

## Integrating a Federated Healthcare Data Query Platform With Electronic IRB Information Systems

Shan He MS<sup>1</sup>, John F. Hurdle MD PhD<sup>1</sup>, Jeffrey R. Botkin MD MPH<sup>2</sup>, Scott P. Narus PhD<sup>1</sup>

<sup>1</sup> Department of Biomedical Informatics, <sup>2</sup> Division of Medical Ethics and Humanities,  
University of Utah, Salt Lake City, UT

### Abstract

*Human subjects are indispensable for clinical and translational research. Federal and local agencies issue regulations governing the conduct of research involving human subjects in order to properly protect study participants. Institutional Review Boards (IRBs) have the authority to review human subject research to ensure concordance with these regulations. One of the primary goals of the IRB oversight is to protect research participants' privacy by carefully reviewing the data used and disclosed during a study. However, there are major challenges for IRBs in the typical research process. Due to the information disconnect between the data providers (e.g., a clinical data warehouse) and the IRB, it is often impossible to tell exactly what data has been disclosed to investigators. This causes time-consuming, inefficient, and often ineffective monitoring of clinical studies. This paper proposes an integrated architecture that interconnects a federated healthcare data query platform with an electronic IRB system.*

### Background

Clinical research is aimed at investigating new approaches to the treatment and prevention of human disease, and it plays an important role in improving the quality of healthcare. Translational research has proven to be a powerful process that drives the clinical research engine (1). Breakthroughs in basic biomedical sciences have provided an unprecedented supply of information for improving human health. Translating the information gained through these basic discoveries into knowledge improving clinical practice and human health requires clinical research involving human subjects (2).

Human subject protection is an obligation essential to the clinical and translational research endeavor, much of which is governed by such rules as Department of Health and Human Services' (HHS) Federal Policy for the Protection of Human Subjects (also known as the "Common Rule") (3); the Food and Drug Administration's (FDA) Protection of Human Subjects Regulations (4); and the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (also known as Standards for Privacy of Individually Identifiable Health Information) (5).

Although the Privacy Rule does not apply to researchers unless they are themselves covered entities or workforce members of covered entities, the researchers may be indirectly affected by the privacy rule if covered entities supply their data (6). In the course of conducting research, researchers may create, use, and/or disclose individually identifiable health information known as protected health information (PHI). The Privacy Rule requires covered entities to obtain appropriate documentation from researchers before disclosing PHI to them, and to scrutinize researchers' requests for access to health information more closely (7). The National Institutes of Health (NIH) also identifies clinical research ethics as a key function of a Clinical and Translational Science Research center (8).

The primary mechanism ensuring appropriate protection of human subjects is review by an Institutional Review Board (IRB) and/or a Privacy Board.<sup>1</sup> The federal regulations delegate authority to IRBs, which are committees that have been formally designated to approve, monitor, and review any research involving humans (3). IRBs not only approve projects, but they also monitor ongoing research to assure that all applicable safeguards are continually followed (9). The Common Rule requires an IRB to conduct continuing review of research at intervals (but at least annually) and to observe the consent process when necessary.

### Problem Statement

IRBs confront major challenges during oversight of human subjects protection in clinical and translational research. One of the challenges is the bureaucratic and time-consuming review procedure, especially when the volume and complexity of current research is taken into account. Though most researchers agree that a human subject protection system is an essential safeguard in clinical research (10), many are frustrated by the burdensome paperwork and long waiting time for IRB approval.

---

<sup>1</sup> For the sake of simplicity, we will use the term IRB in this paper to represent the review body which could also be a Privacy Board when referring to its authority to approve a waiver or an alteration of the Privacy Rule's Authorization requirement.

Researchers frequently view IRB review as a barrier to be overcome rather than as a constructive process that minimizes risks and enhances safety (11).

Another challenge is the difficulty of adequate continuing review to ensure investigators comply with their approved protocols. Institutions vary in their efforts and abilities to monitor investigator compliance, from those that have no monitoring programs to those that conduct random audits (11).

Finally, there is a growing trend towards multi-site biomedical research, particularly emphasized by NIH's Clinical and Translational Science Award program, which offers numerous scientific advantages over single-site studies (12). There are federated approaches which provide informatics infrastructure such as the cancer Biomedical Informatics Grid<sup>®</sup> (caBIG<sup>®</sup>) (13) and the Informatics for Integrating Biology and the Bedside (i2b2) project (14) to facilitate research across multiple institutions. However, the process of obtaining ethics review to address privacy issues for a multi-site study can be a daunting task, consuming time, money, and energy (15).

We will illustrate the research oversight problem using a specific clinical research example, which could be either a prospective or retrospective study. To conduct a prospective study, a common approach taken by investigators is to first query against certain data source(s) to find potential participants who meet some inclusion criteria, followed by a contact to invite enrollment. In a retrospective study, the investigator mainly deals with historical data and there are confidentiality risks such as inadvertent disclosure of sensitive information. Therefore, both types of studies need to link with data providers, the IRBs, and the participants or their data. We will illustrate the challenges mentioned above from the point of view of the investigator, the data provider, and the IRB.

Before conducting research, an investigator will submit an application to his/her local IRB (or multiple applications for a multi-site study). A critical part of the study protocol is the inclusion and exclusion criteria, including such factors as conditions, age, gender, ethnicity, etc., which are used to select the desired cohort. One of us [JFH] served as IRB Chair and has observed that sometimes study inclusion and exclusion criteria are not explicit enough for the IRB to make a decision, a challenge we address below. After the investigator receives IRB approval, subjects meeting the inclusion and exclusion criteria are found either by direct

recruitment (e.g., posting flyers) or, increasingly, by submitting a data request to a clinical data source administrator. The investigator also needs to submit the IRB approval documentation to the data provider. This process is more complex when the study involves multiple sites because duplicate IRB applications and data requests are needed.

When the data administrator receives a request from an investigator, he/she has to ensure the investigator received appropriate approval. According to federal regulations, the data administrator needs to manually review the approval document before disclosing any data to the investigator. However, it is often difficult, if not impossible, to determine the exact data that was approved by the IRB for inclusion in the research protocol. There is often a mismatch of understanding. It is not uncommon that the data provider agrees to whatever data the researcher says he/she needs and assumes that the IRB approval covers the request.

In addition, despite the fact that many institutions have implemented electronic IRB systems that are trying to streamline and automate the oversight process, the IRB only makes the decision to approve or deny an application based on the study protocol and supplemental documentation submitted by a researcher. It has no control over the actual data access for the study because the data access procedure is separate from the IRB review system. In such cases, it is difficult and time-consuming for an IRB to do the ongoing monitoring of the study that is required by the Common Rule.

The IRB *should* contact the data provider(s) to get information (which may not be available) concerning whether the investigator extracted only the data that is needed and approved for the study. Usually, the IRB just sends out a request to the investigators requiring them to renew the application after a certain period of time and update the application whenever the study protocol is changed. Obviously, fulfilling the data review requirement is completely dependent upon the investigators, which means breaches of regulations can easily happen.

Although there are emerging platforms and tools to assist investigators with access to, and management of, data for clinical research (e.g., caGrid which provides investigators with shared data resources from multiple institutions, and caBIG Clinical Trial Suite which provides comprehensive management of clinical trials and associated data), the inherent complexities in the effort to protect participant health information privacy and to achieve regulation

compliance are still considered to be a barrier to efficient clinical research. This paper proposes an integrated system which interconnects a federated healthcare data query platform with an electronic IRB system to streamline the clinical research process, speed research initiation, and increase efficiency and patient privacy.

### System proposal

Researchers in health-related disciplines require access to many sources of health information, from archived medical records and epidemiological databases to disease registries, tissue repositories and hospital discharge records (2). Researchers may require data drawn from multiple sites in order to garner the advantages provided by a multi-site study (12). In such a context, a federated data repository can virtually integrate disparate data sources. It provides researchers with a unified query interface that returns normalized, federated results joining the query results from each data source. The Federated Utah Research and Translational Health e-Repository (FURTHeR), hosted in the Center for Clinical and Translational Science (CCTS) at the University of Utah is an example of such a system. Its overarching aim is to create for Utah an informatics platform that federates Utah's largest patient data warehouses (University of Utah Healthcare, Intermountain Healthcare, and the Salt Lake City Veterans Administration Medical Center); public health data from the State of Utah Department of Health; and data from the Utah Population Database (an extensive genealogic and demographic resource) (16).

A critical component of FURTHeR (or any other federated health data environment) is the security model that protects patient health information and limits accessibility to authorized entities to ensure privacy. A federated security model is complex. Various mechanisms need to be implemented to achieve the security goals such as confidentiality, integrity and availability. In a federated data repository and query environment like FURTHeR, the safeguard mechanisms include, but are not limited to, message encryption so that people who may intercept messages can not read them, digital signature to make sure health information is not modified; federated authentication to authenticate users from different institutions without creating a central user registry; and federated authorization.

Federated authorization is the most complex component because users from different institutions have different roles where each role has different defined privileges. During the process of our work on the federated authorization model for FURTHeR, we

realized that the authorization issue is so complex that simple role-based access control is not adequate for the federated query service to make the right authorization decision. As noted above, the IRB's decision on a study plays an important role in determining whether to authorize the investigator of the study to access certain data elements. This paper will focus on healthcare data access control that restricts researchers to precisely the data they need, and to do so in a way to make that access explicit to IRB reviewers. A secondary benefit is that data resource personnel know precisely what data has been authorized, as well.

To better describe the general system architecture, we use FURTHeR as an example for the federated data repository and query platform. The Web portal, MyRA, is the user interface to FURTHeR for researchers, patients, and providers. We will mainly focus on its query functionality as oriented to researchers. The ERICA system, which is the University of Utah's electronic IRB system, is the example IRB system to be integrated with the federated query platform. The integrated system architecture is shown in Figure 1.

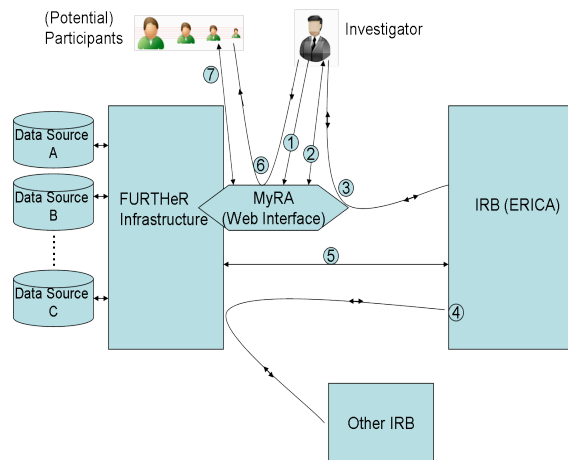


Figure 1. Integrated System Architecture

The process flow for the proposed system is as follows:

- 1) The researcher would login to MyRA. MyRA uses a federated authentication mechanism to uniquely identify the researcher. After logging in, the researcher can view all his/her studies submitted to the IRB in MyRA.
- 2) After agreeing with certain data use contracts, the researcher can query the aggregated number of potential subjects meeting selection criteria of a pre-research study. MyRA will return the result without requiring an IRB exemption, but it will

- log these pre-research queries and provided them to ERICA for IRB review if needed. (This “pre-authorization” for pre-research queries assumes prior agreement from the data providers with the data access methodology and data use contracts.)
- 3) The researcher can initiate an IRB application based on the initial query in step 2). The inclusion and exclusion criteria and the data sources that will be accessed in the study protocol will be automatically generated from the user query constructed in MyRA. The user will be directed to ERICA to enter other required information. This can save time and resources spent on document preparation for review boards and speed the commencement of research activities.
  - 4) The IRB(s) will make a decision on whether to approve or deny or request modification of the study protocol through a regulation-compliant review process. In order for FURThER to make strict and accurate access control based on the IRB review result, the IRB decision about which data elements can be accessed by a researcher should be as structured and as specific as necessary. This can be implemented by adapting the current IRB system (ERICA) to accept more structured application information instead of free-text documents, especially in data access description sections. In addition, for multi-site studies, although previous research suggests using a centralized IRB which oversees multiple institutions (15), until this organizational change takes place, FURThER could act as a central exchange point for local IRBs so that they can exchange review decisions with each other to reduce duplicate reviews and applications.
  - 5) If the researcher wants to query health information beyond the pre-research phase, he/she can construct a query or recall a previous query and indicate an associated IRB approval in ERICA. MyRA will send a query to ERICA about the indicated IRB approval status. If no study information can be found, or the study is not approved or already expired, MyRA will return a notification to the user that IRB approval is required. If data elements requested in the subject query are beyond those approved by the IRB, then only those elements permitted are displayed to the user. In this way, individual data providers are relieved from the burden to review and approve the data access.
  - 6) If the IRB approval allows subject contact, the researcher can send contact letters and consent forms to individuals who meet the study selection criteria directly through MyRA if the subjects’ email information is available. A

reference number is generated for each subject who receives such an email. The individuals who are interested in the study can log in to MyRA using the reference number and sign the consent form electronically. This also provides the IRB the ability to monitor the study’s consent process.

- 7) Research subjects can use MyRA to view all the studies in which they participate and the accounting of disclosure of their information through the auditing features of FURThER.

### System Implementation

The message communication between the federated healthcare data query platform and electronic IRB systems will use the standard clinical study information content currently being developed by HL7 Regulated Clinical Research Information Management (RCRIM)(17) technical committee to increase interoperability between systems. Standard web services interface for querying protocol information from individual IRB systems and interoperation among IRB systems is defined but implementation of the web services is dependent upon the local IRB platforms. This leverages the legacy IRB systems and ensures good scalability via standard communication among systems.

### Discussion

The federated query platform needs to be integrated with multiple electronic IRB systems to support the authorization requirement of users from different institutions that may have their own IRBs. Not every institution has an electronic IRB system. Not every electronic IRB system can be easily adapted to integrate with the federated query platform. How to make this integration effort as easy and as flexible as possible is a critical task to make the ideas proposed here practical and realizable.

The authorization decision on a researcher’s request of health information access relies on the IRB’s review result. However, this can not be the only safeguard mechanism to ensure information security and patient privacy. Other security steps such as federated authentication, message encryption, digital signature, role-based access control and patient consent-based access control should be implemented together with IRB system integration.

This system could be integrated with clinical trial management systems so that more detailed clinical study information can be viewed through the user interface of the federated query platform. The clinical trial data and reports can be linked to participants’ medical records and contributed to the original data source.

In addition, an obstacle for study investigators is the inability to recruit enough participants. One reason behind this is public concern over the risk of the study because of distrust of the safeguard mechanisms to ensure privacy (18). The federated query platform can provide the general public with an access interface to make regulatory information of studies accessible to potential participants. Therefore, public trust in clinical and translational research would be enhanced and study enrollment is expected to be improved.

### Conclusion

IRBs need to monitor investigators conducting their research studies to enhance investigator compliance with regulations for human subject protection. But the monitoring approach should not place more burdens on the IRBs, since they are already facing critical workload and efficiency issues. The proposed integration between a federated healthcare data query platform and electronic IRB systems will perform a stricter and more accurate access control on investigators when they are trying to access participants' health data. It also streamlines the clinical research process and increases efficiency by automating information exchange among different parties involved in the research process. Finally, ethics review for multi-site studies would be made easier and more efficient by utilizing the federated query platform as a hub to allow IRBs to exchange information with each other.

### Acknowledgements

This investigation was supported by Public Health Services research grant UL1-RR025764 from the National Center for Research Resources, and by the National Library of Medicine grant 1RC2LM010798 from the National Institutes of Health.

### References

1. Re-engineering the Clinical Research Enterprise: Translational Research. NIH Common Fund; [updated November 5, 2009; cited 2010 20th, February].
2. Sung N, Crowley WF J, Genel M, Salber P, Sandy L, Sherwood L, et al. Central Challenges Facing the National Clinical Research Enterprise. *JAMA*. 2003;289(10):1278-87.
3. Title 45 of the Code of Federal Regulations, Part 46, Subpart A.
4. Title 21 of the Code of Federal Regulations, Part 50.
5. Title 45 Code of Federal Regulations Part 160 and Subparts A and E of Part 164.
6. Clinical Research and the HIPAA Privacy Rule. National Institute of Health; 2004 [updated February cited 2010 February 8th]; Available from: [http://privacyruleandresearch.nih.gov/clin\\_research.asp](http://privacyruleandresearch.nih.gov/clin_research.asp).
7. Research Repositories, Databases, and the HIPAA Privacy Rule. National Institute of Health; 2004 [cited 2010 February 8th]; Available from: [http://privacyruleandresearch.nih.gov/research\\_repositories.asp](http://privacyruleandresearch.nih.gov/research_repositories.asp).
8. Institutional Clinical and Translational Science Award (U54). National Institutes of Health (NIH); 2007; Available from: <http://grants.nih.gov/grants/guide/rfa-files/rfa-rm-08-002.html>.
9. Bauchner H. Protecting Research Participants. *PEDIATRICS*. 2002;110(2):402-4.
10. Whitney SN, Alcer K, Schneider CE, B L, McCullough, McGuire AL, et al. Principal Investigator Views of the IRB System. *International Journal of Medical Science*. 2008(5):68-72.
11. Ethical and Policy Issues in Research Involving Human Participants.: National Bioethics Advisory Commission2001.
12. Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, Giobbie-Hurder A, et al. Multisite Randomized Controlled Trials in Health Services Research: Scientific Challenges and Operational Issues. *Medical Care*. 2001;39(6):627-34.
13. Cancer Biomedical Informatics Grid (caBIG®). National Cancer Institute; Available from: <https://cabig.nci.nih.gov/>.
14. Informatics for Integrating Biology and the Bedside (i2b2). Available from: <https://www.i2b2.org/>.
15. Gold JL, Dewa CS. Institutional Review Boards and Multisite Studies in Health Services Research: Is There a Better Way? *Health Serv Res*. 2005;40(1):291-308.
16. Bradshaw RL, Matney S, Livne OE, Bray BE, Mitchell JA, Narus SP. Architecture of a Federated Query Engine for Heterogeneous Resources. *Proceedings of the 2009 AMIA Annual Symposium*. 2009.
17. HL7. Regulated Clinical Research Information Management. Available from: <http://www.hl7.org/Special/committees/rcrim/index.cfm>.
18. Roan S. Medical clinical research slows for lack of patients. *Los Angeles Times*. 2009 March 14.