The University of Kansas Medical Center's Healthcare Enterprise Repository for Ontological Narration (HERON): an i2b2-based Platform for Clinical, Translational, and Informatics Research

Lemuel R. Waitman, Arvinder S. Choudhary, Daniel W. Connolly
Division of Medical Informatics, Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS

Introduction

Motivated by a desire to be competitive for an National Institutes of Health Clinical and Translational Science Award (CTSA), the University of Kansas Medical Center (KUMC) created a Division of Medical Informatics within the Department of Biostatistics in February 2010. A cornerstone for informatics was to develop a strategy for reusing clinical data. This required inter-organizational consensus as the University is separate from the covered entities which maintain clinical systems: University of Kansas Hospital (KUH) and the University of Kansas Physicians, Inc. (UKP)

HERON in Context

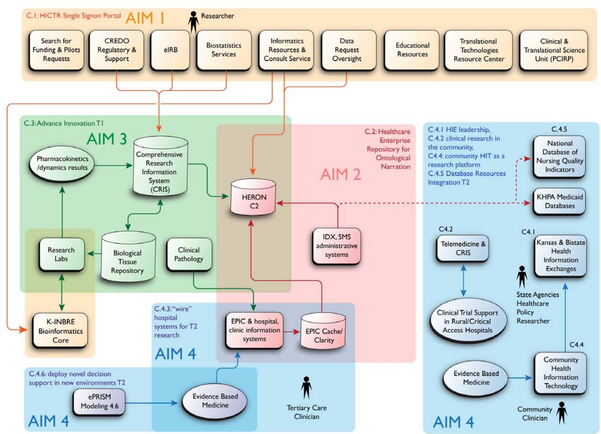
The Informatics CTSA aims also provide a base for biomedical informatics research and services for other university strategic objectives (NCI Cancer Center designation, Institute for Advancing Medical Innovation, developing a School of Public Health). There are four specific aims:

1.Provide a HICTR portal for investigators to access clinical and translational research resources, track usage and outcomes, and provide informatics consultative services.

2.Create a platform, HERON (Healthcare Enterprise Repository for Ontological Narration), to integrate clinical and biomedical data for translational research.

3.Advance medical innovation by linking biological tissues to clinical phenotype and the pharmacokinetic and pharmacodynamic data generated by research cores in phase I and II clinical trials (addressing T1 translational research).

4.Leverage an active, engaged statewide telemedicine and Health Information Exchange (HIE) effort to enable community based translational research (addressing T2 translational research).



Governance to allow "fishing"

In May 2010, i2b2 was operational for a small registry of research participants. In September 2010, the hospital, clinics and university signed a master data sharing agreement to create the repository.

- The HERON executive committee is composed of senior leadership (e.g. chief information, operating, financial, executive officers and chief of staff) from the hospital, clinic and medical center and provides governance for institutional data sharing.
- Establishing business processes and servicing research requests is conducted by the Data Request Oversight Committee (DROC) which reports to the HERON executive committee.
- The repository's construction, oversight process, system access agreement, and data use agreement for investigators were approved by the KUMC Institutional Review Board.

Since HERON is currently funded by KUMC and contains medical center and affiliated clinics and hospital data, access requires a medical center investigator. As additional institutions and health care organizations provide support and contribute data, they will be incorporated with multi-institutional oversight provided by the HERON executive committee and DROC.

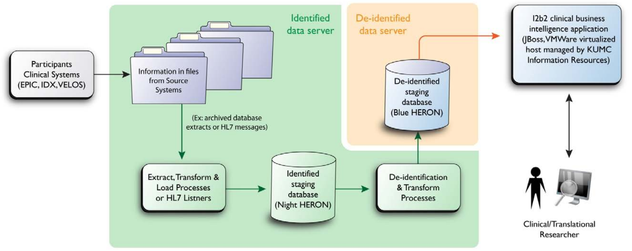
Initial Four Use Cases:

1. After signing a system access agreement and automated verification of human subjects research training, cohort identification queries and view-only access is allowed but logged and audited. Qualified Faculty may also sponsor non-faculty for view-only access.
2. Requests for de-identified patient data, while not deemed human subjects research, are reviewed.
3. Identified data requests require approval by the Institutional Review Board prior to data request review. Medical informatics will generate the data set for the investigator.
4. Contact information from the HICTR Participant Registry have their study request and contact letters reviewed by the CTSA Participant and Clinical Interactions Resources Program.

Technical Approach

- Separate servers (HP DL180s) for identified data from de-identified in the same data center as clinical organizations systems. Using SUSE Linux (KU enterprise license and support)
- Monthly process. We make a complete rebuild of the repository, not HL7 messaging based. Get new files, bake another copy. When the new copy is ready, switch over i2b2 to point at the new version.
- Export KUH Epic Clarity relational database instead of Cache/MUMPS. KUH Clarity is Oracle based so we went Oracle also because of site license and team expertise.
- Monthly file from UKP clinic billing system (GE IDX) that contains demographics, services, diagnoses (ICD9) and procedures (CPT), and HICTR research participant flag.
- ETL processes largely SQL (some Oracle PL/SQL) wrapped in python scripts.

Architectural Decisions

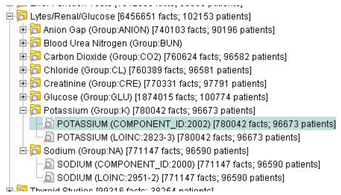


HIPAA Safe Harbor De-identification:
We remove 18 identifiers and date shifting by 365 days back resulting in non-human subjects research data but treated as a limited data set from a system access perspective. System users and data recipients agree to treat as a limited data set (acknowledging re-identification risk)

To be addressed:
•For now, we won't add free text such as progress notes with text scrubbers (DeID, MITRE Identification Scrubber toolkit)
• Currently have "obfuscation" turned on (No sets < 10 and sets randomly perturbed ± 3 patients)
• While de-identified, access to timeline functionality provides individualized patient "signatures". Will likely revisit as organizational familarity grows.

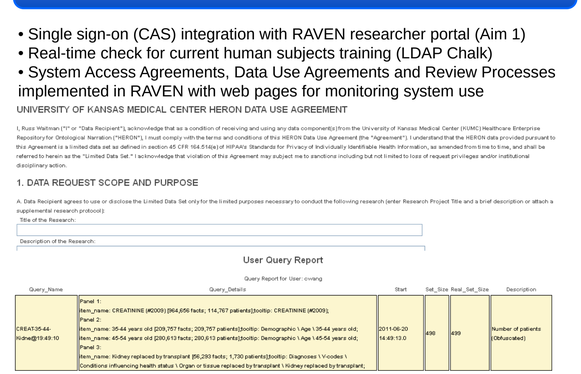
"Lazy" Load supports alternative views of reality

- Load with the local terminology first.
- Map concepts to standards secondarily in the concept space
- Allows multiple ontologies for observations and works around mapping challenges with contributing organizations



Current Functionality

- Single sign-on (CAS) integration with RAVEN researcher portal (Aim 1)
- Real-time check for current human subjects training (LDAP Chalk)
- System Access Agreements, Data Use Agreements and Review Processes implemented in RAVEN with web pages for monitoring system use


UNIVERSITY OF KANSAS MEDICAL CENTER HERON DATA USE AGREEMENT

I, Russ Waitman ("I" or "Data Recipient"), acknowledge that as a condition of receiving and using any data component(s) from the University of Kansas Medical Center (KUMC) Healthcare Enterprise Repository for Ontological Narration ("HERON"), I must comply with the terms and conditions of this HERON Data Use Agreement (the "Agreement"). I understand that the HERON data provided pursuant to this Agreement is a limited data set as defined in section 45 CFR 164.514(e) of HIPAA's Standards for Privacy of Individually Identifiable Health Information, as amended from time to time, and shall be referred to herein as the "Limited Data Set." I acknowledge that violation of this Agreement may subject me to sanctions including but not limited to loss of request privileges and/or institutional disciplinary action.

1. DATA REQUEST SCOPE AND PURPOSE

A. Data Recipient agrees to use or disclose the Limited Data Set only for the limited purposes necessary to conduct the following research (enter Research Project Title and a brief description or attach a supplemental research protocol):

Title of the Research:

Description of the Research:

User Query Report

Query Report for User: cwang

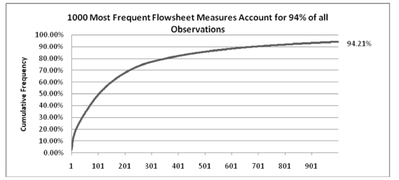
Query_Name	Query_Details	Start	Set_Size	Real_Set_Size	Description
CREAT:35-44-Kidney@19:49:10	Panel 1: Item_name: CREATININE (#2009) [964,656 facts; 114,767 patients][tooltip: CREATININE (#2009)], Panel 2: Item_name: 35-44 years old [209,767 facts; 209,767 patients][tooltip: Demographic \ Age \ 35-44 years old; Item_name: 45-54 years old [280,613 facts; 280,613 patients][tooltip: Demographic \ Age \ 45-54 years old; Panel 3: Item_name: Kidney replaced by transplant [56,293 facts; 1,730 patients][tooltip: Diagnoses \ V-codes \ Conditions influencing health status \ Organ or tissue replaced by transplant \ Kidney replaced by transplant,	2011-06-20 14:49:13.0	490	499	Number of patients [Obfuscated]

- Half a billion facts (calculated at load and updated in ontology)
- Only beta test usage so far
- Research participant flag integrated

HERON System Usage Users

Total Users Count: 16

User_Id	User_Name	Logon_Count	Query_Count
rwaitman	Russ Waitman	350	397
issimpson3	Steven Simpson	2	3
dconnolly	Dan Connolly	264	140
jdenton	Jo Denton	51	22
aallen	Angelica Allen	1	0
achoudhary	Arvinder Choudhary	96	61
dzhu	Dongsheng Zhu	342	305
jwarren2	Judith Warren	12	1
scarlson	Susan Carlson	2	0
cglennon	Cathy Glennon	1	0
lmanos	LaVerne Manos	13	0
mmishra-aff	Meenakshi Mishra	8	6
rbarohn	Richard Barohn	1	0
bbarnes	Brian Barnes	4	13
cwang	Connie Wang	2	1
drobbins	David Robbins	4	14
evidoni	Eric Vidoni	11	79
jburns2	Jeff Burns	9	6



Preliminary research on nursing flowsheets will be presented at AMIA this Fall.

Planned Enhancements

HERON Executive Committee approval June 2011 for incorporating:

- Hospital billing ICD9, MS-DRG, Insurance Status (possibly use data source provided to UHC from Seimens SMS)
- Hospital Tumor registry (Aim 3) and Cerner CoPath pathology system
- University Biospecimen Repository (Aim 3)
- University REDCap and Velos Registries and Clinical Trials systems (Aim 3)
- Social Security Death Master File

Data request fulfillment tool development (possibly a plug-in)

Enhance system scalability with Fusion-IO NAND flash cards
http://www.fusionio.com/products/iodrive-duo/

Explore collaboration with other organizations

- Potential feeds from Medicaid and State Insurance Commission (Aim 4)
- SHRINE and support for other CTSA affiliates in KC area (Children's Mercy, UMKC, KCUMB, KU-Wichita, KU-Lawrence, Cerner)

Acknowledgements

This project has been supported in part by NIH grant UL1 RR033179-0
Follow our progress and download code from our website at:
http://informatics.kumc.edu/work/wiki/HERON
Presented at i2b2 Academic Users Group 1st Annual Conference
June 28, 2011 Boston, MA

KU DEPARTMENT OF BIOSTATISTICS
The University of Kansas