Biobanking Informatics Infrastructure to Support Clinical and Translational Research

Bernie LaSalle¹, Michael Varner², Jeff Botkin³, Marc Jackson⁴, Louisa Stark³, Melissa Cessna⁵, Carolyn Orthner¹, Nathan Hulse⁴, Aldo Bernasconi¹, Randy Madsen¹, Dustin Schultz¹, Richard Bradshaw¹, Joyce Mitchell¹

¹ Department of Biomedical Informatics, ² Department of Obstetrics and Gynecology, ³ Department of Human Genetics, ⁷ Nursing Informatics
University of Utah, Salt Lake City, UT 84112

⁴ Maternal Fetal Medicine, ⁵ Intermountain Biorepository, ⁶ Clinical Knowledge Management
Intermountain Healthcare, Salt Lake City, UT 84111

Abstract
The University of Utah Health Sciences (UUhSC) and Intermountain Healthcare (IH) support high value clinical and translational research programs. The Utah BioHealth Initiative will facilitate next generation research by leveraging the combined resources of both institutions through an infrastructure which links biospecimens and electronic health records (EHR).

During phase I of the Utah BioHealth Initiative (UBI) the participating institutions developed a legal, regulatory and information technology infrastructure that supports clinical and translational research, and advances our understanding of health and disease, improves healthcare value and health for current and future generations of Utahns. We used the Federated Utah Research and Translational Health electronic Repository (FURTHeR)¹ to combine EHR and biospecimen data from an actual study populated by both institutions to demonstrate the robustness of the infrastructure.

Introduction
Medical research and practice are being transformed by the development of increasingly complex and affordable analytical technologies²,³ such as microarrays and next generation sequencing. These advances are expected to create a revolution in medical care by introducing information from a wealth of molecular markers into the process of making informed medical decisions about risk, prevention, diagnosis, prognosis and treatment.

These technologies and their application to Personalized Healthcare have brought increased attention to the importance of human biospecimens⁴. For example, the development of next-generation sequencing technologies and protocols has made it possible to extract and sequence DNA from formaldehyde-fixed paraffin-embedded (FFPE) samples⁵. This development, which would have been unthinkable just a few years ago, has turned existing FFPEs - routinely collected by Pathology departments and clinical care institutions - into valuable resources for molecular research. Increasingly, research and clinical care institutions are realizing the value of their existing and future human biospecimen collections and the need to develop sustainable policies and procedures for their acquisition and management. However, the management of human biospecimens presents a number of legal, regulatory, technical and analytic challenges that span each sample's entire lifetime, including collection, handling, processing, storage, analysis and reporting.

Intermountain Healthcare (IH) and the University of Utah Health Sciences (UUt) have successfully collaborated to create a clinical data infrastructure that researchers from both organizations can use to analyze phenotypic and biospecimen data across institutions. The new infrastructure implements a real-time federation of two datasets, one managed and maintained at Intermountain, and one managed and maintained at the University of Utah. Data and service communications are managed over a direct and secure VPN channel between the two organizations.
Pilot Project
In order to demonstrate the ability of FURTHeR to support real time queries between institutions (EHRs, biospecimens), an Adverse Pregnancy Outcomes (APO) research project was selected as the pilot project. This drove requirements for the infrastructure, which supports both phenotypic and biospecimen data for the APO study, including subject demographics, diagnoses, laboratory test results, medications, illicit drug use, pregnancy, labor, and infant-related observations. Patient/subject phenotype data were linked to DNA specimens and/or serum specimen samples enabling researchers to find specimens based on specific patient phenotypes.

The APO project involved identifying women with pre-eclampsia and controls (pregnant women without signs or symptoms of pre-eclampsia) from Intermountain Healthcare and the University of Utah Health Sciences. Clinical and biospecimen data were extracted for consented study participants and placed in a separate repository for analysis.

The FURTHeR 1.2 system (Figure) is a logical representation of the query process. A query is sent to the Federated Query Engine (FQE) from the user interface (UI). The FQE is then responsible for distributing the query to all listening or selected (from the UI) data sources. The figure shows 4 participating data sources: APO Boutique Database at Intermountain Healthcare, University of Utah Enterprise Data Warehouse (UUEDW), APO Clinical and Biospecimen database, and the Utah Population Database Limited (UPDBL).

Utilizing data source adapters, data sources, which can answer the query, adapt the logical query into a native query that the corresponding data source can understand. Data sources differ in how they operate with some using Structured Query Language (SQL) and others using web services. Each data source’s results are then sent back to the FQE where they are translated to a common data model and aggregated together. The common data model is similar to the Observational Medical Outcomes Partnership (OMOP)6,7 data model with extensions to include biospecimen data, federated data sources and multiple data sets.
The APO pilot project proved to be an effective strategy for driving the project to a completion point that provides useful functionality. The clinical data models that were created were APO project-specific and the patient population (~176 pregnant women with preeclampsia and matching non-preeclamptic pregnant controls) was obtained from two delivery services (University of Utah Health Sciences Center and Intermountain Medical Center – an Intermountain Healthcare facility) who average in total ~10,000 deliveries per year and who both maintain ongoing perinatal research data and sample registries.

**Methods**
The cross-institutional UBI team evaluated and selected software for the management of biospecimen resources. The program caTissue was chosen for specimen tracking at both the University of Utah and Intermountain Healthcare. Once the biospecimen data management software was installed and configured, the UBI tested the new environment by importing a complex repository (DNA samples from the Phenotype Core – part of the University of Utah’s Center for Clinical and Translational Science) into caTissue. This biorepository stores 6000+ samples, covering more than 20 diseases. Following this successful test biospecimens from participants of the APO study were identified both participating institutions, matched to the corresponding phenotype data using medical record numbers. The UU data were then de-identified and linked (phenotype-biospecimen) using a de-identified unique identifier. A parallel biospecimen data management instance was created at IH and linked to phenotype data from the study participants.

In order to link the parallel UU and IH UBI data projects at the participating institutions several technologies and practices had to be created or modified. Implementation of the FURTHeR Query Language (FQL) at Intermountain Healthcare required that a query translation, data service and transport service be put in place. In order to achieve transmission of a secure query and payload, a static virtual private network (VPN) connection was implemented. This network level encryption allows for tight security of connectivity. This foundational network architecture also allows for multiple methods of connection via the secure network layer. As such, a transport service has been implemented at Intermountain Healthcare to proxy request and response traffic over the VPN connection.

Additionally, many architectural enhancements for security were implemented to support the authorization of secure query and response messages. Authorized messages were vetted against access control lists (ACLs) maintained at both institutions for security authorization integrity. In order to restrict access to those named on the participating IRB projects, these ACLs were manually maintained for the scope of this project. We found that a robust federated security model was needed to support identity exchange between the participating institutions for role-based authorization.

The FURTHeR query interface was modified to support the joining of data through iterative queries, which allowed patient demographic, clinical and biospecimen data to be joined into one result set from the participating sources. Restrictions for this process were implemented to follow data access requirements as stipulated by each data source. These stipulations outline which data sources can be joined and were implemented via business logic in the filter chain of the FQE of FURTHeR.

Biospecimen data (a dimension of patient data) were added to each source, as a first step for joining patient and biospecimen information. In order to better support biospecimen data in future queries, a biospecimen common model will need to be integrated in the FURTHeR logical model. As this model matures, independent data services to biospecimen management systems can be implemented to support more complex queries and joining of patient clinical and biospecimen data via FURTHeR.

**Summary**
Managing biospecimens requires not only documentation of the nature, quantity and quality of the samples, but also curated information about the participants who donated them, including clinical and family history, genealogical and demographic information, and consent tracking. It also requires maintenance of links to associated data, including medical and genealogical records (e.g. Utah Population Database (UPDB)), results from molecular analyses (e.g. DNA sequence and expression), and results and conclusions from clinical studies.
Creating tools to analyze and share biospecimen and clinical (phenotype) data across institutions proved to be feasible from a technical standpoint. Despite employing standard HIPAA-compliant techniques to de-identify all data from each data source and utilizing best-of-breed security measures, the most significant challenges encountered were within the legal and compliance offices of both organizations.

One of the accomplishments of the UBI was the development of a Memorandum of Understanding between the University of Utah and Intermountain Healthcare, that acknowledges the value of this partnership and provides a formal understanding for continued development of the UBI. The development of a functional Utah Biohealth Initiative, particularly when combined with the FURTHeR project and the joint Center for Clinical and Translational Sciences will allow translational and discovery researchers from both Institutions to be synergistically more productive.

References

1. Federated querying architecture with clinical & translational health IT application. Livne OE, Schultz ND, Narus SP. J Med Syst. 2011 Oct;35(5):1211-24. Epub 2011 May 3
2. The Transformation to Value-based, Personalized healthcare – Implications for Translational and Clinical Research. Davis, BJ, Sheldon, J. iHealth Connections, 2011;1(1):24–7
3. Forging a pathway to quality cancer care. Farina K. Am J Manag Care. 2012 May;18(3 Spec No.):SP116-8
4. Biobanking: the foundation of personalized medicine. Hewitt RE. Curr Opin Oncol. 2011 Jan;23(1):112-9. Review.
5. Making the most of pathological specimens: molecular diagnosis in formalin-fixed, paraffin embedded tissue.
6. Ryan PB, Griffin D, Reich C, et al. OMOP common data model (CDM) Specifications, 2009. http://omop.fnih.org/CDMandTerminologies (accessed 28 Nov 2009).
7. Reisinger SJ, Ryan PB, O’Hara DJ, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. J Am Med Inform Assoc2010;17:652–62.
8. http://biospecimens.cancer.gov/meeting/brnsymposium/2010/docs/Watson%20ОбР%20Ta lk%202010.pdf