CROSS-NETWORK DIRECTORY SERVICE PROJECT

FINAL REPORT

Prepared by: U.S. Food and Drug Administration, Sentinel Operations Center and Cross-Network Directory Service Collaborating Partners

September 14, 2018

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I. This project was funded by the FDA through HHS Mini-Sentinel contract number HHSF223200910006I. This work was supported by the Office of the Secretary PCORTF under Interagency Agreement #750115PE060020.

Cross-Network Directory Service Project

Final Report

Table of Contents

I. OVERVIEW AND OBJECTIVES

- ***Multiple distributed research networks exist***
- ***Integrating these networks can help create a true Learning Health System***

The growing adoption of distributed networks within the Department of Health and Human Services (HHS) to facilitate large-scale comparative safety and effectiveness studies, as well as other public health surveillance activities, creates an opportunity to leverage those investments to create a national resource that enables a true Learning Health System. U.S. Food and Drug Administration (FDA), Patient-Centered Outcomes Research Institute (PCORI), National Institutes of Health (NIH), Centers for Disease Control and Prevention (CDC) and others are all supporting various forms of distributed health data networks. Together, these networking infrastructure investments can be integrated to support needs across HHS agencies.

This project creates an open source interoperable service that allows: 1) data partners to easily participate in multiple data research networks, 2) queries to seamlessly move across such networks, and 3) users to share analytic capabilities and knowledge across networks. This project pilot tested the Cross-Network Directory Service (CNDS) across two existing networks: FDA's Sentinel and PCORI's National Patient-Centered Clinical Research Network (PCORnet).

The objectives of the project were to:

- Design: Identify the key functionalities and create a technical design for a CNDS
- Implement and Test: Develop and test a detailed design for the CNDS with at least two existing distributed research networks
- Document and Release: Produce user and technical materials.

The project was implemented in three phases: 1) Discovery and Design, 2) Implementation, and 3) Testing and Release. During the Discovery and Design phase, requirements gathering took place that resulted in identification of business and technical requirements and use cases. Based on this work a technical design was created and approved.

During the Implementation phase, software was developed and tested to demonstrate functionality.

During the Testing and Release phase identified issues were addressed and compatibility and performance testing took place along with a Failure Mode and Effects Analysis (FMEA). Technical documentation and a user manual were completed.

Table 1. Schedule of IAA Deliverables to ASPE

Phase	Objectives	Deliverables
Phase 1: Discovery & Design (6 months)	• Identify the key functionalities and an overarching technical design for a CNDS	• Use cases; a brief report of the overarching design, business and technical requirements to be addressed, and description of key functionalities
Phase 2: Implementation (12 months)	• Develop and test (with at least two existing networks) a detailed design for the CNDS	• Detailed design and technical documentation • Software development conducted and testing plans developed and implemented • A report describing results of testing, demonstrating functionality to key stakeholders, and issues to be addressed
Phase 3: Test & Release (6 months)	• Release CNDS • Conduct additional analyses of the robustness of the CNDS and produce user materials	• Software development completed, implementing the CNDS functionality for at least 2 established networks, FDA's Mini-Sentinel (or its successor Sentinel System) and PCORnet • Brief report on the Failure Mode and Effects Analysis (FMEA) and implementation process for extending the CNDS to other networks • Updated documentation & user materials
		• Copy/copies to ASPE of project contract/s • Quarterly Reports to ASPE • Final Report to ASPE

II. WHAT IS CNDS

The PopMedNet™ (PMN) software application supports distributed within-network querying for Sentinel, PCORnet, MDPHnet, Health Care Systems Research Network (HCSRN), the HCSRN Cancer Research Network, the Biologics and Biosimilars Collective Intelligence Consortium (BBCIC), the Reagan-Udall Foundation's Innovation in Medical Evidence Development and Surveillance (IMEDS), and the NIH Health Care Systems Research Collaboratory. It enables creation, operation, and governance of distributed health data networks and distribution of data queries within these networks.

- ***Discover and query data sources***
- ***Respect local governance rules***
- ***Easily integrate new data partners***

The Cross-Network Directory Service (CNDS) extends PMN's functionality to enable cross-network discovery of potential collaborators and data sources and querying of those sources while enforcing local governance rules. CNDS is built within the PMN infrastructure and leverages PMN functionality. CNDS is implemented as a set of services that can be invoked by PMN instances. In slightly more technical terms, CNDS provides a standard set of functions that PMN can call upon through application programming interfaces (APIs). This design limits the need for software upgrades to networks wishing to take advantage of CNDS capabilities.

Through its APIs, CNDS offers functionality to:

- Allow users to register in CNDS
- Capture metadata describing users, organizations, registries, research data sets, and queryable data sources
- Enable users to search organization metadata (to identify potential collaborators) and data source metadata (to explore characteristics of electronic healthcare data sources) across networks
- Route requests and responses across networks

CNDS is powered by metadata—standardized data elements about organizations and data sources. It provides storage and retrieval of metadata about organizations and data sources. Visibility metadata are used to determine what organization and data source metadata can be seen by whom. The data model for storing metadata is designed for flexibility—it enables changes to metadata elements without software redesign or programming. CNDS metadata management functionality allows system administrators to quickly and easily add, delete, or modify metadata elements.

- ***Flexible data model***
- ***Metadata changes require no programming***
- ***Participants choose what information to reveal***

A. KEY FUNCTIONAL REQUIREMENTS

The components of CNDS are—Registration, Discovery, Communication, and Governance as described below.

Registration: Enables users to request an account and enter metadata about themselves, their organizations and their data resources, and determine what metadata others can see.

Discovery: Enables users to search the metadata, entered as part of registration, to find new data sources and potential organizations with which to collaborate.

Like the CNDS data model, Discovery is designed flexibly so that the application does not require re-programming as the metadata change. That is, the list of elements that can be searched is automatically generated from the metadata stored in the database. The result set returned from a search is constrained by the visibility level set by the metadata owner.

Communication: Enables users to send and receive queries across networks. PMN provides functionality for creating, distributing, and responding to a variety of request types. It sends related email notifications within a single PMN network. CNDS extends these capabilities across networks by mapping common request types used by multiple networks. Using CNDS, users can send and receive requests, regardless of network affiliation, according to the governance rules of the recipients.

Governance: Enables granular governance via visibility settings and access controls.

Visibility rules are entered in metadata (via the Registration function). These rules are enforced when users search for organizations or data sources (via the Discovery function). Visibility rules identify who is authorized to see each organization and data source metadata element. Users can tag metadata elements as being visible to:

- No one (i.e., just myself and the system administrators)
- Registrants in my network
- Registrants in any PMN network
- All CNDS Registrants

Because CNDS integrates with PMN, PMN's extensive user access control system (i.e., permission system) is also available to CNDS. These permissions allow for robust governance by controlling every aspect of use of the application, for example, who can: add, edit, delete, and view users, organizations, and DataMarts; respond to, reject, and upload results; manage security; and run audit reports. CNDS adds access controls to govern actions such as who can manage metadata or send a cross-network request.

III. PROBLEMS ADDRESSED AND IMPACT OF SOLUTIONS

The growing adoption of distributed networks within HHS to facilitate large-scale evidence generation studies, as well as other public health surveillance activities, has created a variety of research networks. FDA, PCORI, NIH, ONC, CDC, AHRQ, and others are supporting various forms of distributed health data networks. These networks are independent even as they address related questions of healthcare research, and public policy.
Five factors keep these networks disconnected from each other and preclude their integration:

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 common data models.

CNDS addresses factors 1-3 through its Registration, Discovery, Communication, and Governance capabilities.

- Registration enables an organization to identify itself across all participating networks and to describe its research capabilities.
- Discovery enables organizations to search across networks for organizations with particular capabilities and expertise or for data sources with specific types of information.
- Communication enables organizations to send data requests to other organizations both within their own network and outside of it and to track response activity consistently.
- Governance enables organizations to decide which of their research capabilities they wish to expose and with whom to share that information. Choices currently are not to share, to share only within an organization's own network, to share with any PopMedNet network, or to share with anyone registered in CNDS.

The ASPE-funded project "Data Standards and Metrics" addresses factor 4. The section of this report titled Proposal for Future Work addresses factor 5.
Benefits of CNDS for improving the Patient Centered Outcomes Research (PCOR) research infrastructure can be substantial:

- Currently, the only way to identify potential collaborators is to already know people or to know someone who knows the right people. CNDS assists both sides of a collaboration. Organizations with research skills can advertise through CNDS, and investigators seeking specific skills can find them.
- Despite considerable overlap between their common data models (CDMs), PCORnet and Sentinel data partners cannot send requests to each other. CNDS delivers a framework that enables cross-network communication. The next step of enabling a query to span different CDMs is discussed in the section titled Proposal for Future Work.
- PCORnet offers a large and diverse inventory of EHR-based clinical information, and Sentinel does the same for claims-based information. There are hundreds if not thousands of important data collections and registries, and for some research interests these other data sources are of far greater significance than Sentinel or PCORnet. CNDS can be extended to encompass any or all such data sources through simple modifications of the configuration of the CNDS software (i.e., no additional software programming required).

IV. ACCOMPLISHMENTS

A. GOAL

- ***Bridge disparate research networks***
- ***Share research expertise***
- ***Leverage diverse data sources (e.g. claims, EHR, patient-reported, registries)***

The goal of this project was to create an open source interoperable service that enables: 1) data partners to easily participate in multiple data research networks, 2) queries to seamlessly move across such networks, and 3) users to share analytic capabilities and knowledge across networks. In achieving this goal, CNDS provides bridges among distributed research networks that were developed at separate times, with differing rules for participation and governance, and were deployed to separate research audiences that are not fully aware of others' research capabilities. Success required achieving three objectives: 1) identify the key functionalities and produce an overarching technical design for a CNDS, 2) develop and test a detailed design for the CNDS with at least two existing distributed research networks, and 3) conduct additional analyses of the robustness of the CNDS and produce user and technical materials.

B. OBJECTIVES

1. Phase 1: Identify Key Functionalities and Technical Design for a CNDS

Requirements gathering took place resulting in identification of business and technical requirements and use cases. Initial input was gathered during two stakeholder meetings.[1] The Workgroup (composed of representatives from Harvard Pilgrim Health Care Institute, FDA, a Sentinel data partner – Humana, a PCORnet data partner – University of Michigan, and the software developer) met regularly (9 times) between August 18, 2016 and October 10, 2017 to contribute ideas, view the project as it developed, give feedback, and test the software.

A data modeler and software developers were engaged to create software based on this input. CNDS was built within the PopMedNet™ (PMN) infrastructure to take advantage of its already mature functionality for securely distributing data requests within a distributed health data network. The architecture underlying CNDS makes it easy to add new partners. Given that the metadata inventory for CNDS will change and grow over time, the data model was designed and the software built to accommodate change without the need for additional programming. Whenever it is determined that there needs to be a change in the metadata collected, an administrative user with minimal training can quickly make changes to the underlying metadata tables, and the data entry screens are designed to automatically display what is in the underlying tables.

Table 2. Deliverables for this Objective

Deliverable	Description	Audience and Use	Location
Use cases	A brief report of the overarching design, business and technical requirements, and description of key functionalities	For the more technical user who wishes to understand the background of how CNDS was built	This was the first draft of the Phase 3 deliverable "Detailed design and technical documentation" which is available on the Sentinel website

2. Phase 2: Develop and Test Detailed Design with Existing Networks

As the software was being developed, new functionality was regularly presented at Workgroup meetings. Data partners had the opportunity to try out the software and provide feedback. For beta testing, the Sentinel data partner, Humana, and the PCORnet data partner, University of Michigan, accessed CNDS through web-based portals created for them.

In the first round of testing, the data partners successfully registered and entered their metadata, including visibility settings governing who could see their metadata. The metadata values and visibility settings they entered were scripted so that a comprehensive test of the Discovery and Governance functions could be performed at a later stage.

[1] Stakeholder meetings were attended by: Aetna, America's Health Insurance Plans, Bohn Epidemiology, Cincinnati Children's Hospital Medical Center, Duke University, FDA, Harvard Pilgrim Health Care Institute, Health Care Systems Research Network, HealthCore, Hospital Corporation of America Health Care, Humana, IQVIA (IMS Health and Quintiles), Kaiser Permanente Center for Health Research, Kaiser Permanente Colorado, Kaiser Permanente Mid Atlantic, PCORI, RTI International, University of Iowa, and University of Michigan.

Delivered software application with:

- ***Flexible data model***
- ***Components for Registration, Discovery, Communication, and Governance***

When software development was complete, data partners proceeded with the second and final round of testing, which was designed to simulate actual use of CNDS. Each partner performed several metadata searches looking for data sources with specific characteristics (that is, data sources with particular values of certain metadata elements). As noted in the preceding paragraph, the metadata values and visibility settings entered were designed so that the correct outcomes were known in advance. For example, in some cases a data source had the desired characteristic, but the visibility setting entered should prevent it from being discovered if the software was working properly.

Each data partner successfully tested the full cycle of normal CNDS usage from Registration to Discovery to Communication—while adhering to the other's Governance rules. The data partners: discovered data the other did have and was willing to share out of network, sent the other partner a data request, and received a response to the request. Both partners received automatic notification of each of these events. Further, 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.

Table 3. Deliverables for this Objective

Deliverable	Description	Audience and Use	Location
Detailed design and technical documentation - including a report describing results of testing and issues to be addressed*	A brief report of the overarching design, business and technical requirements, description of key functionalities, and issues to be addressed as a result of testing and feedback from stakeholders	For the more technical user who wishes to understand the history of how CNDS was built.	This was the second draft of the Phase 3 deliverable "Detailed design and technical documentation" which is available on the Sentinel website

*Note that the technical documentation and testing report appear as separate deliverables in the Schedule of IAA Deliverables, but were delivered as one document.

3. Phase 3: Release CNDS and Produce User and Technical Materials

Delivered:
- ***User guide***
- ***Design and technical documentation***

A detailed user guide, technical guide, and final report are publicly available. The user guide was tested by the data partners, that is they used it to carry out software testing.

Table 4. Deliverables for this Objective

Deliverable	Description	Audience and Use	Location
Source code and documentation	The source code for completed CNDS software and a readme file, describing how to implement it, is available as open-source code.	For those wishing to implement and use the CNDS software themselves. Users should be Git-literate and technical in general.	Available through GitHub
User materials*	A step-by-step guide for using the CNDS software with screenshots and examples. It shows how to enter metadata, search for data sources and potential collaborators, govern access to metadata, and send and receive requests.	For users to learn how to use the CNDS software	Available on the Sentinel website
Detailed design and technical documentation*	Detailed technical documentation of the CNDS software with an overview of the architecture and description of how the Governance component works with the Registration, Discovery, and Communication components to enable data source owners to govern what they share and with whom.	For the more technical user who wishes to understand how CNDS is built and for those who need to administer CNDS	Available on the Sentinel website
Final Report	A description of the goals and accomplishments of the CNDS project with sections on lessons learned and recommendations for future work.	For those interested in an overview of the creation of CNDS.	Available on the Sentinel website

*Note that these appear as a single deliverable in the Schedule of IAA Deliverables, but were delivered as two separate documents.

C. DISSEMINATION

In addition to the contracted deliverables for this project, the following activities were conducted to disseminate the work.

1. Conferences

- Poster presentation:
  AMIA 2018 Informatics Summit
  San Francisco, March 12-15, 2018
- Oral presentation:
  Health Care Systems Research Network (HCSRN) Conference
  Minneapolis, April 11-13, 2018

2. Paper in Progress

- Learning Health Systems – Technical Report

V. LESSONS LEARNED

In this section, we describe lessons learned through the CNDS project and how we might carry this learning through to other projects. Because the concept of metadata lies at the core of CNDS, it is not surprising that most lessons learned related to metadata.

A. DATA VS. METADATA

In our discussions with workgroup members, especially concerning which metadata we should capture about data sources and organizations, we found that the distinction between data and metadata was not always clear, and in fact people held different views on the subject. For example, is an organization's address data or metadata? From an implementation perspective the answer is unimportant, as our model accommodates any type of information we wish to capture about an organization. From a methodological perspective, the answer may be important, as it relates to the internal consistency of the metadata model.

In future work, we will likely draw the distinction between data held in a data source and descriptors of those data. For example, CNDS is currently designed to record the fact that a particular data source collects death data, but not that those data are missing 60% of the time or come from death certificate records.

B. METADATA MODEL VS. TAXONOMY

A taxonomy is a consistent, preferably exhaustive, framework for classifying information or knowledge. The CNDS project did not require creating a taxonomy of metadata about data sources and organizations. Some workgroup participants thought that the project would have benefited from an explicit taxonomy. From the perspective of this project the discussion of taxonomy was out of scope. The database built on the CNDS metadata model works equally well whether its structure follows a taxonomy or not. In fact, it would work equally well if collecting information about countries (instead of organizations) and their vacation opportunities (instead of data sources). Developing a stable taxonomy and a robust metadata dictionary is future work.

C. COMPLEXITY OF MAPPING BETWEEN NETWORKS

We expected the CNDS project to be challenging, but until we delved into the technical specifications we did not appreciate the complexity of mapping requests between networks. PopMedNet provides great flexibility to structure networks, organizations, and projects, and that flexibility is reflected in the amount of configuration needed to successfully transport data requests between networks. This level of configuration, in turn, is the primary driver behind our future work recommendation to establish a small CNDS operating center across affiliated networks, with a key responsibility for operating inter-network request mappings.

VI. PROPOSAL FOR FUTURE WORK

The CNDS project has demonstrated the feasibility of enabling Discovery (search) and Communication (querying) across independent distributed research networks. These capabilities were demonstrated on the Sentinel and PCORnet networks. CNDS was implemented outside the main line of PMN software to avoid requiring all Sentinel and PCORnet participating data partners to revise their configurations. Consequently, test instances of Sentinel and PCORnet were used in the project.

In this section, we describe follow-on work of two types. The first type is work necessary to realize the full value of the CNDS as currently implemented. The second type is work that will significantly expand CNDS capabilities.

A. REALIZING FULL CNDS VALUE

We recommend the following activities:

- Convene a workgroup, including representatives from both Sentinel and PCORnet, charged with identifying a basic set of metadata about data sources and organizational capabilities that all Sentinel and PCORnet data partners will be expected to maintain. This activity is necessary so that investigators will have the ability to search for collaborators in a meaningful way across the networks and to enable cross-network querying and collaboration.
- Undertake software development to integrate CNDS into the main line of the PopMedNet software code and to create a utility that simplifies migrating existing Sentinel and PCORnet metadata into the CNDS metadata model.
- CNDS itself remains a separately implemented set of web-based services. Verify that CNDS services can be invoked from applications other than PopMedNet, e.g. an Informatics for Integrating Biology & the Bedside (i2b2) instance.
- CNDS requires creating and maintaining mappings of governance rules and other configurations so that networks can communicate with each other. Create a small CNDS operating center to fulfill this maintenance role.
- Outreach to the Observational Health Data Sciences and Informatics (OHDSI) and i2b2 communities to enlist their participation in CNDS. With current CNDS capabilities, that participation will be limited to enabling Discovery. Enabling Communication will be discussed below under “Extending CNDS Capability”.
- Create and support an open source community for use of CNDS through development of presentations, training materials, and improved implementation documentation.

B. EXTENDING CNDS CAPABILITY

During this project, we have identified several avenues for extending and enhancing CNDS capabilities. Especially important are opportunities to improve the metadata model that lies at the heart of the Discovery function, the capabilities of Discovery itself, and the ability to Communicate with data partners other than those in a PopMedNet network.

1. Improve the CNDS Metadata Model

- Expand the metadata model to capture date-related information about metadata elements. Specifically, enable an "effective dating" capability, whereby one can track when values of metadata elements change. For example, a data partner establishes a new research practice, say in pediatrics. Prior to that date, the partner's metadata value for "Knowledge of pediatric research" would have been "No," but after that date the value would be "Yes." Investigators looking for collaborators in pediatric research will want to know how long another organization has had such skills in place.
- Enable importing lists of values for reference sets. Currently, if the metadata model wished to provide a list of U.S. state codes as a reference set for addresses, or a list of ICD-10 diagnosis codes, those code sets would have to be entered manually, which is feasible but clumsy.

2. Improve Discovery and Communication

- Several opportunities exist for improving search functionality:
    - Enable saving and re-using search criteria.
    - Support complex Boolean logic.
    - Add visualizations to better show which criteria have the greatest effect on search results. For example, the universe of data sources is 104 distinct sources, criterion 1 eliminates 51 of those sources, criterion 2 eliminates an additional 29 sources, etc. Use visualization techniques to compare different sets of criteria.
- Enable CNDS to distribute Menu-Driven Queries (MDQs) to data sources that are held in relational databases.
    - Enable Sentinel data partners to send MDQs to PCORnet data sources.
    - Add "model adapters" to PopMedNet for OMOP and i2b2 data sources.
    - Develop a software utility that enables creating a PMN model adapter for any data source housed in a relational database. This would be an extremely valuable addition to CNDS, albeit a challenging technical task.

VII. SUMMARY

The CNDS project gathered 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 functionalities 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.

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.

VIII. ACKNOWLEDGEMENTS

The Sentinel Operations Center would like to thank the CNDS Workgroup for their work on this project:

Collaborating Partner	Participant
Sentinel Operating Center	• Jeffrey Brown, PhD Principal Investigator and Workgroup Co-lead • Chayim Herzig-Marx, PhD Workgroup Co-lead • Jenny Hochstadt, MS Technical Project Manager • Kim Barrett, MPH • Jessica Malenfant, MPH • Bridget Nolan, BS • Zac Wyner, MPH
Humana, Inc., Comprehensive Health Insights, Miramar, FL	• Vinit P. Nair, BPharm, MS, RPh Principal Investigator • Thomas Harkins, MA, MPH • Qianli Nair, MS
University of Michigan	• Charles Friedman, PhD Principal Investigator • James Estill, PhD • Lisa Ferguson, MSI • Maria Flores, MSW • Marcelline Harris, PhD, RN, MS
Developer	• Daniel Dee, MS