9-19-2013

Pathways to Success for Multi-Site Clinical Data Research

Deven McGraw
Center for Democracy and Technology, deven@cdt.org

Alice B. Leiter
Center for Democracy & Technology, aleiter@cdt.org

Recommended Citation

Pathways to Success for Multisite Clinical Data Research

Deven C. McGraw JD, LLM, MPH; Alice B. Leiter JD[i]

Abstract

There are numerous and significant challenges associated with leveraging electronic clinical data (ECD) for purposes beyond treating an individual patient and getting paid for that care. Optimizing this secondary use of clinical data is a key underpinning of many health reform goals and triggers numerous issues related to data stewardship and, more broadly, data governance. These challenges often involve legal, policy, and procedural issues related to the access, use, and disclosure of electronic health record (EHR) data for quality improvement and research. This paper contributes to the ongoing discussion of health data governance by detailing the experiences of nine multisite research initiatives across the country. The rich set of experiences from these initiatives, as well as a number of resources used by project participants to work through various challenges, are documented and collected here for others wishing to learn from their collective efforts. The paper does not attempt to catalog the full spectrum of governance issues that could potentially surface in the course of multisite research projects using ECD. Rather, the goal was to provide a snapshot in time of data-sharing challenges and navigation strategies, as well as validation that privacy-protective, legally compliant clinical data sharing across sites is currently possible. Finally, the paper also provides a foundation and framing for a broader community resource on governance—a "governance toolkit"—that will create a virtual space for the further discussion and sharing of promising practices.

Introduction

Federal efforts to reform the nation's health care system are driven by the combined goals of improving patient safety and health, promoting efficiency and accountability, and, overall, achieving a "high-value" health care system.[1] In support of these goals, the government has invested tens of billions of dollars to digitize health data and facilitate its "meaningful use" for enhancing care, improving quality, and driving toward greater efficiency in the health care system. Federal and state tax dollars also are funding the development of infrastructure to support the electronic exchange of health information and its use for robust analysis in support of rapid-cycle innovation in care delivery and payment.[2]

However, there are numerous and significant challenges associated with leveraging ECD for purposes beyond treating an individual patient and getting paid for that care, which is sometimes referred to as secondary use or "reuse" of clinical data. Optimizing the reuse of data is a key underpinning of many health reform goals, and it triggers numerous issues related to data stewardship and, more broadly, data governance.[3] These challenges often involve legal, policy, and procedural issues related to the access, use, and disclosure of electronic health record (EHR) data for quality improvement (QI) and research, and these challenges and issues have been described extensively in the literature.[4,5]

This paper contributes to the ongoing discussion of health data governance by detailing the experiences of nine multisite research initiatives across the country (see Table 1). Extensive telephone interviews and in-person discussions were conducted to understand the perspectives of this diverse group of projects at the cutting edge of applied health research using ECD. The rich set of experiences from these initiatives (hereafter referred to as "project participants"), as well as a number of resources used by project participants to work through various challenges, is documented here for others wishing to learn from their collective efforts. Such resources represent key building blocks of governance and include data use agreements (DUAs), model privacy policies, and common data models.

The paper is intended to be a learning tool and an inspiration for nascent initiatives. It illustrates that, despite real and at times significant challenges, institutions are able to share clinical data across sites for research and other purposes, and they are able to do so successfully in ways that are privacy protective while also accomplishing study goals.

This paper does not attempt to catalog the full spectrum of governance issues that could potentially surface in the course of multisite research projects using ECD. Rather, this paper is designed to provide a foundation and framing for a broader community resource on governance—a "governance toolkit"—that will create a virtual space for the further discussion and sharing of promising practices. Commissioned by the

[i]Center for Democracy and Technology

Table 1. Project Participants

- Building Modular Pediatric Chronic Disease Registries for QI and CE Research (at Cincinnati Children's Hospital Medical Center [CCHMC])
- Comparative Outcomes Management with Electronic Data Technologies (COMET)
- High Value Health Care Collaborative (HVHC)
- Mini-Sentinel
- Pediatric Health Information System Plus (PHIS+)
- SCAlable National Network for Effectiveness Research (SCANNER)
- Surgical Care and Outcomes Assessment Program Comparative Effectiveness Research Translation Network (SCOAP CERTAIN)
- VA Informatics and Computing Infrastructure (VINCI)
- Washington Heights Initiative Community-Based Comparative Effectiveness Research (WICER)

Descriptions of CCHMC, COMET, SCANNER, and WICER can be found at www.academyhealth.org/files/FileDownloads/edmprofile-snov2012.pdf. More information about HVHC can be found at www.dartmouth-hitchcock.org/about_dh/hvhc_collaborative.html; more information about Mini-Sentinel can be found at www.mini-sentinel.org; more information about PHIS+ can be found at: http://prisnetwork.org/research/phis_plus.html; and more information about VINCI can be found at http://www.hsrd.research.va.gov/for_researchers/vinci/default.cfm#.Ua5IJYLudNs.

Electronic Data Methods (EDM) Forum, the toolkit will consist of a series of linked papers, issue briefs, and other resources, each of which will explore specific governance issues in more depth.

Paper Overview

This paper is organized around a number of themes identified for further exploration in the governance toolkit. The themes illustrate that, despite the very real challenges of implementing multisite data-sharing research initiatives, data reuse for systematic learning can and is being done, with great success.

Project participants encountered nine common challenges:

- Legal and regulatory concerns
- Governance implication of data network architecture
- Structure and role of governance bodies
- Institutional Review Boards
- Governance issues unique to specific data properties (e.g. behavioral health)
- Data sharing approaches and considerations
- Governance issues related to competitive marketplace situations
- Stakeholder engagement and participation
- Sustainability

Project participants also shared potential solutions and strategies to resolve or mitigate these concerns, including:

- Capitalizing on pre-existing relationships;
- Starting small, then expanding (with respect to participant number, data type, or data use);

- Developing legal and policy documents with participant input; Exchanging de-identified data; and
- Structuring governance bodies carefully and with broad representation.

A starting point for the broader governance toolkit effort, this paper sheds light on a number of the challenges and potentially promising solutions or strategies to facilitate the reuse of clinical data to drive discovery and improve care. As the number of entities engaged in such research efforts evolves, the continued collection, organization, and discussion of resources, relevant literature, and practical experience will pave the way for future thought and further exploration.

Legal Landscape

A number of data governance issues are related to compliance with existing legal requirements designed to protect the rights of patients. When it comes to reuse of health information, the two most relevant federal laws are the Health Insurance Portability and Accountability Act (HIPAA)[6] and the Common Rule.[7] The HIPAA Privacy Rule permits "covered entities," which include most health care providers and health care institutions, to access, use, and disclose protected health information (PHI) for treatment, payment, and health care operations (TPO) without the need to first obtain a patient's authorization.[8,9] Included in the category of "health care operations" is the performance of quality assessment and improvement activities, as long as the primary purpose of such activity is for internal use and does not involve contributing to "generalizable knowledge."[10]

The Common Rule covers only "research" conducted with the support of federal funding from certain agencies, including the Department of Health and Human Services (HHS), the Department of Veterans Affairs (VA), and the Department of Energy[11] As with the HIPAA Privacy Rule, the Common Rule defines research as "systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[12]

Thus, the intent to contribute to generalizable knowledge, not the specific methods or tests applied to the data, is the test for whether a use or disclosure of information is categorized as research or QI. QI is defined as internal, operational improvement work; thus, cross-organizational or collective studies may not be viewed as QI under current regulations.

When an activity is considered "research," a number of specific legal provisions likely apply, under both HIPAA and the Common Rule.[13] See Table 3 for more detailed information about these requirements.

Because information that is de-identified or less identifiable is subject to less regulation under HIPAA and the Common Rule, researchers often strive to access this type of data. However, in

Table 2. Data Governance Definition and Working Taxonomy

In this context, the term *data governance* refers to the policies, protocols, and practices necessary to:
- Establish trusted relationships,
- Define roles and accountability, and
- Formulate approaches to information stewardship and management.

Such policies, protocols, and practices must consider the influences of:
- Federal and state laws,
- Organizational culture and context,
- Biomedical ethics, and
- Business models.

Given this working definition of data governance, a "governance taxonomy" will include the following domains:
- Legal and regulatory concerns
  - Data network design (or architecture)
  - Data linkage
- Structure and role of governance bodies
  - Data security
- Institutional review board (IRB) issues
- Data properties
  - Dealing with questions about "ownership"
- Data-sharing approaches and considerations, including consent
- Competitive marketplace
  - Rights to and process for publication and/or broader dissemination.
- Stakeholder engagement and participation
- Sustainability

many cases, the need to maximize the utility of data (which can be compromised when data are de-identified or stripped of common identifiers) and the need to assure compliance with applicable law (which assumes a need to comply with the more stringent rules applicable to identifiable data) rendered issues of data identifiability less relevant than expected.

HIPAA and the Common Rule are not the only laws or rules that may apply to the exchange of health information for research or other secondary purposes. For example, there are federal regulations governing some substance abuse treatment data, and there are a number of state medical laws that typically govern certain types of sensitive health information. A discussion of these laws is beyond the scope of this paper.[14]

Common Governance Challenges

Reusing clinical data for purposes of QI or research commonly poses a number of governance challenges, which are often cited as having a chilling effect on a health care entity's willingness to share or grant access to health data for the purpose of anything other than TPO. These high-level challenges arise and are managed in differing ways depending on the specifics of a particular data-sharing network or collection of clinical sites.

Uncertainty about the Reuse of Data

As discussed above, the primary federal provisions governing use of data for secondary purposes attempt to draw a line between information collection, use, or disclosure that intends to contribute to the generalizable knowledge base of the health care community (research) and information evaluation activities for internal use (operations) or application to a specific patient population (QI). Drawing this distinction in circumstances where the consequences of getting it "wrong" can result in legal liability poses a challenge—one that entities with little experience in reusing their clinical data (or making it available for others to use) find particularly daunting.

As a result, the applicable or relevant legal framework governing data sharing is frequently noted to be a challenge for multisite initiatives. In other words, opportunities for multiple institutions to share data for secondary purposes, including research, are hampered by nonuniform policies across institutions that can make legal compliance confusing or challenging, as institutions fear insufficient protection for the data.

Though such collaboration was time-intensive, the **Comparative Outcomes Management with Electronic Data Technologies** (COMET) initiative in particular indicated that investing time and, in equal measure, working through the legal implications of the project with the institutional attorneys at each of the individual participant sites helped assure legal compliance and also built a more trusting relationship among them.

When a project expands with respect to either the number of participants in each initiative or the type or use of data, existing DUAs or business associate agreements (BAAs) often need to be altered. Several initiatives noted that drafting these documents broadly at the outset, such as to cover both administrative and clinical data, made subsequent project expansion and evolution much easier. The Pediatric Health Information System Plus (**PHIS**+) initiative found success with this strategy; the PHIS+ database, which contains both clinical and administrative data, evolved from the pre-existing PHIS database, a comprehensive pediatric data resource containing details of more than 6 million patient cases, all derived from administrative data.

However, the effort it takes to customize these types of documents for specific partners should not be underestimated. We heard from the Washington Heights Initiative Community-Based Comparative Effectiveness Research (WICER) that ensuring partners are engaged and understand governance requirements from the outset can greatly reduce the amount of time to bring new partners and sites on board.

Distributed Versus Centralized Networks

When considering whether or not to make data available for secondary purposes, data-holding entities often express concern about losing control of information over which they have both legal and ethical privacy and security obligations. Some

organizations are more comfortable with research arrangements where the patient data are maintained and can only be accessed internally or behind institutional firewalls. In this model, individual patient records are not allowed to flow beyond the firewall, so only aggregate results can be shared externally. This model is known as a distributed, or federated, model, and organizations choosing this approach offer a rationale of being able to better maintain control of information uses and compliance with applicable federal and state law. Most federal and state privacy laws apply only to identifiable data. In a distributed model, because identifiable data remain in the state where the organization is physically located, only that particular state's privacy laws apply (in addition to federal laws).

A second option, known as the centralized model, involves sending individual patient records to a centralized database, which can then be used to support research. Adherents to this approach indicate that it allows greater ease of analysis because the data are all in one place; consequently, the ability to achieve consistency of analysis is improved as well. The data are housed centrally on behalf of participants from multiple states; consequently, the central database may need to comply with multiple state rules, in addition to federal rules, that govern the disclosure of identifiable data for research purposes. This is also an important consideration if cloud computing is used, since the physical placement of the data and backups could be distributed in geographically distant data centers.[15]

Participants using each model cite the advantages of their chosen approach, but they also acknowledge the drawbacks, both from a policy and technical perspective, that have proved frustrating to many.

Distributed Networks
A number of project participants using distributed network models cited "consistency of analyses across sites" as an issue to be resolved; distributed networks by design require that analyses be performed numerous times and in numerous places. One approach to improve consistency was employed in the SCAlable National Network for Effectiveness Research (SCANNER) initiative, which aims to develop a secure, scalable distributed infrastructure that facilitates collaborative comparative effectiveness research (CER), where each participating institution retains patient data due to the individual site rules.

To effectively aggregate data, the site-specific variables must be harmonized before analytics can be performed on each institution's data set. Many analytical methods can be deconstructed such that analyses performed locally at each institution can be combined, generating results that are essentially identical to those that would be derived from a centralized data set.

To achieve its goal of consistent analysis across the sites, SCANNER developed a "virtual machine" that is exported to each site. The virtual machine includes the programs and operating systems necessary to conduct data analysis, and thus enables identical local analyses at each site. Each site partner approves the installation of

Table 3. Legal Requirements Applicable to Research

- Under the Health Insurance Portability and Accountability Act (HIPAA), use or disclosure of identifiable health information for research purposes usually requires specific authorization from the individual, unless the research meets certain exceptions or the requirement for authorization is waived by an institutional review board (IRB) or privacy board.
- Under the Common Rule, the patient's authorization also is required if the researcher is receiving and using identifiable information, as is IRB approval, with some exceptions and qualifications.
- Under HIPAA, an IRB or privacy board can waive the requirement for authorization where (1) the research raises minimal risk to privacy, (2) the research could not practicably be done without a waiver, and (3) the research could not practicably be conducted without the identifiable health information.
- Under the Common Rule, authorization can be waived by the IRB in similar circumstances: (1) the research involves no more than minimal risk, (2) the waiver will not adversely affect the rights and welfare of the patients, (3) the research could not practicably be carried out without the waiver, and (4) when appropriate, the subjects are provided with additional pertinent information after participation.

HIPAA has two categories of data sets that can be used for research purposes and are subject to less—or no—regulation:

1. A limited data set may be used by covered entities, or contractors (business associates) acting on their behalf, for health care operations, research and public health purposes, subject to a data use agreement.
2. Data that qualify as "de-identified" under the HIPAA Privacy Rule are not subject to further regulation, and can be accessed, used, or disclosed for any purpose.

this virtual machine behind its firewalls to analyze the data (which can require numerous levels of approval) to ensure that the results of the analyses do not compromise the privacy of the records.

Leads of the Mini-Sentinel project took a similar approach by developing a common data model (the Mini-Sentinel Common Data Model, or MSCDM) and requiring the participating sites to use it. The MSCDM is a data structure that standardizes administrative and clinical information across data partners and makes it possible to execute standardized data analyses consistently.[16] Investigators from the Mini-Sentinel project noted that use of a distributed approach eliminated "an astonishing number of barriers," including those related to participant willingness to share data, retention of individual partner control, and legal obstacles.

Project participants also noted drawbacks to distributed networks. For example, it takes extensive resources for centers to participate in and maintain a number of different distributed networks. With such a structure, institutions must actually conduct the data analytics themselves, rather than merely contributing data to a centralized system. Another issue is the time and money it takes to actually manage multiple sites, which can act almost as a "tax" on network participants.

Centralized Networks
Other initiatives have employed techniques for managing data access in centralized models. For example, in the VA Informatics and Computing Infrastructure (VINCI), data are accessible

through a centralized secure enclave behind a firewall. Access requests to this data are made using the Data Access Request Tracker (DART) application, which coordinates the requests through VA officers. Through permissions based on the DART approval process, VINCI controls the data such that only authorized users have access, and only for specific research projects under an active IRB protocol. This practice is designed to prevent researchers from accessing data for one project and then reusing data for multiple other projects without appropriate review and approval.

The High Value Health Care Collaborative (HVHC) also employs a centralized database that currently is part of a limited data set; members send encrypted, coded patient identifiers and provider identifiers in crosswalk tables, but they retain the codes needed to "un-encrypt" at their own sites. The database is set to expand as clinical data from the Centers for Medicare & Medicaid Services (CMS) are pulled into the initiative as part of a new grant from the CMS Innovation Center. In order to join member data with CMS data, identifiers will be sent via crosswalk tables; once the data are joined, identifiers will be stored at separate sites. Members will not be allowed to download patient-level data to their individual sites; all analysis will be conducted on the centralized database.

Migrating from One Model to Another
Migrating from one architectural model to another can also prove challenging. For example, the project leads at Cincinnati Children's Hospital Medical Center (CCHMC) were tasked with "enhancing" the registry used by ImproveCareNow, a QI and research collaborative of 52 centers focused on improving the care and outcomes of children with pediatric inflammatory bowel disease. To accomplish this enhancement, the CCHMC project leads aimed to migrate from a centralized registry to a distributed (or at least partially distributed) one. Migrating an existing registry that was established for QI purposes meant that registry data would be collected directly in the EHR and then pulled into a local version of the registry database. When data were needed for reporting purposes, each center would be queried using distributed queries, returning the relevant results.

The CCHMC project leads admitted that the migration from this centralized model has proved difficult. The existing registry used a centralized data management group to calculate and report the various monthly QI measures. The procedures used in these calculations were not run in an automated fashion and would need to be modified as the network launched new QI activities. Moving to a distributed model without a having a dedicated data manager at each site has posed a number of challenges, including how to distribute updated calculations or modifications to existing reports.

Currently, the project is taking a two-pronged approach to dealing with these issues. First, the full registry is being implemented in a "centrally distributed" fashion, meaning that it is using a distributed architecture but is housed entirely at CCHMC. This allows

the project team to work through the logistics of automating the measure calculations without affecting grant timelines or overly taxing resources at the various partner sites. Second, the initiative has created a pilot network of six geographically dispersed centers that would fully implement a distributed report for a couple of key outcome and process measures. This will allow the network to gauge how successful the distributed model ultimately is.

Project Expansion and/or Evolution
A common theme that arises in discussions of governance challenges with respect to data sharing is that experience with one data type, one data use, or even one particular set of existing partnerships does not necessarily translate into instant comfort or trust with another. As a result, even when projects evolve from existing studies, networks, or relationships, the evolution is not always seamless.

Expanding from Administrative to Clinical Data
An example of a project expansion that posed some challenges is the PHIS+ initiative.[17] The PHIS+ enhanced database involves the addition of select clinical information. Although this significantly enhances the utility and scope of possible research, it triggers not only new work flow issues but also new privacy challenges. Clinical data tend to be more sensitive than administrative data because they include more specific patient details (e.g., the results of a particular test, rather than the mere fact that the test was performed). Thus, in the event of a security breach, the potential impact on a patient may be greater than an unauthorized release of claims or billing data.

As a result, although the PHIS+ project leads did not see any additional security risk to including clinical data and had not foreseen any concern related to the expansion, the participating hospitals were hesitant to participate in the enhanced database. The institutions had developed a level of comfort with sending administrative data. But the sharing of clinical data required the approval of the medical faculty, who were less familiar with the history of administrative data sharing and the safeguards that had been deployed to protect the data; these "new partners" had heightened concerns about sharing potentially more sensitive patient information. The PHIS+ leads were able to address this concern by assembling a presentation and a document for distribution that detailed the precise security procedures in place to ensure data privacy and security. The materials were well received and effectively alleviated the worries of the medical staff, ultimately resulting in widespread buy-in to the expanded scope of data sharing.

Expanding from QI to Research Use
A shift in focus or emphasis (e.g., from QI to research) can also pose obstacles—particularly when such a move involves project staff new to research. Such has been the experience of the Surgical Care and Outcomes Assessment Program CER and Translation Network (SCOAP CERTAIN).

The express purpose of the CERTAIN grant was to expand an existing QI registry, known as the SCOAP registry. Registries are collections of data related to patients with a specific diagnosis, condition, or procedure. SCOAP is a registry of patients undergoing surgical and interventional procedures. CERTAIN, a partner to SCOAP, works with clinicians and hospitals to optimize health care delivery through both QI initiatives and CER activities.

In launching CERTAIN, the project leads engaged in a formal recruitment process, reaching out to clinicians, chief medical officers, and chief information officers at institutions that were members of SCOAP. Those institutions that signed on to the CERTAIN DUA did so largely because of existing working relationships developed through the SCOAP project. Not only did these clinicians and health care administrators have familiarity with the CERTAIN project leads, but they also had developed a substantial level of trust in the team based on their successful QI data sharing.[18]

As project implementation progressed, the CERTAIN team found that collaborating in new ways posed unanticipated challenges. With activities classified as research requiring further levels of institutional approval, including IRB approval, many of the clinicians and health care administrators were unfamiliar with additional requirements or how to navigate institutional processes. The CERTAIN leads had simply not anticipated that a relationship involving data exchange for one purpose would not translate seamlessly when the data were exchanged for a different purpose.

Data "Overprotectiveness"

Some project participants we interviewed noted that clinicians, hospital administrators, or researchers frequently are reluctant to share data. This reluctance is at least in part based on their legal and ethical obligations to protect patient data; however, competitive pressures among researchers (e.g., a desire to receive credit for a particular finding or scientific development or to maintain sole access to or familiarity with a valuable data resource) also contribute to that reluctance. In research collaborations among multiple institutions, concerns about possible use of data for competitive market advantage can be a factor. Project participants reported in addition that smaller entities with less experience were more likely to initially view the invitation to join a multisite initiative with skepticism, particularly if the initiative involved participants that were larger in size and had more resources.

These issues can be overcome when participants have a history of working together, as was the case with CERTAIN and SCOAP, but such pre-existing relationships do not always exist, and even if they do, they may not preclude these issues. Further, delivery systems invest millions of dollars to build the systems to generate data and can be reluctant to "share" beyond trusted partners and in the absence of a thorough set of governance agreements that provide resolution for these issues.

With respect to reluctance to share data, WICER leaders noted that such concerns escalated after passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act as part of the American Recovery and Reinvestment Act in 2009.[19] The HITECH Act imposed new civil monetary penalties for unauthorized uses and disclosures on entities covered by HIPAA, and as a result some entities grew more cautious about sharing data for research purposes. Though liability for misuse of data has always been present in HIPAA, project participants noted that decisions to share data that previously may have been made informally were now requiring approval at the highest institutional level, with extensive conversations among privacy boards, privacy officers, and lawyers.

The ImproveCareNow enhanced registry involved migration from a QI registry to one that could be used for both research and clinical support, in the form of tools for pre-visit planning and population management. In the course of this expansion, the initiative began to exchange PHI. This led to concerns from the clinicians that their patients might be targeted for fundraising by the network or that centers with better outcomes might try to do targeted outreach to patients from those centers that were not performing as well. To address this issue, the network drafted a "frequently asked questions" (FAQ) document that explained the changes and the rationale behind them. This information was sent to the network and was followed by a series of webinars, where clinicians were encouraged to raise any questions that remained unresolved. These questions, and their responses, were compiled and added to the FAQ.

To head off its own data-sharing concerns when HVHC was expanding its network to a larger number of participants, it conducted individual site visits to promote the steps it was taking to ensure patient privacy and data security. HVHC contracted with an outside vendor and attorney to lead these visits, outlining HVHC's data trust environment and mapping new trust policies to the relevant regulations. Thus, potential participants were aware from the outset that patient privacy was paramount and an explicit focus of the initiative, which served to increase participant confidence.

Differing Priorities or Understanding of Technologists and Clinicians

A number of the project participants discussed the tension, miscommunication, or both that can arise among the technologists who design the data-sharing network and develop its exchange capabilities, the researchers who design the study or studies, and the clinicians who collect the data. According to the project participants, technology and clinical experts do not necessarily share the same views on the sensitivity of patient data.

Clinicians and researchers have legal obligations to protect patient privacy and have professional reputations to uphold; consequently, they are frequently cautious when it comes to data sharing. In

contrast, technologists take pride in their expertise in security safeguards and often have more confidence in the data security environment they helped to create; as a result, they tend to be more enthusiastic about data sharing. When clinicians are asked to cede control over privacy and security decisions that they may nevertheless be held responsible for, an understandable tension can develop, especially in the absence of trust established over a history of working together or previous collaborations without security incidents.

Further, because data-sharing networks frequently are not designed by the actual users (clinicians), there can be a resulting tension between design and functionality. As noted by the leads of the VINCI initiative, close communication between those who understand the functional requirements of a data-sharing research network and those responsible for the technical implementation is key, whether conducted through supervision of contracted vendors or collaboration among clinicians and technologists.

A number of project leads, including those at WICER, noted that, in hindsight, the chief medical officers at the various sites involved in their initiatives may have been more appropriate collaborators than the chief information officers. On the other hand, the CERTAIN initiative and the registry team at CCHMC found the opposite to be true; the extensive work they did to build relationships and achieve buy-in from clinicians was certainly important, but it occasionally proved challenging to obtain the same support from the technologists.

More broadly, the CERTAIN project leads noted that there are competing priorities at any health care entity, and achieving the support of one group does not necessarily translate into institution-wide support. Regardless of clinician backing, if administrators do not give high priority to a particular research or QI endeavor, the project may lack the necessary funding or staff support. Similarly, if the institution's technologists are busy with competing priorities, such as the implementation of medical record systems to comply with HITECH financial incentives, clinician support alone will not enable a fully functional and operational network.

The PHIS+ initiative found that forming an oversight committee that included both technologists and clinical researchers helped bridge the gaps between design, implementation, and usability. Regular meetings of this group combined the knowledge of the two project branches and provided translation when necessary. Meetings also served a trouble-shooting purpose, helping to assure that issues were continuously identified, discussed, and solved early, rather than after they posed a consequential problem.

Project participants generally agreed that issues related to differing priorities or understanding can boil down to one of leadership, and whether someone in a position of authority is willing to devote his or her time and influence to making sure the launch of an initiative is put high on the institution's priority list. Two participants also mentioned the potential benefit of prioritizing

the creation of a system and process for launching data-sharing initiatives so that efforts similar to theirs would, in the future, require less time and fewer resources.

Challenges Related to IRBs

Organizations typically rely heavily on IRB review as a gatekeeper to reuses of data, in order to comply with legal requirements and institutional policies designed to ensure ethical uses of patient data. All participants expressed at least some frustration in dealing with IRBs. For example, in multisite research studies it is possible that authorization could be required for use of information from one institution but not required for use of information from another, because of different or varying interpretations of legal requirements and institutional policy. Researchers seeking to conduct research at multiple institutions have also noted the difficulty of achieving approval from multiple IRBs, each of which has its own procedures, timelines, and standards of review.

Project participants agreed that it is common to have numerous individual IRBs involved in a multisite research or QI initiative, particularly in academic settings. How the project leadership and individual sites forge relationships with each other and their respective IRBs, as well as the details of the particular studies involved, can help ameliorate potential challenges. Some common issues, as well as some potential approaches to address them, are described below.

Local IRBs, IRB of Record, or Both

Because multisite research initiatives often involve numerous IRBs, many choose to designate an "IRB of Record" (sometimes called a "parent" or "master" IRB) that will take the lead and is responsible for overall approval of an initiative. When a primary IRB is designated, individual partners can choose either to rely on the IRB of Record or to use their own IRBs; a localized IRB structure instead means that each site relies on its own IRB. Different projects can therefore end up with various combinations of centralized and/or local IRB reliance, particularly in the case of review of collaborative research.

For example, specific research questions initiated and conducted by individual partner institutions on a centralized database might be reviewed by that institution's local IRB, although at times an IRB of Record will be used for this purpose as well, depending on whether the dataset is identified or de-identified.

The scenario in the preceding paragraph is the case for the PHIS+ database, the CCHMC enhanced registry, and the VINCI project. In the case of VINCI, IRBs are not standardized across sites—some VA facilities have stand-alone IRBs, whereas others are affiliated with university partners. It is recommended that multisite studies be reviewed by the Central IRB, governed by the Veterans Health Administration's Office of Research and Development, and requests for each specific data use be submitted and reviewed as needed.

HVHC navigates the IRB approval process by sending IRB protocols to its individual members 30 days in advance of when an application is reviewed by the IRB of Record (Dartmouth). This allows the local IRBs to review and make any suggested changes or have questions answered before the protocol gets its final, definitive review. This process was outlined in the project's Master Collaborative Agreement, which states that all collaborative studies are performed under the authority of the Dartmouth College IRB, to which all members agreed to defer.

The SCANNER initiative relies on individual site IRB approval, although for particular studies this can be accomplished in a collaborative fashion: one site's IRB initiates a chain of approval, and then subsequent sites' IRBs follow suit. Similarly, the CERTAIN initiative, which currently relies on a combination of review by an IRB of Record (at the University of Washington) and local IRB review, is in the process of developing a collaborative IRB agreement that would allow individual IRBs to approve a project, but with some consistency of approval process and criteria. COMET also involves an IRB master protocol and template consent that each site uses; this is especially important since these documents have had three amendments as the CER study progressed.

Even with an IRB of Record, individual institutional variation can cause problems or delays. As noted by several projects, when one site's IRB takes more time with its review, for whatever reason, the entire project can be affected by creating a delay in the project meeting its milestones.

Initiatives exchanging data for secondary purposes are sometimes able to facilitate the process of participating institutions meeting their individual compliance needs. In an effort to avoid future IRB challenges related to regulatory audits, staffing issues, lack of previous relationships among some IRBs and lack of experience of others, and increasingly long periods of time necessary for study approval, the CERTAIN project leaders have held a series of meetings to gain a better understanding of what was causing these issues and delays. The meetings have involved CERTAIN's scientific director and various IRB chairs of participating hospitals, and the formation of an IRB Symposium has been planned for the fall of 2013.

QI Versus Research Designation
As noted earlier, designation of a project as involving either QI or research has major implications for the conduct of the studies and the level of involvement of an IRB, because the two are treated differently under both the HIPAA Privacy Rule and the federal Common Rule. This difference in treatment can be exacerbated when a number of independent IRBs review the same collaborative grant, because the interpretations of what category of data use is involved vary widely—particularly if the network is already an established QI network or a research network. IRBs may also designate something as a research project but allow for a waiver exempting researchers from obtaining patient authorization.[20]

In the experience of CERTAIN, despite the previous trust established through the operation of SCOAP, challenges arose when trying to distinguish what electronic data would be used for QI activities and what would be allowed to be used for research or "learning" purposes. This led to a nine-month discussion process across the awardee institution's (the University of Washington) Quality Safety and Research Oversight Workgroup, the Clinical Data Research Oversight Committee and its investigators, and the IRB to clearly define the project process. Ultimately a "wall" between identifiable data and de-identified data was constructed, and it was agreed that only de-identified, aggregate data would be acceptable for research. The decision led to a Memorandum of Understanding that was signed by four institutions and ultimately accepted by other participating institutions.

The CCHMC enhanced registry project encountered this issue as well. Although the registry began as a QI project, the IRBs of the participating centers had a range of interpretations on how to characterize the project, from requiring consent from all participants to finding that the efforts of the network did not qualify as human subjects research.

A number of project participants shared the view that larger institutions, especially those with a great deal of experience with data-sharing projects, are more comfortable and confident designating a particular project as "quality improvement" and therefore not requiring patient authorization. This results from a level of ease on the part of the clinical team, combined with confidence on the part of the legal team, based on deep familiarity with and experience in making the distinction. In contrast, project participants agreed that smaller organizations that have less familiarity with the regulations and little experience interpreting them tend to feel more comfortable taking a conservative approach, assuming that nearly all reuses of clinical data are research, requiring IRB review and patient authorization (unless waived by the IRB).

This "better safe than sorry" policy also provides legal coverage for institutions when a project changes from one focused on QI to one that also involves the conduct of research; in such a case, newly relevant IRB requirements have already been satisfied and authorization has either been obtained or waived. Consequently, this conservative approach feels prudent for a number of research initiatives. The HVHC team, for example, obtained agreement for a centralized IRB and revisits specific projects as its work expands. Though elements of their work could be characterized confidently as QI, HVHC also aims to contribute to generalizable knowledge and therefore chose to obtain IRB approval for all aspects of their project.

Two project participants noted the challenges posed by the "for generalizable knowledge" prong of the legal definition of research, which people translate as requiring IRB approval for conducting research that could result in publication of study results. As such, the benefits of soliciting IRB review (or exemption, as the case may be) at the outset go beyond advance planning for future research uses. In particular, IRB review preserves the option of publishing

study results. The HVHC IRB found a minimally disruptive way to handle this, creating a separate category of review for the QI protocols so they could meet professional and academic journal requirements.

Patient Authorization

When data initially collected for QI purposes—and thus used without first obtaining patient consent or authorization—later become relevant for a research study, challenges related to patient authorization can arise. As mentioned earlier in this paper, the ImproveCareNow enhanced registry encountered this issue. The CCHMC's IRB (the project's IRB of Record) determined that, while the project was initially established for QI purposes and there is some latitude for the reuse of QI data for research, this enhanced registry "felt different" than other QI registries expanded to research uses (like PHIS or national joint registries).

Because the clinicians in ImproveCareNow had a long-standing relationship with their patients, seeing them every three to six months, it was not impractical to obtain patient consent. They considered whether it was necessary to go back to all patients whose data had already been collected in the registry to obtain authorization for research purposes. To help them make this determination, the CCHMC's IRB developed a "decision tree," or flow chart, that included such factors as the relevant IRB decision with respect to the initial registry (i.e., whether, as described above, authorization was required by an institutional IRB for the initial QI use of the data), and whether data are allowed to leave the institution. Applying this decision tree, the project leads and the IRB of Record determined that going back to obtain consent would in fact be necessary.[21]

Fortunately, CCHMC found that most of its partner centers did not need to seek the additional authorization from study participants, because the centers had done so when setting up the initial registry. This demonstrates a potential advantage to treating all (or most) reuses as research even in circumstances where such treatment is not required. HIPAA research authorizations are required to be study specific, which may make it hard to adapt an authorization given for QI purposes into one for research. However, recent changes to the regulations, which went into effect on March 26, 2013, allow for authorizations that cover future research uses, as long as they describe the purpose of the authorization so that the individual would reasonably expect that his or her PHI could be used or disclosed for future research.[22]

Because the transactional costs of seeking authorization can be steep, CCHMC developed an e-consent tool as a means of reducing the burden, allowing patients to provide authorization online. The tool is Web based and can be configured for individual projects, and it allows individual centers to establish their own processes for obtaining patient authorization. Some will send a letter home with patients explaining the study and consent process and providing a link to the e-consent tool. Others will choose to send patients (or their parents) an email or will have them provide e-consent via a tablet device in a clinical setting. This allows individual site IRBs to determine their own procedures while still complying with the requirement to obtain patient authorization.[23] The e-consent tool has been used by a number of studies since it went live in early 2013; centers in ImproveCareNow that are relying on CCHMC as their IRB of Record started using the tool in June 2013.[24]

Factors Contributing to Success

Although all large-scale data-sharing projects experience occasional delays or pitfalls, the project participants identified a number of successful strategies for avoiding or ameliorating some of the more difficult issues described above. The most promising strategies contributing to success are listed below and then discussed in the following sections:

1. Leveraging pre-existing relationships
2. Starting small, then expanding
3. Developing governance documents with individual site input
4. Exchanging de-identified data
5. Establishing governance bodies.

Leveraging Pre-existing Relationships

Perhaps the single unifying theme revealed by the nine site interviews was that successful data-sharing initiatives tend to be based on pre-existing relationships. They evolved out of networks previously built for other purposes, piggybacked on established data-sharing arrangements, and required only modification of existing agreements to encompass new data types and permitted uses. Often the participants asked to join a collaboration are chosen in part because of long-standing or pre-existing professional relationships, shared views regarding data sharing and governance, and high levels of performance in completing research projects successfully and on time. This was true for both the COMET and SCANNER initiatives.

As experienced by the CERTAIN initiative, early collaborations provide a foundation for further collective work, because the participants have a track record of working well together and have gone through the process of coming to agreement on the elements of a working relationship. According to project participants, this foundation contributes to a willingness (even an eagerness) to contribute resources to a research initiative, to rely on a central or external IRB, to move forward in some cases without formal governance arrangements, and to execute common agreements with little negotiation of terms—or even to forgo formal participation agreements altogether in some cases.

In addition to shared values or philosophies, we were told that it can be helpful if collaborating organizations have broadly shared missions or occupy similar positions in their fields. For example, the founding members of HVHC shared a belief that they had a responsibility to be examples and leaders in health and delivery reform, as well as sharing a philosophical commitment to innovation and continuous QI.

Starting Small, Then Expanding

Expanding a project beyond the initial, relationship-based network can mean adding additional members, adding additional data types, changing the purpose of data use, or launching an entirely new study. When more sophisticated projects are built from small or simple pilots, there is already tangible proof that collaboration is possible. The smaller projects pave the way for larger initiatives, having demonstrated that data sharing across multidisciplinary, multisite teams can be done in a way that both protects privacy and security and achieves mutually beneficial results.

Many of the multisite initiatives we interviewed found success by originally launching with a handful of trusted institutions and expanding when a founding member invited a new site participant. HVHC employed such a strategy, building out their initiative slowly, from the four founders to the current 15 members. They felt this allowed their mission to remain central, as new participants were required to adhere to participation conditions that were mission driven and had proved to be successful.

Project leaders of the HVHC and PHIS+ database expansions found it was relatively easy to progress from the exchange and use of billing or administrative data to clinical data. Although clinical data are potentially more sensitive, the institutions had established a track record of exchanging less sensitive data that paved the way to moving to the exchange of more sensitive information.

Developing Governance Documents with Individual Site Input

Most multisite research or QI initiatives have in place some combination of BAAs, DUAs, master collaborative or participation agreements, and privacy policies. The development of these governing documents has the potential to prove contentious, because there are a number of individual institutional goals, norms, policies, and procedures to reconcile. Those projects that reported having engaged in lengthy and close collaboration with all member sites prior to finalizing any contract or policy all touted the effectiveness of their legal or governance documents.

HVHC found that involving each of its individual sites in the development of its DUA was a key element in expanding its initial pilot project to include clinical as well as billing data, and to add 11 new members. HVHC developed a new Master Collaborative Agreement, a process that took nine months and numerous visits to the participant sites; these investments in time and money were essential to the agreement's development and adoption, and they paid off.

The final master collaborative agreement is a straightforward legal contract that governs data use, the path of exchange, and access. HVHC found that its careful front-end work made the founding members more comfortable with bringing in new members as the project evolves and expands, given how familiar with and confident they are about the terms to which new members are agreeing.

The team at the COMET initiative picked its initial members based on long-standing relationships, and it too worked with each site partner to develop its DUA. Although the project leads emphasized that this process can be onerous because each site's lawyers must play a role, the project leads of all initiatives that have gone through the process of getting each participant's buy-in to the terms, policies, and procedures say it eliminates downstream issues. They also affirm that it promotes trust through transparency and collaboration.

The COMET team noted that it is important to identify the right point person at each institution with whom to work directly when attempting an endeavor as consequential and labor-intensive as drafting a multi-site DUA. Given the potential for constant revisions (both big and small), the smaller, more efficient, and "more appropriate" the team, the more smoothly the process proceeds and the lighter the burden. The COMET team learned that the optimal path for their study was to draft the multisite DUA at the parent site (Stanford) through meetings with the principal investigator, the study data manager, the contract/compliance officer, and the technology and licensing attorney, and then to have each site tailor the DUA to the site's requirements through discussions among its lawyer, principal investigator, and study coordinator.

Exchanging De-identified Data

Privacy risks decrease when data are accessed and shared in ways that significantly reduce the possibility of re-identifying particular patients. Data that meet HIPAA de-identification standards are not subject to regulation, and nonidentifiable data are not subject to the Common Rule. Many project participants noted that relying on the use of de-identified data for a study eliminates up front a number of the previously discussed governance challenges.

HVHC employs a centralized database model and exchanges only de-identified data, as described earlier in this paper. Members send to each other encrypted, coded patient and provider identifiers in crosswalk tables, and the codes are retained at the individual sites. The WICER project also involves a centralized database containing de-identified data. WICER aims to gather longitudinal information about patients across institutions in one community, but this is accomplished in a privacy-preserving fashion by deleting the actual identifiers once a link has been established. The PHIS+ project, too, exchanges only de-identified data.

The Mini-Sentinel project is authorized to exchange fully identifiable data because it has been deemed by the HHS Office for Human Research Protections to be a public health project, rather than a research initiative.[25] However, the project's stated goal from the outset was to minimize the use of PHI and exchange as little of it as possible. To accomplish this, project leads have employed a strategy that involves an anonymous linkage of data to their identifiers, or a "one-way hash."[26] Although Mini-Sentinel is testing such a practice, the project leads have yet to make a decision about whether to adopt it as a standard practice.

Establishing Governance Bodies

Perhaps because so many successful multisite initiatives are built on pre-existing relationships and often pre-existing data-exchange networks, establishment of formal governance bodies happens less frequently (or is less burdensome) than one might presume. However, a number of the initiatives interviewed for this paper have developed governing structures or bodies that have improved the ease of and comfort with data sharing.

ImproveCareNow has an executive committee and board of directors, as well as a separate leadership group overseeing the implementation of the enhanced registry and execution of the grant objectives. In addition, ImproveCareNow has a patient advisory council and parent advisory council, both of which have been particularly meaningful to the network; the initiative has found that engagement of patients and families is key to the successful management of chronic conditions such as inflammatory bowel disease.

The Mini-Sentinel project leads credited its success to several factors related to its governance structure. First, early attention was paid to governance by broadly representative committees. Further, Mini-Sentinel created a privacy committee comprised of privacy experts who were not affiliated with any of the participating organizations.[14] Second, there was ongoing engagement and approval of new policies by a planning board that includes representation from all participating organizations, in addition to a patient representative. Finally, project leads distributed opportunities for substantive work among affiliated investigators, including data partners, who choose to lead or participate in work groups formed on a project-by-project basis and thus feel thoroughly invested in the initiative.

HVHC's Master Collaborative Agreement specifies an organizational structure that includes an executive committee, and within that committee a steering committee with representation from each participating institution, ensuring transparency and equal participation.

Regardless of the precise elements of any governing structure, having bodies or committees in place that are responsible for oversight, accountability, trouble-shooting, and translation between involved parties can help either avoid or resolve challenges that arise in the course of launching or implementing a data-sharing project.

Conclusion

Leveraging clinical data for secondary purposes is key to meeting goals for health care reform (i.e., improving health care quality and reducing costs). Drawing valid, meaningful conclusions from health data will in most cases require that data be gathered and analyzed across care settings (e.g., hospitals and clinics) and communities that may be geographically dispersed. Unfortunately, the governance challenges associated with sharing data collected from EHR can be daunting and difficult to resolve.

Current rules governing reuses of data present an important set of guardrails that help ensure the ethical and responsible use of health data, but differences in interpretation and varying tolerances for risk can serve as obstacles to multisite research collaborations. This paper has explored the experiences of nine multisite comparative effectiveness or health services research initiatives— their collaborations, the challenges they faced, and how they have been able to either manage or overcome those challenges and move forward with their research.

The hope is that this paper and the toolkit to which it will contribute will be instructive and helpful to others seeking to be a contributing part of the learning health care system by exploring and resolving data governance issues.

Acknowledgements

References
1. Strategic plan, fiscal years 2010-2015. Washington (DC): U.S. Department of Health & Human Services; 2011 Sept. 124 p.
2. Kaiser Family Foundation. Federal support for health information technology in Medicaid: key provisions in the American Recovery and Reinvestment Act [Internet]. 2009 Aug [cited 2013 Sept. 13]. 8 p. Available from: http://kff.org/medicaid/issue-brief/federal-support-for-health-information-technology-in-medicaid-key-provisions-in-the-american-recovery-and-reinvestment-act.
3. Data stewardship is broadly defined as the responsibility or accountability to ensure wise and appropriate collection, storage, and/or use of data derived from individual personal health information. Data governance is the process by which responsibilities of stewardship are conceptualized and carried out.
4. Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule; Nass S, Levit LA, Gostin LO, editors. Beyond the HIPAA Privacy Rule: enhancing privacy, improving health through research [Internet]. Washington (DC): National Academies Press; 2009. Available from: http://www.ncbi.nlm.nih.gov/books/NBK9578.
5. Department of Health and Human Services (US). Human subjects research protections: enhancing protections for research subjects and reducing burden, delay, and ambiguity for investigators. Proposed rules. Fed Regist. 2011 July 26; 76(143): 44512-44531.
6. Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. 104-191 (Aug. 21,1996).
7. Standards for privacy of individually identifiable health information. 45 C.F.R. Sect. 46.101 (2010).

8. "Health care operations" are certain administrative, financial, legal, and quality improvement activities of a covered entity that are necessary to run its business and to support the core functions of treatment and payment. These activities are limited to the activities listed in the definition of "health care operations" at 45 C.F.R. Sect. 164.501 (2010).

9. Standards for privacy of individually identifiable health information. 45 C.F.R. Sect. 164.502 (2010).

10. Standards for privacy of individually identifiable health information. 45 C.F.R. Sect. 164.502 (2010).

11. See note 13. We note that U.S. Food and Drug Administration (FDA) regulations conform to the Common Rule to the extent permitted by statute, but the FDA has its own rules governing human subjects research that include a different definition of "research." See Lee BM. Science and research: comparison of FDA and HHS human subject protection regulations. Silver Spring (MD): Food and Drug Administration; 2009 Mar 10 [cited 2013 Sept 13]. Available from: http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/educationalmaterials/ucm112910.htm.

12. Standards for privacy of individually identifiable health information. 45 C.F.R. Sect. 46.101 (2010).

13. Protection of human subjects. 45 C.F.R. Sect. 46.116(c-d) (2010); 45 C.F.R. Sect. 164.512(i)(2)(ii) (2010). See also McGraw D, Leiter A. Legal and policy challenges to secondary uses of information from electronic clinical health records [Internet]. Washington (DC): AcademyHealth; 2012 [cited 2013 Sept. 13]. Available from: http://www.academyhealth.org/files/publications/HIT4AKLegalandPolicy.pdf

14. Rosenbaum S. Data governance and stewardship: designing data stewardship entities and advancing data access. Health Serv Res. 2010 Oct;45(5):1442–55.

15. In fact, this is important for disaster recovery. Recent modifications to HIPAA indicate that cloud providers should act as business associates since data are maintained in their systems, although this does not apply to Internet providers acting as a mere conduit to transport the data. See Department of Health and Human Services (US). Modifications to the HIPAA privacy, security, enforcement, and breach notification rules under the health information technology for economic and clinical health act and the genetic information nondiscrimination act; other modifications to the HIPAA rules. Final rule. Fed Regist. 2013 Jan 25;78(17): 5566–5702, 5571.

16. Mini-Sentinel Principles and Policies (August 2012). Available from: http://mini-sentinel.org/work_products/About_Us/Mini-Sentinel-Principles-and-Policies.pdf

17. Narus SP, Srivastava R, Gouripeddi R, Livne OE, Mo P, Bickel JP, et al. Federating clinical data from six pediatric hospitals: process and initial results from the PHIS+ consortium. AMIA Annu Symp Proc 2011; 2011: 994–1003.

18. Those initiatives that declined to join CERTAIN did so based on competing demands for internal information technology resources, primarily due to meaningful use implementation.

19. See note 5.

20. Marsolo K. Approaches to facilitate institutional review board approval of multicenter research studies. Med Care. 2012: July;50(Suppl):S77–81.

21. Patients whose registry data are being used purely for QI still do not need to provide authorization for that use.

22. Department of Health and Human Services (US). Modifications to the HIPAA privacy, security, enforcement, and breach notification rules under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; other modifications to the HIPAA rules. Final rule. Fed Regist. 2013 Jan 25;78(17): 5566-5702.

23. The use of this tool is somewhat more complicated if a particular research project involves minors as the subjects. In general, parents will be given the e-consent link and will provide the necessary authorization on behalf of their children. Depending on the age range of the study participants and individual site policies, sometimes patient consent (or assent) is involved as well; if the patient is 18 years or older, parental consent is not sought, because the minor has the capacity to authorize the research uses.

24. Although an e-consent tool has numerous efficiency benefits, it also requires development of processes to manage identity and authentication of patients (and parents consenting on their behalf). While federal law may authorize release of data for nonresearch purposes without consent, some entities have institution-specific consent requirements to share data that are either organizational policy or state law.

25. The Mini-Sentinel may not be a typical example, because it was created pursuant to a congressional mandate to build a postmarketing safety surveillance system—a classic activity of public health agencies. Its formal designation as "public health activity" by the Office for Human Research Protections eliminated a number of common governance obstacles.

26. A one-way hash is an algorithm that turns messages or text into a fixed string of digits, usually for security or data management purposes. The "one way" means that it is nearly impossible to derive the original text from the string. If one applies the algorithm to a set of full identifiers (e.g., name, address, social security number), one gets a unique string that has no intrinsic meaning. "One-way" means one cannot practically reverse the process. But if two parties apply the same hashing algorithm to the same name, address, etc., both parties will get the same unique string. The two parties can then exchange these strings to find the matches, without disclosing the identities of any of the individuals who are represented by these strings.

27. One of the authors of this paper, Deven McGraw, served on this privacy committee.