function, showing how the underlying data model is populated via the user interface



**FIGURE 3** Cross-Network Directory Service (CNDS) and PopMedNet (PMN) integration architecture



**FIGURE 4** Cross network request cycle

### 3.3 | Validation testing

User acceptance testing was designed to verify key system functionality:

- Metadata management
  - Network participants can enter values for all metadata fields
  - Metadata fields can be added to the inventory without programming
  - Network participants can set visibility values for each metadata field independently
- Discovery
  - Network participants can search for organizations and data sources based on any combination of metadata fields
  - Discovery correctly returns organizations or data sources whose metadata meet the search criteria

Discovery correctly does not return organizations or data sources whose metadata do not meet the search criteria

Discovery correctly returns results only to participants who qualify based on visibility settings

- Communication
  - Data requests can be routed across networks
  - Notifications of request status are correctly sent to requestors

The pilot CNDS instances of PCORnet and Sentinel were iteratively tested and improved based on feedback and test results from the project team and workgroup. The user interfaces for capturing and exploring information about potential collaborators was validated by the teams. Testing to cover the end-to-end process of setting metadata visibility restrictions, searching, and then successfully querying across networks proved that CNDS can accommodate a wide range of use cases and provide the framework to support



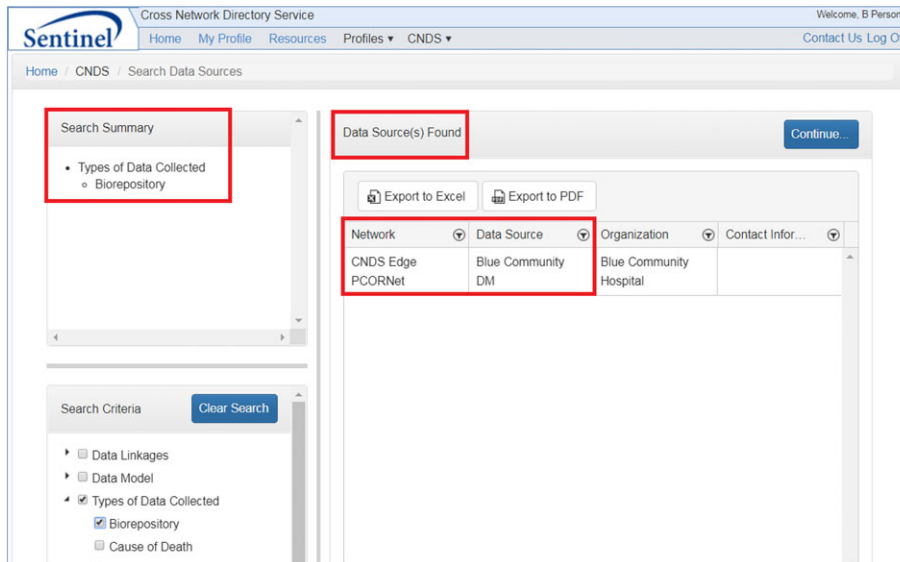


FIGURE 5 Discovery functionality

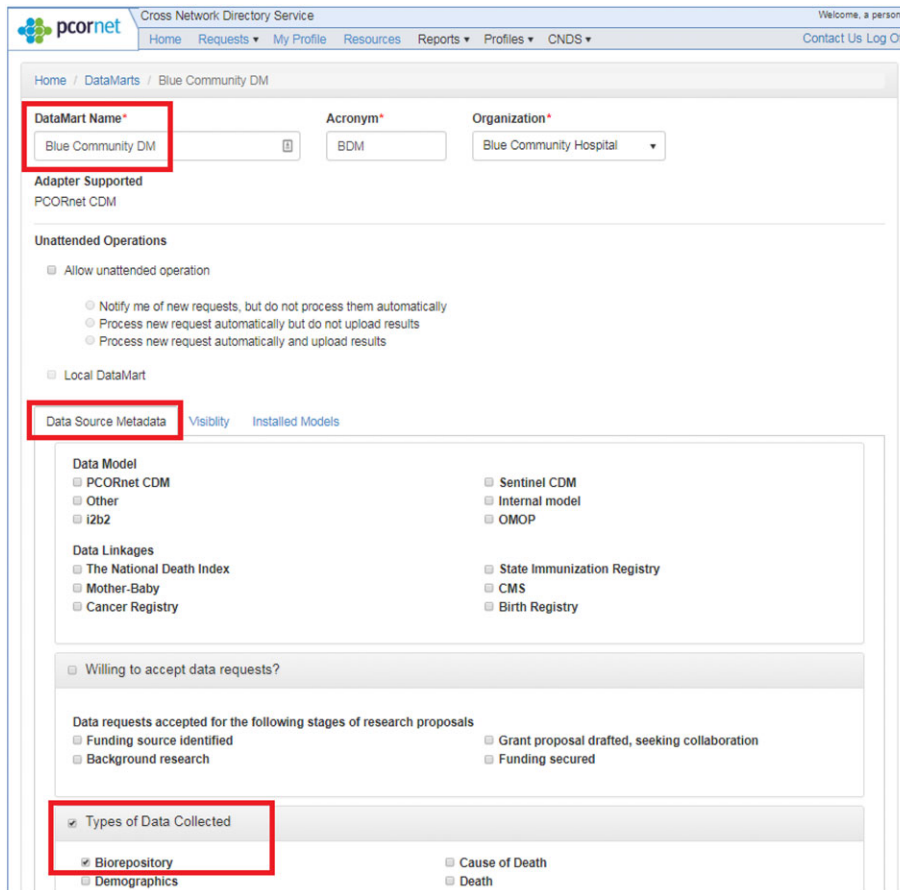


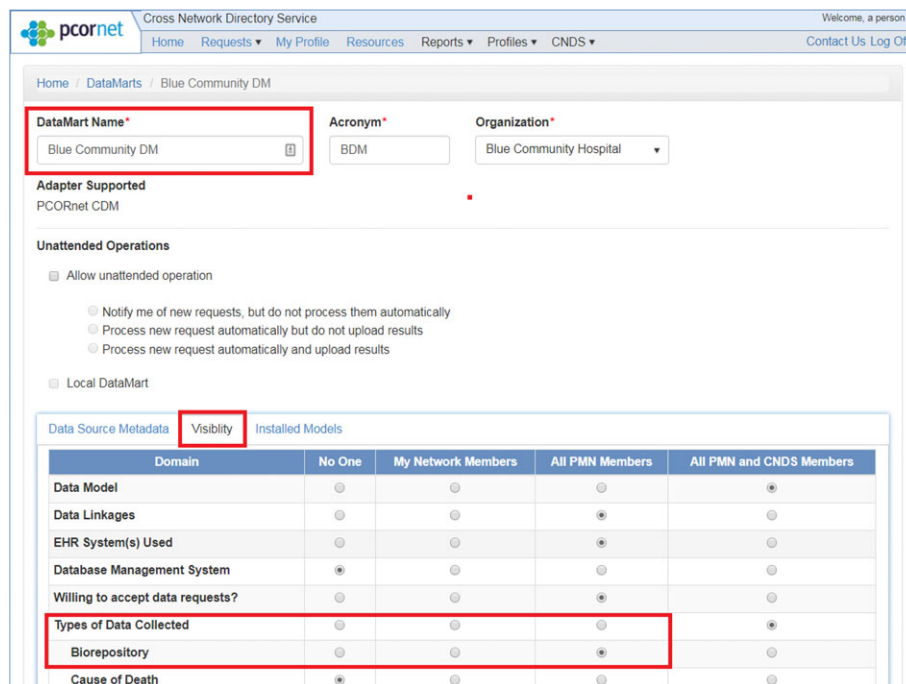
FIGURE 6 Metadata capture interface

via LHS. The system functions were successfully verified through user acceptance testing.

## 4 | DISCUSSION

In this section, we describe lessons learned through the CNDS project and how we might carry this learning through to other projects.

The CNDS project demonstrated the feasibility of enabling Discovery (search) and Communication (querying) across independent distributed networks. These capabilities were demonstrated on test instances of the Sentinel and PCORnet networks, and CNDS was implemented outside the main line of PMN software to avoid impacting Sentinel and PCORnet data partners not participating in the CNDS pilot.



**FIGURE 7** Visibility settings

Observations made during this project provided the teams with insights, ideas, limitations, and challenges that will drive and add value to future work, as described below.

- *Metadata provenance is critical.* While the CNDIS data model was flexible, it did not include effective updating information (eg, date of the update). Future work could include enhancing the CNDIS metadata model to capture provenance information about metadata elements to answer questions that enable users to determine fit-to-purpose characteristics, such as (1) Do the data in the system cover the data ranges of interest for the study, (2) For which data elements are common or standard coded data elements available (eg, LOINC codes), and (3) Are there active researchers in the domain of interest?
- *A formal approach to metadata data curation is needed* to sustain a system like CNDIS. While the value of identifying and defining metadata elements is important for a platform like CNDIS to evolve, this initial project aimed at standardizing the approach to capturing metadata. Sustainability is crucial for success.
- *Expand cross-network query functions and introduce terminology services.* The value of CNDIS can be expanded by integrating CNDIS into the PMN software code and creating a utility that simplifies migrating existing network metadata into the CNDIS metadata model. In addition, there are existing point-and-click query tools used by PCORnet, referred to as Menu-Driven Queries, that would be an added value for CNDIS. CNDIS was developed to securely send files across distinct networks, but if multiple models can be used to answer the same question, more complex request types that can be run directly against source data could be used. For example, if two data models both capture the same values for Race using the same structure in their respective Demographics tables, an SQL-based query that can be executed

directly against the database and return aggregate counts would be a valuable enhancement to CNDIS. Preliminary work on this approach has shown good results.<sup>22</sup>

- *Integration with other health research and collaboration platforms in the United States and abroad.* Future work that involves other collaborative health data initiatives (eg, Informatics for Integrating Biology & the Bedside [i2b2], Observational Health Data Sciences and Informatics [OHDSI], Electronic Health Data in a European Network [EHDEN], Canadian Network for Observational Drug Effect Studies [CNODES], and Collaborative Informatics Environment for Learning on Health Outcomes [CIELO]) will provide important benefits. Technology-wise, CNDIS was built with standard communication mechanisms (eg, APIs), enabling future integration possibilities.<sup>23-25</sup>
- *Health research community engagement.* Development of an open-source community for use of CNDIS through development of presentations, training materials, and improved implementation documentation would help to expand use of the tool to better leverage investments in distributed health data networks.

Of the five factors listed in the introduction that we see keeping distributed healthcare networks disconnected from each other and impeding collaboration, CNDIS directly addresses the first four. CNDIS helps break data network siloes by enabling networks and network participants to securely communicate with each other, discover resources across networks, and even query each other while adhering to appropriate governance.

CNDIS is a prototype because it has not yet been fully implemented in production. It is fully functional because we have demonstrated the ability of CNDIS to connect the Sentinel and PCORnet networks, both for mutual discovery of research capabilities and for making data requests of each other and tracking responses. Factor 5

(There is no reliable mechanism for sending queries across networks) was partially addressed by developing infrastructure to send data requests across distinct DDNs. Factor 4 (There are no operational standards or metrics for describing data at a level that enables researchers to judge fitness-for-use of others' data sources) is the subject of a separate project also funded by PCORTF through the U. S. Office of the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services (HHS) and through the FDA.

A flexible data model was developed to store the information entered. CNDS is not meant to re-create other professional networking platforms or registries, but to set a foundation upon which future integrations with such systems is possible via APIs.<sup>19-21,26</sup> The system was designed with the knowledge of related projects focused on professional collaboration efforts; projects such as ORCID, eagle-i, CIELO, and related LHS initiatives could potentially be integrated with CNDS via standard web services. The project team is exploring options to make CNDS a significant and sustainable part of LHS infrastructure. The teams envision CNDS being integrated with and leveraging such initiatives.

## 5 | SUMMARY

The CNDS project gathered functional requirements from stakeholders and collaborating partners to build a software application to enable cross-network data and resource sharing. The two partners—one from Sentinel and one from PCORnet—tested the software. They successfully entered metadata about their organizations and data sources. They were then able to use the Discovery and Communication functionality as both requesters and data sources. This means that each partner was able to discover only the information the other had designated they had and were willing to share out of network, send the other partner a data request, and receive a response to the request.

This pilot project aimed to leverage the HHS investments in health data networks by creating an open source tool that advances distributed analytics, data-sharing methods, and health research. The CNDS software can help integrate disparate health data networks by providing a mechanism for data partners to participate in multiple networks, share resources, and seamlessly send queries across those networks.

CNDS provides an elaborate yet easy-to-use system for sharing information across networks while maintaining local control over who can access it. Although the enabling software and data models are publicly available, fully realizing the value of CNDS, and the multiple health data networks in the United States and beyond, will require identifying use cases that demonstrate clear value for CNDS. Many collaborative opportunities exist to demonstrate value, for example, collaboration across Sentinel and PCORnet to further the goals of each network, or across networks in the United States, Canada, Asia, and Europe to further medical product safety surveillance. But realizing the value of CNDS to support these collaborations will require an investment in time and resources, coupled with a vision for how collaboration can benefit all parties.

## ACKNOWLEDGMENTS

The authors would like to thank the following people for their work on this project:

From Humana, Inc., Comprehensive Health Insights: Thomas Harkins, MA, MPH; Qianli Nair, MS. From the University of Michigan: Charles Friedman, PhD; James Estill, PhD; Lisa Ferguson, MSI; Maria Flores, MSW, From the Harvard Pilgrim Health Care Institute: Adam Paczuski, BS; From NTT Data Services: Matt McManus, BS.

Funding for this work was provided by the US Food and Drug Administration through the Department of Health and Human Services (HHS), as agents for the Patient-Centered Outcomes Research Trust Fund (Contract number HHSF223201400030I/HHSF22301006T).

## CONFLICT OF INTEREST

Jeff Brown, Jessica Malenfant, Jenny Hochstadt, Kimberly Barrett, Zachary Wyner, Chayim Herzig-Marx, and Bridget Nolan are employees of Harvard Medical School and Harvard Pilgrim Health Care Institute (HPHCI)-Department of Population Medicine. They have no conflicts of interest to declare.

Software development partners Daniel Dee and Dean Corriveau were employees of Avacoda LLC during the project period. They are now employees of General Dynamics Information Technology, Inc. (GDIT). The principal investigator for the project and subject of the article spoke on behalf of Dee and Corriveau that, as employees of GDIT, they had no conflicts of interest. Through contracts with HPHCI, GDIT supports development of PopMedNet and CNDS software platforms.

Marcelline Harris is employed by the University of Michigan; she has no conflict of interest to declare.

Vinit Nair is employed by Humana, Inc.; he has no conflict of interest to declare.

## ORCID

Jessica M. Malenfant  <https://orcid.org/0000-0001-8381-4496>

## REFERENCES

1. Institute of Medicine Roundtable on Evidence-Based M. The National Academies Collection: Reports funded by National Institutes of Health. In: Olsen LA, Aisner D, McGinnis JM, eds. *The Learning Healthcare System: Workshop Summary*. Washington (DC): National Academies Press (US) National Academy of Sciences; 2007.
2. Friedman C, Rubin J, Brown J, et al. Toward a science of learning systems: a research agenda for the high-functioning learning health system. *J. Am. Med. Inform. Assoc.: JAMIA*. 2015;22(1):43-50.
3. Friedman CP, Allee NJ, Delaney BC, et al. The science of learning health systems: foundations for a new journal. *Learn Health Syst*. 2017;1(1):e10020.
4. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48(6 Suppl):S45-S51.
5. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med*. 2009;151(5):341-344.
6. Toh S, Platt R, Steiner JF, Brown JS. Comparative-effectiveness research in distributed health data networks. *Clin Pharmacol Ther*. 2011;90(6):883-887.

7. McNeil MM, Gee J, Weintraub ES, et al. The Vaccine Safety Datalink: successes and challenges monitoring vaccine safety. *Vaccine*. 2014;32(42):5390-5398.
8. Steiner JF, Paolino AR, Thompson EE, Larson EB. Sustaining research networks: the twenty-year experience of the HMO Research Network. *EGEMS (Washington, DC)*. 2014;2(2):1067.
9. Platt R, Carnahan RM, Brown JS, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf*. 2012;21(Suppl 1):1-8.
10. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J. Am. Med. Inform. Assoc.: JAMIA*. 2014;21(4):578-582.
11. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff. (Project Hope)*. 2014;33(7):1178-1186.
12. Platt R, Lieu T. Data enclaves for sharing information derived from clinical and administrative data. *JAMA*. 2018;320(8):753-754.
13. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574-578.
14. PopMedNet [Website]. 2018 [cited 2018. Available from: <https://www.popmednet.org>.
15. Davies M, Erickson K, Wyner Z, Malenfant J, Rosen R, Brown J. Software-enabled distributed network governance: the PopMedNet experience. *EGEMS (Washington, DC)*. 2016;4(2):1213.
16. PopMedNet Request Types [Website]. 2018 [cited 2018. Available from: <https://popmednet.atlassian.net/wiki/spaces/DOC/pages/8880286/Request+Types>.
17. Her QL, Malenfant JM, Malek S, et al. A query workflow design to perform automatable distributed regression analysis in large distributed data networks. *eGEMS*. 2018;6(1):11.
18. PCORnet Data Network Request: PCORI; 2018 [updated 03-06-2018; cited 2018 08-20-2018]. Available from: <https://pcornet.org/data-network-request/>.
19. What is Eagle-i? : The Harvard Clinical and Translational Science Center; 2018 [cited 2018 08-01-2018]. Available from: <https://www.eagle-i.net/about>.
20. Professional Networking and Expertise Mining for Research Collaboration: The Harvard Clinical and Translational Science Center; [cited 2018 08-01-2018]. Available from: <http://profiles.catalyst.harvard.edu/>.
21. Payne P, Lele O, Johnson B, Holve E. Enabling Open Science for Health Research: Collaborative Informatics Environment for Learning on Health Outcomes (CIELO). *J Med Internet Res*. 2017;19(7):e276.
22. Klann JG, Buck MD, Brown J, et al. Query health: standards-based, cross-platform population health surveillance. *J. Am. Med. Inform. Assoc.: JAMIA*. 2014;21(4):650-656.
23. Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. *J. Am. Med. Inform. Assoc.: JAMIA*. 2016;23(5):909-915.
24. European Health Data & Evidence Network 2018 [cited 2018 12/20/2018]. Available from: <http://www.ehden.eu/>.
25. Suissa S, Henry D, Caetano P, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med: A Peer-Reviewed, Independent, OAJ*. 2012;6(4):e134-e140.
26. ORCID: Connecting Research and Researchers [website]. ORCID; 2018 [cited 2018 08-20-2018]. Available from: <https://orcid.org/>.

**How to cite this article:** Malenfant JM, Hochstadt J, Nolan B, et al. Cross-Network Directory Service: Infrastructure to enable collaborations across distributed research networks. *Learn Health Sys*. 2019;e10187. <https://doi.org/10.1002/lrh2.10187>