3-30-2016

Software-Enabled Distributed Network Governance: The PopMedNet Experience

Melanie Davies
Harvard Pilgrim Health Care Institute, melaniedavies51@gmail.com

Kyle Erickson
Harvard Pilgrim Health Care Institute, kyle_erickson@harvardpilgrim.org

Zachary Wyner
Harvard Pilgrim Health Care Institute, zachary_wyner@harvardpilgrim.org

Jessica M. Malenfant
Harvard Pilgrim Health Care Institute, jessica_malenfant@harvardpilgrim.org

See next pages for additional authors

Recommended Citation

Davies, Melanie; Erickson, Kyle; Wyner, Zachary; Malenfant, Jessica M.; Rosen, Rob; and Brown, Jeff (2016) "Software-Enabled Distributed Network Governance: The PopMedNet Experience," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 4: Iss. 2, Article 5.
DOI: http://dx.doi.org/10.13063/2327-9214.1213
Available at: http://repository.edm-forum.org/egems/vol4/iss2/5

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Software-Enabled Distributed Network Governance: The PopMedNet Experience

Abstract

Introduction: The expanded availability of electronic health information has led to increased interest in distributed health data research networks.

Distributed Research Network Model: The distributed research network model leaves data with and under the control of the data holder. Data holders, network coordinating centers, and researchers have distinct needs and challenges within this model.

Software Enabled Governance: PopMedNet: The concerns of network stakeholders are addressed in the design and governance models of the PopMedNet software platform. PopMedNet features include distributed querying, customizable workflows, and auditing and search capabilities. Its flexible role-based access control system enables the enforcement of varying governance policies.

Selected Case Studies: Four case studies describe how PopMedNet is used to enforce network governance models.

Issues and Challenges: Trust is an essential component of a distributed research network and must be built before data partners may be willing to participate further. The complexity of the PopMedNet system must be managed as networks grow and new data, analytic methods, and querying approaches are developed.

Conclusions: The PopMedNet software platform supports a variety of network structures, governance models, and research activities through customizable features designed to meet the needs of network stakeholders.

Acknowledgements

The work described was supported by the following grants and contracts: Efforts to Develop the Sentinel Initiative, FDA (HHSF223200910006I); PCORI National Patient-Centered Clinical Research Network Coordinating Center, PCORI (PCO-COORDCTR2013); Health Care Systems Research Collaboratory - Coordinating Center, NIH (1U54AT007748-01)

Keywords

Governance, Informatics, Platforms, Research Networks, Research Operations, Health Information Technology, Data Ownership

Disciplines

Medicine and Health Sciences

Authors
Melanie Davies, *Harvard Pilgrim Health Care Institute*; Kyle Erickson, *Harvard Pilgrim Health Care Institute*; Zachary Wyner, *Harvard Pilgrim Health Care Institute*; Jessica M Malenfant, *Harvard Pilgrim Health Care Institute*; Rob Rosen, *Lincoln Peak Partners*; Jeff Brown, *Harvard Pilgrim Health Care Institute*.

eGEMs
Generating Evidence & Methods
to improve patient outcomes

Software-Enabled Distributed Network Governance: The PopMedNet Experience

Melanie Davies;[i] Kyle Erickson;[i] Zachary Wyner, MPH;[i] Jessica Malenfant, MPH;[i] Rob Rosen, MBA;[ii] Jeffrey Brown, PhD[i]

ABSTRACT

Introduction: The expanded availability of electronic health information has led to increased interest in distributed health data research networks.

Distributed Research Network Model: The distributed research network model leaves data with and under the control of the data holder. Data holders, network coordinating centers, and researchers have distinct needs and challenges within this model.

Software Enabled Governance: PopMedNet: The concerns of network stakeholders are addressed in the design and governance models of the PopMedNet software platform. PopMedNet features include distributed querying, customizable workflows, and auditing and search capabilities. Its flexible role-based access control system enables the enforcement of varying governance policies.

Selected Case Studies: Four case studies describe how PopMedNet is used to enforce network governance models.

Issues and Challenges: Trust is an essential component of a distributed research network and must be built before data partners may be willing to participate further. The complexity of the PopMedNet system must be managed as networks grow and new data, analytic methods, and querying approaches are developed.

Conclusions: The PopMedNet software platform supports a variety of network structures, governance models, and research activities through customizable features designed to meet the needs of network stakeholders.

[i]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, [ii]Lincoln Peak Partners

Introduction

As health researchers, academic research institutions, regulatory agencies, and others seek to leverage the expanded availability of electronic health information, the viability of, and interest in, health data research networks has grown. Health data research networks are critical to achieve a successful learning health system and hold great promise for many research (e.g., comparative effectiveness research) and public health initiatives aiming to improve the health of patients and populations.[1-4] In the United States, these initiatives include President Obama's Precision Medicine Initiative, the National Institutes of Health's Big Data to Knowledge (BD2K) initiative, and initiatives led by the Patient-Centered Outcomes Research Institute.[5-8] The Innovative Medicines Initiative, specifically the Electronic Health Records Systems for Clinical Research (EHR4CR) project and the European Union's TRANSFoRm project, are examples of European efforts to leverage electronic health data for similar purposes.[9-11] The learning health system is already being realized in a number of ongoing health data research networks, including those referenced in this paper.[12-14]

Health data networks are needed when no single data source can address the intended research needs. However, multi-institutional health research collaborations create many issues related to patient privacy; regulatory compliance; data security; and safeguarding proprietary, competitive, or otherwise sensitive information. Addressing these critical concerns requires comprehensive governance approaches.[15-18]

This paper describes critical data partner, coordinating center, and researcher needs as they relate to governance and operations with distributed health data networks, and how the PopMedNet software platform helps address those operational and governance matters. The paper focuses on specific implementation approaches and challenges as they relate to distributed network implementation and use using PopMedNet; others have described general governance issues related to distributed networks.[15,19-21] We use several examples of comparative effectiveness research, medical product safety surveillance, and public health monitoring to illustrate the applicability of the architecture to health data research networks across a range of network sizes, governance models, data sources, funding models, and research areas.[8,13,14,22,23] Additionally, we identify the opportunities and barriers to implementing PopMedNet for distributed research. Detailed information on implementing PopMedNet may be found at https://popmednet.org.

Distributed Research Network Model

In a distributed research network, data are held and managed by the institution that collected the data or that is otherwise responsible for its management. Data remain behind the institution's firewall, ensuring the security of protected health information; the data holder maintains complete operational control over the data and all uses. In the distributed model, research questions are answered by sending a "query" (request) to the data partners for local execution. Query results are returned for final analysis.[17,18,24-30] This distributed model raises a series of important considerations for stakeholders such as data partners, network coordinating centers, and researchers. Table 1 summarizes the concerns described below.

Typical multi-institutional research projects require sharing of individual-level data with researchers to create a single analytic data set. This approach raises many concerns that could make it difficult for data partners to participate in research. Specifically, sharing person-level data across institutions risks the accidental release of protected health information,

Table 1. Summary of Considerations

PARTY	CONCERNS
Data Partner	• Protection of PHI • Loss of control over proprietary information • Vulnerability to malicious programs
Coordinating Center	• Consistency of information across data partners • Quality of data • Flexibility to handle different network structures with overlapping data partners • Ability to track and report operational activities • Managing multiple projects
Researcher	• Lack of direct access to data • Differences among networks' data models • Possible limitations on number of data requests allowed

eliminates the ability to control and monitor data uses (and reuses), and raises potential proprietary and competitive concerns. Data security is also a serious concern, both during data transit and security of the final analytic data set.

A successful distributed research network requires coordination to ensure network efficiency and functionality. Typically, networks rely upon a coordinating center to manage day-to-day operations, including administrative and project management tasks, maintenance of the technical infrastructure, research operations, activity tracking and reporting, and capture and dissemination of lessons learned through a robust knowledge management system. Another coordinating center obligation is to provide network users with up-to-date information about network data availability and quality. Coordinating centers also need the ability to configure networks' flexibly to enable various architectures and governance models to meet the needs of data partners and researchers. Finally, coordinating centers are typically responsible for ensuring the security of the network. The Federal Information Security Management Act of 2002 (FISMA) lays out a framework for managing information security and charges the National Institute of Standards and Technology with defining the standard and guidelines for compliance. FISMA requirements must be met by any contractor working on behalf of a federal agency. Therefore, any distributed research network supported by federal funds must operate within a FISMA-compliant information architecture.

Although distributed networks can expand the scope, depth, and breadth of data available for research, there are important complexities associated with conducting research using a distributed data approach. Researchers do not have direct access to the data, making exploratory analysis more difficult. In addition, researchers in distributed networks should fully specify feasibility questions and analytic plans because the burden of each additional look at the data is higher within a distributed environment. Depending upon network governance, researchers may be limited as to the number of queries they may distribute, so as not to overburden the data partners in networks that require manual intervention for response. Variation in data partners' local environments also increases the complexity of distributing queries that will execute successfully at each site. In addition, although the data may be rich, network governance may limit querying of potentially identifiable data.

Software Enabled Governance: PopMedNet

PopMedNet is a software platform designed to facilitate the creation and operation of distributed health data networks. The software platform was designed to meet the needs of disparate data partners, coordinating center models, and researchers. Research data networks exist only if the data partners are willing to participate, so meeting their needs is critical. The PopMedNet platform's flexible architecture and governance models enable network designs that meet the critical needs of data partners within distributed networks, including data privacy and security requirements, system security requirements, governance and operational requirements, regulatory and workflow requirements, and monitoring of network functions. In networks with 50 or 100 institutions, developing an approach that meets the security needs of every individual institution is a substantial undertaking. Additionally, PopMedNet includes a number of features designed to facilitate research and network learning more broadly than simply distributing queries. These features aid researchers in identifying potential collaborators, discovering prior research conducted within their network, and understanding more about the data available within their network. Some of the key PopMedNet platform capabilities designed to meet the needs of stakeholders are described below and summarized in Table 2.

The PopMedNet platform consists of two interrelated components: a web-based portal for distributing requests and administering the network, and the DataMart Client (see DataMart Client section, below). These are illustrated in Figure 1, which represents an implementation of PopMedNet within a secure architecture. Other implementation options are possible.

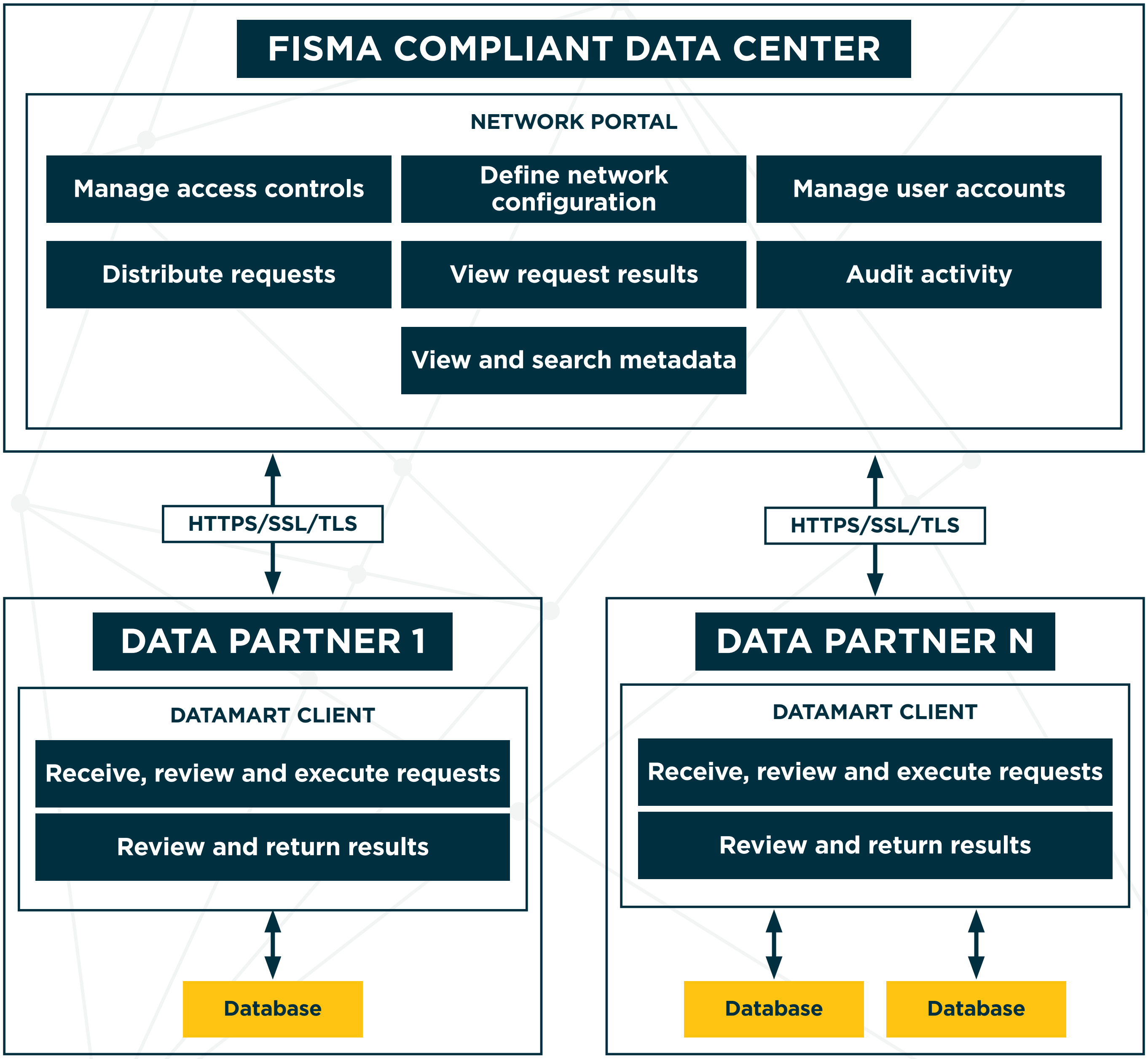
DataMart Client

PopMedNet was designed to overcome data partners' security, operational, confidentiality, and privacy concerns.[18,24-27] PopMedNet uses a publish-and-subscribe approach that does not require any open ports or Virtual Private Networks (VPNs),

Table 2. Summary of PopMedNet Features to Address Stakeholder Needs

PARTY	FEATURE
Data Partner	• Secure DataMart Client installed behind firewall • Customizable request response workflow • Network governance enforcement
Coordinating Center	• Customizable network configuration and access controls • Auditable request activity • Reporting capabilities
Researcher	• Menu-driven query interfaces • File distribution capabilities • Searchable metadata

Figure 1. A Common Secure Implementation of the PopMedNet Architecture



eliminating a critical security concern for data partners. It is installed on a data partner end user's local machine, behind the data partner firewall. There is no direct external access to local data and all queries from the network portal are pulled into the local environment rather than being pushed through an open port. The system ensures that all communications between the DataMart Client application and network portals use HTTP/SSL/TLS connections to securely transfer requests and results.

The PopMedNet architecture allows data partners to use their existing internal workflows related to use and release of data. The DataMart Client acts as an inbox for data partners to receive, review, and respond to queries distributed from a network portal. This enables data partners to review the details of all requests, including request metadata such as the name and email address of the requester, a description of the request, the purpose of the request, and the request parameters. After review, the data partner may choose to execute the query, hold it for further review, or reject it. After execution, the data partner may review the results and then decide whether to return them. Additional review or workflows consistent with local policies can also be implemented. This asynchronous approach to querying is a feature of the system that provides the data partners with complete control over their data and all its uses. Data partners can choose to automate many of the query processing steps, or choose to use the manual process to ensure compliance with local requirements. This level of data partner control creates the governance necessary to encourage data partner participation in research networks.

Network Portal

The PopMedNet network portal is used to configure and manage the network, distribute and track queries, and receive and review results.

Networks require strong governance structures that describe the nature of the network, set rules for use of data, set expectations, and outline requirements for all network participants, including researchers, funders, and data partners. The flexible architecture of PopMedNet allows implementation of a variety of governance models that can be enforced through software configurations established by the coordinating center and through coordinating center policies. PopMedNet uses a sophisticated and flexible role-based access control system to define the permissions within each network, allowing implementation of a wide range of network structures and governance models.

Network administrators define the relationships between network entities, including Organizations, DataMarts, Projects, and Users to create a custom network configuration. In addition, the system uses customizable security groups to define specific permissions and accesses with a network. For example, a security group may define the permissions for an Investigator in a project. The PopMedNet networks described here use a common set of roles to define permissions, (Table 3). An individual user's rights within a PopMedNet network are determined by the security groups applied to the user's account, enabling users to hold more than one role within a network from the same account. Permissions are granular to the degree that a security group may be granted permission to submit only a single request type to a single DataMart within the context of a single project.

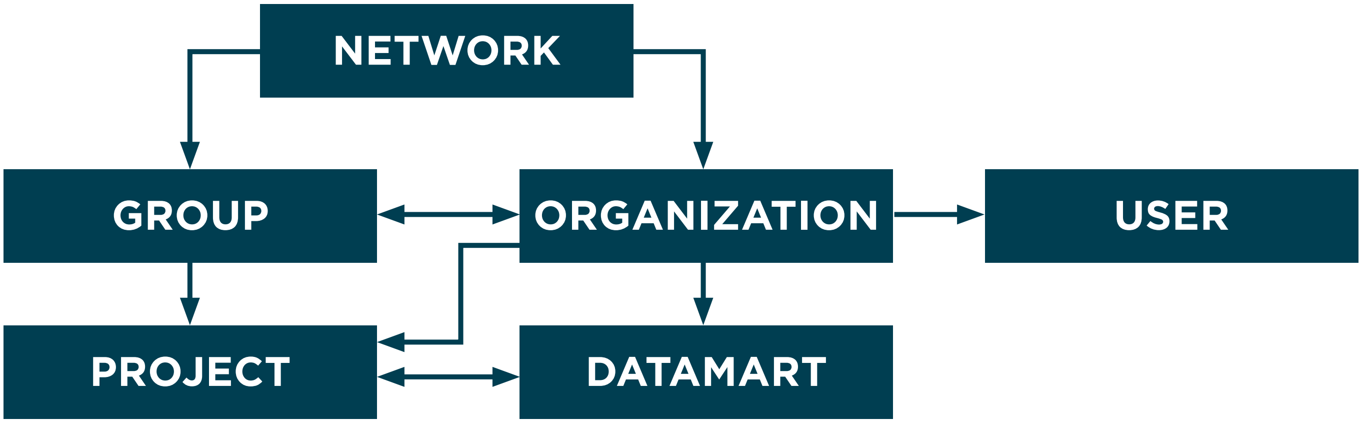
Table 3. Common Roles and Security Groups

COMMON ROLES	PERMISSIONS
DataMart Administrator	Review and respond to requests via the DataMart Client. Depending on network governance, DataMart Administrators may also manage the metadata for their DataMarts and may submit requests to their own DataMarts.
Investigator	May submit requests, and may review and export aggregated (not site-specific) results within a Project.
Enhanced Investigator	May submit requests, and may review and export disaggregated (site-specific) results within a Project.
Network Administrator	Manage the network, including creating network entities, managing access controls, and approving or creating users.
Observer	View and audit network or Project activity, excluding request results.
Enhanced Observer	View and audit network or Project activity, including request results.
Organization Administrator	Manage the metadata for their Organization and DataMarts. Monitor their DataMart activity.
Request Reviewer	Review requests before they are released to any DataMart.
Response Reviewer	Review responses for a specified DataMart or group of DataMarts before they are released to the Investigator. Depending on network governance, Response Reviewers may also have the option to group responses from multiple DataMarts into aggregate result sets before release.

Figure 2 is a stylized depiction of the major entities in a network[ii]. A Network can include several Organizations, each of which can include several DataMarts (representing a single database or data resource). One Project can include one or more DataMarts, and one DataMart can include one or more Projects. Users belong to Organizations; an Organization can include one or more users, but each user belongs to only one Organization. The PopMedNet distributed architecture is made up of a collection of Groups, Organizations, Users, Projects and DataMarts. Together, these entities can be flexibly configured to meet the needs of the network partners; different networks will have different configurations as determined by network needs and governance.

Network: a set of business entities that join together in pursuit of a common interest, typically by exchanging information with each other and collaborating to produce a work effort. In the context of this paper, networks are established at the request of a funding agency that wishes to pursue a research

Figure 2. PopMedNet Network Entity Structure



or public health initiative, such as performing postmarket medical product safety surveillance, tracking and reporting incidents of communicable diseases, or identifying patients for a clinical trial.

Organization: a collection of Users and DataMarts that model real-world business entities. Organizations may have zero, one, or more sub-organizations, but a sub-organization may have only a single parent organization. Establishing an organizational hierarchy allows some PopMedNet features to extend to sub-organizations. For instance, users with access rights to review and approve requests submitted by users in their organization can also view and approve requests of users in sub-organizations.

Group: a collection of Organizations, which enables an administrator to establish permissions and workflow across business entities that are not directly related. A group may have one or more organizations and an organization may be a member of multiple groups. Groups enable the formation of subnetworks. Subnetworks enable deployment and operation of activities that can be isolated and managed within a secure environment through access control settings.

Project: establishes security policies for composing, reviewing, and executing requests against one or more DataMarts assigned to the Project. Projects are created within an organizational group. Groups have one or more member organizations whose users and datamarts may participate within projects owned by the group.

DataMart: represents data sources used to respond to requests issued by Investigators. A DataMart is owned by a single Organization.

Each entity within a network has a profile page to capture organizational and datamart metadata that can then be used to identify potential partners. Organizational metadata include elements such as an organization description, available data resources and data models, local expertise, and supported data models. Datamart metadata include information regarding the data model utilized, data elements captured, and periods of data capture. Researchers may search for organizations and datamarts to discover potential sources of collaboration. Additionally, researchers can search request metadata to help them identify prior network activity that may be relevant for their current planned activities, helping to leverage prior work for additional analyses.

The flexible network configuration and access control system enables enforcement of varying governance policies. Examples of these governance policies from existing networks are listed below.

- Network Administrators are identified and have responsibility for network operations.
- New data partners and network users can only be added to the network by the Network Administrator and in accordance with network governance policies.
- Users can have one or more roles in the network; those roles are assigned via access controls that give users permission to perform certain functions; the Network Administrator is responsible for managing User access control settings.
- Specific users (or groups of users based on role) may view site-specific results. Other users are able to view only aggregated results. Network rules ensure results cannot be disaggregated.
- Data partners appoint one or more individuals to serve as DataMart Administrators for their sites. DataMart Administrators are responsible for responding to queries distributed to their DataMart through the network.
- DataMart Administrators retain full control over access to their data and of the transmission of query results. They have the ability to accept or reject each query on a case-by-case basis.
- Data partners may use the network to query their own data. Some networks allow data partners to query across the networks, others do not.
- DataMart Administrators can, at any time, create audit reports of activity related to their datamart.
- DataMart Administrators determine their datamart access settings, including contact information and the users and organizations who are able to send queries. These settings can be changed at any time.
- Network Administrators will not alter any datamart settings without prior approval of a DataMart Administrator; DataMart Administrators can opt to be alerted via email when any datamart settings change.
- Query results may not be used in a proposal or in any report without the consent of the network member organization where the data originated.
- For feasibility requests, no publication or external report other than use in research proposals is permitted.
- All use is monitored and reviewed every three months.
- Data on the network are deleted periodically, as determined by the network.

Distributed Queries from the Network Portal

The PopMedNet platform provides both the structure of a distributed research network and the querying capabilities for the network's activities. PopMedNet uses "request types" to help differentiate types of requests (i.e., queries) sent by researchers. This enables data partners to customize response workflows based on the type of request and the requester. PopMedNet request types include a variety of menu-driven queries, file distribution, and modular program distribution requests. Menu-driven queries provide a simple interface to define query parameters for a select set of data types within a common data model. File Distribution and Modular Program Distribution request types enable more complex querying. Modular Programs are a specific set of file distribution requests used by the United States Food and Drug Administration's (FDA) Sentinel project.[31]

Operational network and querying metadata are captured at every step of the query life cycle. The system automatically captures all request metadata (e.g., request name, requester, project and grant information, priority, due date, and request description) and additional information at each step in the life cycle, including submission date, completion date, sender, locations sent to, and responders. Email notifications containing this metadata are triggered by actions within the life cycle. Once captured, all of this information can be reviewed and reported by network participants, including the coordinating center, data partners, and others (e.g., funders).

The PopMedNet networks described in this paper are hosted within a FISMA-compliant data center. In addition to physical and operational security enforced via FISMA, PopMedNet has undergone several third-party software security reviews and penetration tests. This level of secure application hosting and security review has been undertaken to provide assurance to data partners and funders regarding the security of the network and any information that flows across the networks.[32]

Selected Case Studies

Curtis et al. (2014) describes four health data networks implemented using PopMedNet.[14] The case studies below reference the same networks, providing details on how PopMedNet is used to enforce the networks' governance models. The governance models of the four networks are similar in that all queries (i.e., requests) originate from a single institution or partner and data partners are not permitted to send queries to each other. Alternative models, such as the model developed by the Health Care Systems Research Network (formerly the HMORN), may allow for cross-data partner querying without coordinating center oversight.[16]

Sentinel

In 2008 the FDA announced the Sentinel initiative and established goals for the program, including the ability to actively monitor medical products safety by querying the electronic health data of at least 100 million patients.[33,34] The network consists of 18 data partners that have responded to over 1,000 requests.[35] The Sentinel system utilizes a strict governance process to maintain quality, accuracy, and consistency across queries. All queries are initiated by the FDA and sent via the PopMedNet system, giving data partners a single source of queries (the Sentinel Operations Center) and a single contact point (the DataMart Client) for all project queries. Governance does not permit network partners to send queries to each other; a restriction enforced via access controls and other network configuration settings. Governance rules specify the level of data aggregation permitted for each request, how data partners identities are protected, and the appropriate uses of query results. These governance policies eliminate uncertainty regarding who is distributing a query and what will happen with the response.

PCORnet

The goal of the national Patient-Centered Clinical Research Network (PCORnet) is to create a national, representative, patient-centered, clinical research network capable of supporting high-quality, observational comparative effectiveness research and clinical trials that are embedded in clinical care settings. Data are held within Clinical Data Research Networks (CDRNs) and Patient Powered Research Networks (PPRNs).[36,37] PCORnet is using PopMedNet to create the technical infrastructure to enable the network to collaborate across these disparate institutions and issue queries across over 60 distinct datamarts. As possible, PCORnet network governance (once finalized) will be implemented using PopMedNet functionality, including network configuration, access controls, and permissions.

MDPHnet

MDPHnet is a public health surveillance network led by the Massachusetts Department of Public Health (MDPH). It supports routine surveillance

Table 4. Overview Comparison of Select Networks

NETWORK	STRUCTURE	COORDINATING CENTER RESPONSIBILITIES	DATA PARTNER OBLIGATIONS	POPMEDNET IMPLEMENTATION
Sentinel	Single coordinating center. Eighteen data partners, 5 that form a subnetwork.	Network administration. Distribution, processing, and monitoring of all network queries.	Respond within a time frame defined as part of an annual contract.	Modular Program, File Distribution, Data Checker, and menu-driven queries. Subnetwork utilizes response approval workflow.
Patient-Centered Clinical Research Network (PCORnet)	Single coordinating center. Thirteen Clinical Data Research Networks (CDRN), totaling over 60 distinct data partners.	Network administration. Distribution, processing, and monitoring of all network queries.	Governance policies under development.	CDRN can include multiple sub-organizations. File Distribution, Data Checker, and menu-driven (beta-testing) queries.
MDPHnet	Separate coordinating center and requester organizations. Three data partners.	Network administration. Request distribution to validate new features. Coordinating center does not issue queries; all queries initiated by the Massachusetts Department of Public Health.	Agree to review requests, but not obligated to respond.	Menu-driven queries and raw SQL code.
National Institutes of Health (NIH) Health Care Systems Research Collaboratory Distributed Research Network (NIH DRN)	Network coordinating center, 11 data partners participating in one or more projects.	Network administration. Distribution, processing, and monitoring of network queries.	Agree to review requests, but not obligated to respond.	Individual projects determine permissible request types—including File Distribution and Modular Program Distribution requests—and menu-driven querying.

and evaluation of public health interventions.[13,38,39] The network consists of three clinical practice data partners (representing over 1.5 million patients), investigators from MDPH, and a coordinating center. In contrast to the other networks referenced here, the coordinating center is not the sole originator of queries. Rather, investigators at MDPH query the data partners directly, with the coordinating center providing support as needed. MDPH investigators may use a menu-driven query interface or distribute custom analytic SQL code. An advisory panel oversees the network governance of MDPHnet.[13]

National Institutes of Health (NIH) Collaboratory Distributed Research Network (NIH DRN)

The National Institutes of Health's (NIH) Health Care Systems Research Collaboratory Distributed Research Network (NIH DRN) is designed to support multicenter studies supported by the NIH.[40] The primary use of the NIH DRN is for feasibility and preparation-to-research queries. A governance document describes the uses of the network and the network participant responsibilities.[41] The governance requirements are implemented by the network coordinating center through PopMedNet configuration settings, access controls, and permissions. Governance requires that the coordinating center distribute all queries on behalf of authorized requesters, a policy enforced by the PopMedNet software implementation. Security and query metadata policies are also enabled by the software implementation and architecture setup by the coordinating center.

Issues and Challenges

While there are several examples of successful distributed research networks, implementing a successful governance model within a distributed research network represents a significant challenge. Although data remain behind local firewalls, participating institutions still need to understand who has the ability to send queries and need to trust that the network is configured in accordance with policies. In our experience, trust is not gained via legal agreements, but is built over time from interactions and experiences. PopMedNet's asynchronous query approach (i.e., use of manual steps to review queries and send responses) and publish-and-subscribe architecture enables data partners (and networks) to use their own internal governance models rather than having to adopt a one-size-fits-all model inconsistent with their institutional policies. This lowers barriers to participation in research networks. Over time, as trust is earned, data partners may become increasingly willing to participate further, such as by automating query responses or accepting requests from additional requesters.

Scalability is also a consideration when formulating a governance model. Request workflows, roles, and business processes must be clearly defined to ensure that the network can easily expand to new data partners; the more granular and flexible the system, the more complex it is to manage. Additionally, as more types of data, analytic methods, and querying approaches are developed, both governance approaches and the software itself must adapt to account for these changes. For example, methods for privacy-preserving distributed regression have the potential to completely enable automated, distributed comparative effectiveness research.[42-46] Although this would be a great advance in our ability to conduct distributed research, it fundamentally changes the nature of "query" and would require new workflows, metadata, approval mechanisms, and notifications to take full advantage of the new querying capability. The PopMedNet platform is being enhanced to allow efficient implementation of these newly developed, privacy-preserving, distributed queries capabilities.

Conclusion

The PopMedNet software platform supports the governance of distributed research networks through a variety of features designed to meet the disparate needs of network stakeholders. The platform enables implementation of a highly secure, customizable network infrastructure that provides granular and flexible access controls to enforce network governance; querying capabilities; and network monitoring, tracking, and reporting functions. PopMedNet encourages data partner participation by enabling questions to be sent to data that remain under the complete operational and physical control of the data holders, and allow data partners to maintain their existing internal workflows related to data access, use, and sharing. The metadata captured by the system contains rich information on request activity and on the participating sites and data resources in a network. This metadata may be monitored, queried, and used to generate reports. As evidenced by the multiple distributed research networks using PopMedNet, the system can support a variety of network structures, governance models, and activities.

References

1. Friedman CP, Wong AK, Blumenthal D. Achieving a Nationwide Learning Health System. *Sci. Transl. Med*. 2, 57cm29 (2010).
2. Oye, K., Jain, G., Amador, M., et al. (2015), The next frontier: Fostering innovation by improving health data access and utilization. *Clinical Pharmacology & Therapeutics*, 98: 514–521. doi: 10.1002/cpt.191
3. 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*. 2015 Jan;22(1):43-50. doi: 10.1136/amiajnl-2014-002977. Epub 2014 Oct 23.
4. Collins FS. Exceptional opportunities in medical science: A view from the National Institutes of Health. *JAMA*. 2015; 313(2):131-132. doi:10.1001/jama.2014.16736.
5. The White House. Precision Medicine Initiative: Privacy and trust principles. Available at: https://www.whitehouse.gov/sites/default/files/microsites/finalpmiprivacyandtrustprinciples.pdf. Published November 9, 2015. Accessed December 30, 2015.
6. National Institutes of Health. About the Precision Medicine Initiative Cohort Program. Available at: https://www.nih.gov/precision-medicine-initiative-cohort-program. Accessed December 30, 2015.
7. National Institutes of Health. About BD2K. Available at: https://datascience.nih.gov/bd2k/about. Accessed December 30, 2015.
8. 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*. 2014 Jul-Aug;21(4):578-82. doi: 10.1136/amiajnl-2014-002747. Epub 2014 May 12.
9. Innovative Medicines Initiative. EHR4CR: Electronic Health Records System for Clinical Research. Available at: http://www.imi.europa.eu/content/ehr4cr. Accessed December 30, 2015.
10. TRANSFoRm. Project Overview. Available at: http://www.transformproject.eu/about/project-overview/. Accessed December 30, 2015.
11. Delaney BC, Curcin V, Andreasson A, et al., "Translational Medicine and Patient Safety in Europe: TRANSFoRm—Architecture for the Learning Health System in Europe," *BioMed Research International*, 2015; 2015:1-8. doi:10.1155/2015/961526
12. AMCP Task Force on Biosimilar Collective Intelligence Systems, Baldziki M, Brown J, Chan H, et al. Utilizing data consortia to monitor safety and effectiveness of biosimilars and their innovator products. *J Manag Care Spec Pharm*. 2015 Jan; 21(1):23-34.
13. Vogel J, Klompas M, Brown JS, Land T, Platt R. MDPHnet: Secure, Distributed Sharing of Electronic Health Record Data for Public Health Surveillance, Evaluation, and Planning. *Am J Public Health*. 2014 Dec;0: e1-e6.
14. Curtis LC, Brown JS, Platt R. Four Health Data Networks Illustrate The Potential For A Shared National Multipurpose Big-Data Network. *Health Affairs*. 2014; 33(7):1178-1186.
15. Holmes JH, Elliott TE, Brown JS, et al. Data warehouse governance for distributed research networks in the United States: a systematic review of the literature. *J Am Med Inform Assoc*. 2014 Jul-Aug; 21(4):730-6.
16. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: A public data model to support collaboration. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2014; 2(1): Article 2.
17. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf*. 2012;21(S1):23–31.
18. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med*. 2009; 5:341-344.
19. Meeker D, Jiang X, Matheny ME, et al. A system to build distributed multivariate models and manage disparate data sharing policies: implementation in the scalable national network for effectiveness research. *J Am Med Inform Assoc*. 2015 Nov; 22(6):1187-95. doi: 10.1093/jamia/ocv017. Epub 2015 Jul 3.
20. Kim KK, Browe DK, Logan HC, et al. Data governance requirements for distributed clinical research networks: triangulating perspectives of diverse stakeholders. *JAMIA*. 2013; 21:714–719. doi:10.1136/amiajnl-2013-002308.
21. Schilling L, Kwan B, Drolshagen C et al. Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) Technology Infrastructure for a Distributed Data Network. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2013;1(1). doi:10.13063/2327-9214.1027.
22. McMurry AJ, Murphy SN, MacFadden D, et al. (2013) SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies. *PLoS ONE* 8(3): e55811. doi:10.1371/journal.pone.0055811
23. Holmes JH, Nelson AF, Raebel MA, et al. "Scalable PArtnering Network (SPAN) for Comparative Effectiveness Research (CER): Purpose, Structure, and Operations" (2013). Governance Toolkit. Paper 3. Available at: http://repository.academyhealth.org/govtoolkit/3

24. 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:S45-51.
25. Brown JS, Holmes JH, Maro J, et al. Design specifications for network prototype and cooperative to conduct population-based studies and safety surveillance. Effective Health Care Research Report No. 13. (Prepared by the DEcIDE Centers at the HMO Research Network Center for Education and Research on Therapeutics and the University of Pennsylvania Under Contract No. HHSA29020050033I T05.) Rockville, MD: Agency for Healthcare Research and Quality. July 2009. Available at: https://effectivehealthcare.ahrq.gov/reports/final.cfm.
26. Brown JS, Holmes JH, Syat B, et al. Proof-of-principle evaluation of a distributed research network. Effective Health Care Research Report No. 26. (Prepared by the DEcIDE Centers at the HMO Research Network and the University of Pennsylvania Under Contract No. HHSA29020050033I T05.) Rockville, MD: Agency for Healthcare Research and Quality. June 2010. Available at: http://effectivehealthcare.ahrq.gov/reports/final.cfm.
27. Brown JS, Syat B, Lane K, et al. Blueprint for a distributed research network to conduct population studies and safety surveillance. Effective Health Care Research Report No. 27. (Prepared by the DEcIDE Centers at the HMO Research Network and the University of Pennsylvania Under Contract No. HHSA29020050033I T05.) Rockville, MD: Agency for Healthcare Research and Quality. June 2010. Available at: http://effectivehealthcare.ahrq.gov/reports/final.cfm.
28. Klann JG, Buck MD, Brown JS, et al. Query Health: Standards-based, cross-platform population health surveillance. *J Am Med Inform Assoc*. 2014 Jul; 21(4): 650–656.
29. Brown JS, Davidson A, Elliott TE, et al. “Scalable PArtnering Network (SPAN) for Comparative Effectiveness Research (CER): Research User Interface Principles and Requirements” (2013). Governance Toolkit. Paper 2. Available at: http://repository.academyhealth.org/govtoolkit/2.
30. Sittig DF, Hazlehurst BL, Brown JS, et al. A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogeneous clinical data. Med Care 2012;50: S49-S59.
31. Mini-Sentinel. Routine Querying Tools (Modular Programs). c2010-2011, updated 2015 Jun 4. Available at: http://www.mini-sentinel.org/data_activities/modular_programs/default.aspx. Accessed August 6, 2015.
32. PopMedNet. System Security. PopMedNet Wiki; January 28, 2015, updated February 23, 2015. Available at: https://popmednet.atlassian.net/wiki/display/DOC/System+Security
33. Food and Drug Administration. FDA’s Sentinel Initiative. July, 2010, updated November 21, 2014. Available at: http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm. Accessed August 6, 2015.
34. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System - A national resource for evidence development. *N Engl J Med*. 2011; 364:498-499.
35. Woodworth, T. PopMedNet in Mini Sentinel. PopMedNet User Conference; July 27, 2015; Simmons College, Boston, MA.
36. PCORnet. PCORnet: the National Patient-Centered Clinical Research Network. Available at: http://pcornet.org/. Accessed August 6, 2015.
37. Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. *Health Aff*. 2014; 33(7):1212–19.
38. Mass eHealth Initiative. MDPHnet. Available at: http://mehi.masstech.org/programs/past-programs/mdphnet. Accessed August 6, 2015.
39. Department of Population Medicine. ESPnet: EHR Support for Public Health. Available at: http://esphealth.org/. Accessed August 6, 2015.
40. NIH Health Care Systems Research Collaboratory. NIH Collaboratory Distributed Research Network. Available at: https://www.nihcollaboratory.org/Pages/distributed-research-network.aspx. Accessed August 6, 2015.
41. NIH Health Care Systems Research Collaboratory Distributed Research Network. NIH Distributed Research Network: Policies and Procedures. 2013. Available at: https://www.nihcollaboratory.org/Products/NIH-DRN-Governance_v1.0.pdf
42. Jiang W, Li P, Wang S, et al. WebGLORE: a Web service for Grid LOgistic REgression. *Bioinformatics*. 2013; 29(24):3238–3240.
43. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med Care*. 2013 Aug; 51(8 Suppl 3):S4-10. doi: 10.1097/MLR.0b013e31829b1bb1.
44. Karr A. Secure statistical analysis of distributed databases, emphasizing what we don’t know. *Journal of Privacy and Confidentiality*. 2009; (2):197-211. Available at: http://repository.cmu.edu/cgi/viewcontent.cgi?article=1011\&context=jpc.
45. Karr AF, Lin X, Sanil AP, Reiter JP. Secure regression on distributed databases. J*ournal of Computational and Graphical Statistics*. 2005; 14(2):263-279. doi:10.1198/106186005X47714.
46. Karr AF, Fulp WJ, Vera F, Young SS, Lin X, Reiter JP. Secure, privacy-preserving analysis of distributed databases. *Technometrics*. 2007; 49(3):335-345. doi:10.1198/004017007000000209.