Developing Governance for Federated Community-based EHR Data Sharing

Ching-Ping Lin PhD1,2, Kari A. Stephens PhD1, Laura-Mae Baldwin MD MPH1, Gina A. Keppel MPH1, Ron J. Whitener JD1, Abigail Echo-Hawk MPS1, Diane Korngiebel DPhil1

1Institute of Translational Health Sciences, University of Washington, Seattle, WA, 2Global REACH, University of Michigan, Ann Arbor, MI

Abstract

Bi-directional translational pathways between scientific discoveries and primary care are crucial for improving individual patient care and population health. The Data QUEST pilot project is a program supporting data sharing amongst community based primary care practices and is built on a technical infrastructure to share electronic health record data. We developed a set of governance requirements from interviewing and collaborating with partner organizations. Recommendations from our partner organizations included: 1) partner organizations can physically terminate the link to the data sharing network and only approved data exits the local site; 2) partner organizations must approve or reject each query; 3) partner organizations and researchers must respect local processes, resource restrictions, and infrastructures; and 4) partner organizations can be seamlessly added and removed from any individual data sharing query or the entire network.

Introduction

A key aim of the National Institutes of Health (NIH) Roadmap has been to broaden the participation of communities and practice-based care settings in medical and health services. Bi-directional translational pathways between scientific discoveries and primary care are crucial for improving individual patient care and population health. The Institute of Translational Health Sciences (ITHS) developed the Data QUEST (Data QUery Extraction Standardization and Translation) pilot project to build a technical infrastructure to support the sharing of electronic health record (EHR) data across primary care practices and tribal communities. Data QUEST targets engagement with disparate primary care practice based organizations serving small rural populations because they are often excluded from research and dissemination efforts.

While gathering technical requirements, we also deliberately engaged in a time and resource intensive process to learn about governance requirements. We knew that the success of our project depended on our ability to foster trust and consider complicated and independent governance issues across community based practice partners, many of which served underserved populations and some of which involved communities with tribal sovereignty.

Background

Data governance is defined as the process by which the responsibilities of data stewardship (the acquisition, storage, aggregation, de-identification, release, and use of data) are conceptualized and carried out through policies and approaches. Data governance is crucial for maintaining privacy protection for community members at the individual and group level. The research community may be familiar with HIPAA protections that address individual data privacy through protected health information, but anonymous aggregated data across groups can also be damaging by characterizing subsets of patient populations leading to stigmatization (e.g., substance abuse, sexually transmitted infection, mental illness, obesity rates) and thus need similar governance policies.

Creating new data sharing pathways must consider thoughtful changes to data governance, which involve high levels of investment and trust by partner organizations. Data sharing systems can enforce different governance requirements through authentication models and levels of automation for querying and receiving data. However, governance requirements are independent of any specific implementation approach. We needed to establish proper governance requirements within Data QUEST that met the needs of partner organizations before determining our implementation approach.
Methods

We reviewed requirements for gathering data with five WPRN partner organizations and five American Indian/Alaska Native (AI/AN) tribal practices located across the WWAMI (Washington (n = 5), Wyoming (n = 0) Alaska (n = 2), Montana (n = 1), Idaho (n = 2)) region. The purpose of these discussions was to determine both technical and governance readiness for Data QUEST. On average, each non-tribal practice (primary care community-based practice, n = 5) supported 30-35 clinical providers per organization, variably dividing their time between clinical, administrative, and teaching duties. The AI/AN partners (n = 5) included two general types of practices: practices for which tribes received funding from the United States government for clinical operations, but were managed by the local tribal administration and governance and practices funded and managed by the Indian Health Service, a division of the United States Health and Human Services. They supported a range of 3-5 providers generally working full time and served a similar volume of visits per year (ranged from 10,000-40,000).

We spoke to a diverse set of practice leaders, including clinicians, technical staff, and administrators to identify stakeholders and leaders necessary to support and authorize a data sharing project. We presented current governance practices based on existing practice based research networks and their related data sharing efforts and tools to drive discussions of governance requirements (i.e., Group Health Research Institute’s Virtual Data Warehouse10, DARTNet Institute’s data sharing infrastructure12). While working with tribal based practices, we also consulted with tribal leaders and received approval by the tribal governing bodies. We recorded feedback from community liaison and investigator partners.7 As we cycled through iterations with partners in developing the Data QUEST technical architecture, we addressed technical implications of the governance requirements.13 We recorded local governance and engagement procedures, as well as engaged partners in iterating governance requirements for Data QUEST data sharing. Qualitative data gathered through this iterative design process defined governance requirements.

Results

From continued engagement with partners, we identified four governance requirements and governing principles that we outline below:

Governance Requirement 1: Organizations can physically terminate the link to the network and only approved data leaves the local site.

Two common approaches to data sharing are: 1) a centralized approach in which data from each organization are aggregated and stored in a single physical repository for inquiry and 2) a federated approach in which the data remain at each practice with access to the data through a virtual repository.14 Our partners clearly preferred a federated approach.

Interviews with partners revealed concerns over data ownership, control, and security of identifiable data. In centralized architectures, while partners can approve or deny data requests either through technical or social means, they typically do not have physical control over the data and servers, which are managed by the centralized data steward. In a federated system, if necessary, the local data manager can simply physically turn off the database or server, and the local data can no longer be shared. This capability to “flip the off switch” was important to our partners, given that the identified data included Protected Health Information. Another motivation for a federated approach was the desire to only release sanctioned data to the outside as opposed to releasing all the possible sharable data into a centralized repository and rely on a centralized data steward to facilitate access.

Figure 1 shows Data QUEST’s federated architecture. Organization A, Organization B, and Organization C are part of the Data QUEST network, and all the data reside locally at the organization (within the boundaries of the local firewall system) until they decide to share data with outside partners for a specific query or project. No centralized warehouse or database contains all of the data in the network. The “combined health data” are aggregated on a project-by-project basis.
Governance Requirement 1 (physically terminate the link to the network and only approved data leaves the firewall), and Governance Requirement 2 (approve or reject each query)

Governance Requirement 2: Partners must approve or reject each query

Our partners expressed concern that users of the Data QUEST network might “data fish,” that is, download and analyze health data without a constructive or approved research question. Data fishing could cause exploitation or vulnerabilities, due to sensitive disease-based issues that may stigmatize communities or reveal sensitive clinical quality issues. Tribal communities related these preferences particularly in the context of historical contentious relationships with outside researchers, especially in the area of data sharing and publication of results without community oversight, and all of our partners reflected these preferences. Data fishing raises many ethical issues that can be addressed by federation and careful control over data ownership. Our partners stated ownership of the data, identified or unidentified, must remain in the hands of the community or practice.

Our partners also stated they must be able to review every data query request and result, with the option to deny access and/or withdraw from participation at any time, in addition to Institutional Review Board (IRB) or prior approval. Therefore, all data query requests must pass through a designated authority at the partner organization who must explicitly approve all queries before results are delivered.

Governance Requirement 3: Partner organizations and researchers must respect local processes, resource restrictions, and infrastructures

Each practice and community has unique approval processes, human resources, and different methods for engagement. From gathering information on local processes and social structures, we recognized our third governance requirement to respect local processes, resource restrictions, and infrastructures. This requirement is reflected through a set of core data sharing documents that govern data sharing:

- Data Use Agreement (DUA) authorizing the sharing of the health data with partners
- Publication Policy (PP) outlining how publications and presentations presenting data will be vetted
- Memorandum of Understanding (MOU) creating an agreement to participate in specific research projects

In addition to DUAs, PPs, and MOUs, each organization also required Business Associate Agreements with our vendor, Data Transfer Agreements between the vendor and the organization, and contracts between the university and the vendor.

Figure 1. Governance Requirement 1 (physically terminate the link to the network and only approved data leaves the firewall), and Governance Requirement 2 (approve or reject each query)

Figure 2. Heterogeneous local governance infrastructure examples
Figure 2 shows an example of two Data QUEST organizations that have different data sharing governance structures in terms of documents, policies, and governing boards. At both organizations A and B, we have developed Data Use Agreements. Organization B required a Memorandum of Understanding before we could begin to implement our data sharing infrastructure. Our tribal partners required an additional publication policy that the other partner organizations subsequently adapted and adopted. These policies detailed how publications, presentations and abstracts will be vetted, how communities will be described in manuscripts, and an arbitration system for disputes. Figure 2 shows that Organization A uses a research regulatory committee to oversee research activities, while at Organization B, the Tribal Council oversees research activities related to the tribal clinic. The implementation teams for the AI/AN communities and WPRN worked together to share governance materials and processes that facilitated development. Commonalities were identified and eventually we produced specific documents to be used by each group (e.g., DUAs and PPs).

Developing Data QUEST required working within these local processes. Any researchers wishing to partner with Data QUEST must also respect and comply with these established processes.

**Governance Requirement 4: Organizations can be seamlessly added and removed from data queries or from Data QUEST itself**

The goal of the Data QUEST pilot project was to start with a small initial number of community-based partners to develop the foundational infrastructure and proof of concept for a data sharing network, with the intention to grow. We have also engaged national efforts, most notably with the WPRN joining the DARTNet Institute, expanding our data sharing capacity nationally. Anticipating the potential need and desire to grow partnerships and engage in parallel / national efforts, we recognized that the governance process must be flexible enough to support the ability to add (or remove) Data QUEST organizations without interrupting data sharing capabilities for our participating organizations.

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Data QUEST’s governance must permit the addition (Org D) and subtraction (Org C) of organizations easily from individual data sharing projects or from Data QUEST itself.

Partner organizations do not have the ability to vote additional partners in or out, nor have any influence over other organizations’ participation in Data QUEST or in any individual query or project. Instead, all partners are allowed to participate or withdraw at any time for any individual data query / project or from Data QUEST. This flexibility is easily facilitated because of the network’s federated architecture as discussed above. In Figure 3, Org D has decided to join the network and Org C has decided to withdraw from the network. Org C is able to withdraw by simply disconnecting. This “disconnection” may be a physical technical connection that can be turned “off” or it may be an abstract connection with the network or both.

**Discussion**

Community-based partners require local control over their data and are reluctant to allow their data to reside permanently outside their domain of control, even if they have full oversight over the data. In addition, building trust between academic research and community-based partners can be facilitated by developing systems that maximize local control and allowing for expectations of high granular control over the flow of data to researchers. Community-based partners require the ability to control the data flow at the individual query level, not just at the research project level, which may include several queries, though they may choose to grant access to data on a per project basis. Because the Data QUEST network consists of disparate and geographically distant organizations without shared governance, some of which involve tribal nations with complex legal requirements, we needed to
understand and support each partner’s local governance infrastructure and research processes, including our own institution’s requirements for facilitating research or receiving health data from non-affiliated institutions. This process was time consuming and critical to the success of our network development efforts.

**Multiple approaches can fulfill governance requirements**

It is important to re-iterate that none of these requirements define a specific technical solution. They define a set of parameters that any automated or non-automated solution must meet, but they leave open the possibility for multiple approaches. For instance, Governance Requirement 2 (partners must approve or reject every query) can be met through a software functionality where the organization can view and approve the query electronically or through a manual, people-based process.

We did not develop customized data sharing software for Data QUEST due to resource restraints. However, other efforts are underway to develop and distribute software that instantiates many of these requirements (e.g. i2b2\(^1\), PopMedNet\(^6\)). The complexity of adapting one of these budding technologies to meet our immediate governance requirements was not feasible. We therefore used a mixed social and technical approach to support the workflow and governance of the network.

**Continuous and iterative engagement of partners is crucial**

Determining governance and technical dimensions in primary care community-based settings for growing data sharing across EHR data is best addressed through iterative and inclusive involvement with community partners. Community partners must be engaged from the onset in developing technical requirements to ensure that governance requirements are incorporated and implementation of pre-built software meets their needs.

However, finer control also adds to the workload of each partner to manage their local database, queries from researchers, and proposed research projects potentially slowing down the functionality of the entire network while some sites wait on others for decisions. Additionally, Governance Requirements 2 and 3 also create a direct relationship between the number of sites in the network and the efficiency of the system - the more sites, the more site-specific complexity is added to the network.

It takes significant time to meet with partners about their concerns and needs. However, these discussions are crucial because resources vary across organizations. For instance, one organization had a research project coordinator who could serve as a dedicated liaison to manage data requests, whereas another organization did not. This iterative process helped determine whether partners felt they could participate and revealed additional system requirements to be built in the context of each organization’s existing resources. As data sharing networks and the volume of research partnerships grow, so does the need for managing data requests and communication between communities, practices, investigators, and information technology specialists. All stakeholders must be engaged in this process and iterate together.

**Conclusions**

We have presented the governance requirements for Data QUEST, a pilot project building a data sharing architecture of community-based and tribal primary care practices across the northwest region: 1) organizations can physically terminate the link to the network and only approved data leaves the firewall; 2) partners must approve or reject each query; 3) partner organizations and researchers must respect local processes, resource restrictions, and infrastructures; and 4) organizations can be seamlessly added and removed from a query or from the network itself. Support for these requirements may be automated or manual.

It is crucial to build data sharing networks in community-based settings so that translational science can succeed at bridging scientific discovery to front line treatment environments in primary care settings. Researchers can capitalize on these networks to speed significant health impacts by conducting work in these real-world settings. These networks can promote inclusion of underrepresented and rurally located people in research, who are so often missed in research conducted in academic-based settings. Dissemination science is growing and the need to conduct comparative effectiveness trials will be served well by the efficiencies offered within these complex data sharing environments. Time and resources are precious in these settings however, and care and iteration must be engaged to develop these networks effectively.
Acknowledgements

This publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000423. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Many thanks to our partner organizations in the Data QUEST pilot program, Alfred O. Berg, Peter Tarczy-Hornoch, and our colleagues in the Community Outreach and Biomedical Informatics programs of the ITHS.

References

1. Zerhouni EA. US Biomedical Research: Basic, Translational, and Clinical Sciences. JAMA. 2005 September 21, 2005;294(11):1352-8.
2. Westfall JM, Mold J, Fagnan L. Practice-Based Research--"Blue Highways" on the NIH Roadmap. JAMA. 2007 January 24, 2007;297(4):403-6.
3. CTSA Principal Investigators. Preparedness of the CTSA's Structural and Scientific Assets to Support the Mission of the National Center for Advancing Translational Sciences (NCATS). Clinical and Translational Science.5(2):121-9.
4. Lin CP, Black RA, Laplante J, et al. Facilitating health data sharing across diverse practices and communities. AMIA Summits Transl Sci Proc. 2010;2010:16-20.
5. Hodge FS, Weinmann S, Roubideaux Y. Recruitment of American Indians and Alaska Natives into clinical trials. Ann Epidemiol. 2000 Nov;10(8 Suppl):S41-8.
6. Lin CP, Stephens KA, Black RA, et al. Facilitating Health Data Sharing Across Diverse Practices and Communities. 2010 AMIA Summit on Clinical Research Informatics. San Francisco, CA: AMIA; 2010.
7. Stephens KA, Anderson N, Lin CP. Developing best practices for evaluating federated data sharing: Approaches from academic hospital and primary care clinic networks. AMIA Annu Symp Proc. 2010.
8. Rosenbaum S. Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. Health Services Research.45(5p2):1442-55.
9. Lin K, Daemer G. caBIG Security Technology Evaluation White Paper: National CAnceer Institute; 2006 January 23, 2006.
10. Vogt TM, Elston-Lafata J, Tolsma D, Greene SM. The role of research in integrated healthcare systems: the HMO Research Network. Am J Manag Care. 2004 Sep;10(9):643-8.
11. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association. March 1, 2010;17(2):124-30.
12. Pace WD, Cifuentes M, Valuck RJ, et al. An electronic practice-based network for observational comparative effectiveness research. Ann Intern Med. 2009 Sep 1;151(5):338-40.
13. Stephens KA, Lin CP, Baldwin LM, et al. LC Data QUEST: A Technical Architecture for Community Federated Clinical Data Sharing. 2012 AMIA Summit on Clinical Research Informatics. San Francisco, CA: AMIA; 2012.
14. Schatz B, Mischo WH, Cole TW, et al. Federating Diverse Collections of Scientific Literature. Computer. 1996;29(5):28-36.
15. Whitener RJ. Research in Native American Communities in the Genetics Age: Can the Federal Data Sharing Statute of General Applicability and Tribal Control of Research Be Reconciled? J Tech L & PoPy. 2010;15:217.
16. Brown JS, Holmes JH, Shah K, et al. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Med Care. 2010 Jun;48(6 Suppl):S45-51.