KUMC Medical Informatics Update and Roadmap

Russ Waitman, Arvinder Choudhary, and Dan Connolly
Division of Medical Informatics
Department of Biostatistics
August 31, 2011

This project has been supported in part by NIH grant UL1 RR033179-0
Outline for Today’s Presentation

- Core KUMC CTSA Informatics Aim: HERON
- Current Functionality
- Current Process
  - How We Work
- Upcoming Clinical Data in HERON
  - Supporting both the CTSA and Cancer Center
- Future Functionality and Integration
  - 1.6 release versus current 1.4: visits and modifiers
  - i2b2 <-> REDCap
Background: NIH Goal to Reduce Barriers to Research

- Administrative bottlenecks
- Poor integration of translational resources
- Delay in the completion of clinical studies
- Difficulties in human subject recruitment
- Little investment in methodologic research
- Insufficient bi-directional information flow
- Increasingly complex resources needed
- Inadequate models of human disease
- Reduced financial margins
- Difficulty recruiting, training, mentoring scientists
National Clinical and Translational Science Award (CTSA) Objectives:

The purpose of this initiative is to assist institutions to forge a uniquely transformative, novel, and integrative academic home for Clinical and Translational Science that has the consolidated resources to:

1) captivate, advance, and nurture a cadre of well-trained multi- and inter-disciplinary investigators and research teams;

2) create an incubator for innovative research tools and information technologies; and

3) synergize multi-disciplinary and inter-disciplinary clinical and translational research and researchers to catalyze the application of new knowledge and techniques to clinical practice at the front lines of patient care.
Frontiers Biomedical Informatics Aims

1. Provide a 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).
Aim #2: Create a data “fishing” platform

- Develop business agreements, policies, data use agreements and oversight.

- Implement open source NIH funded (i.e. i2b2) initiatives for accessing data.

- Transform data into information using the NLM UMLS Metathesaurus as our vocabulary source.

- Link clinical data sources to enhance their research utility.
Develop business agreements, policies, data use agreements and oversight.

- September 2010 the hospital, clinics and university signed a master data sharing agreement to create the repository.
  - Executive Committee – decides organization/systems expansion
  - Data Request Oversight Committee – guides implementation and approves/monitors use.

- Use Cases:
  - After signing a system access agreement, cohort identification queries and view-only access is allowed but logged and audited
  - Requests for de-identified patient data, while not deemed human subjects research, are reviewed.
  - Identified data requests require approval by the Institutional Review Board prior to data request review.
  - Contact information from the Frontiers Participant Registry have their study request and contact letters reviewed by the Participant and Clinical Interactions Resources Program
Current Functionality

• Single sign-on (CAS) integration with HERON portal linked off Frontiers home page (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
  • To be refactored…
• Demonstration
  • i2b2 and HERON tools
Current Process

- Architecture
- Feeding HERON: Extract, Transform, Load (ETL)
- Software Development assisted by TRAC
Technical Details

• Separate servers (HP DL180s) for identified and de-identified data.
  • In the same data center as clinical organizations systems.
  • Supported by KUMC Information Resources
• Using SUSE Linux (enterprise license & support)
• KUH Clarity is Oracle-based so we went Oracle and also because of site license and team expertise
• Upgraded storage to use FusionIO Duo 1.28 TB NAND cards
  • Solid state storage with direct PCI bus connection.
  • Impact has been dramatic on query performance and ability to support simultaneous queries.
Extracting, Loading,Transforming Data

• Goal: stable monthly process, minimal downtime
  • Complete rebuild of the repository, not HL7 messaging.
  • Two databases: create new DB while old DB is in use.
  • When the new DB is ready, switch over i2b2 to serve it.

• Initial Files from Clinical Organizations
  • Export KUH Epic Clarity relational database instead of Cache/MUMPS.
  • Monthly file from UKP clinic billing system (GE IDX).
    • Demographics, services, diagnoses, procedures, and Frontiers research participant flag.

• ELT processes largely SQL (some Oracle PL/SQL)
  • Wrapped in python scripts.
HERON De-identification Decisions

- HIPAA Safe Harbor De-identification
  - 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”
Other Key HERON decision

- “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.

Further technical details described at: http://informatics.kumc.edu/work/wiki/HERON
ELT Workload Breakdown

- Engage contributing organizations – Arvinder & Russ
- Analysis – Russ, Arvinder, Dan
- Set up databases - Arvinder
- Extract Source Data to Staging Area – Arvinder
- Transform into i2b2 Night HERON – Dan & Arvinder
- De-id transform to Blue HERON – Dan & Arvinder
- Regularize process – Dan develops, Arvinder manages
- Publish release notes – Dan
  - HERON visits new lake or river
- Whole process fairly “lights out” for several months.
  - Takes ~ 80 hours
Software Development Process

- [http://informatics.kumc.edu](http://informatics.kumc.edu)
- Simple Wiki & Blog for documentation
- Checking in code into Mercurial version control
- View changes in TRAC
- Work with tickets, milestones and roadmap
  - New data added to monthly release milestones
  - Other milestones for functionality
  - Tickets for tasks, enhancements, end user problems & bugs
- Share lessons and experiences with others
- TRAC demo (BSR example)
Data Sources for FY2012: Focus on Supporting Cancer Center Initiative

- HERON Executive Committee approval June 2011 for incorporating:
  - University Biospecimen Repository (Aim 3, Cancer Center)
  - Hospital Tumor registry (Aim 3, Cancer Center)
  - University REDCap and Velos Registries and Clinical Trials systems (Aim 3, Cancer Center)
  - Hospital billing ICD9, MS-DRG, Insurance Status
  - Social Security Death Master File (Aim 4, Cancer Center)
  - Cerner CoPath pathology system (Aim 3, Cancer Center)
- Also continue to extract and refine data from Epic EMR
Developing a Rich Description of our Population: Existing and Planned Data Sources for HERON. Existing sources shown in **bold underlined text** and planned in plain text.
An i2b2 query against HERON for currently supported cancer centric data sources

Any neoplasm ICD9 diagnosis (106,000 patients) and a WBC count (121,000) -> 44,000 distinct patients,
*require height (123,000) and weight (154,000) -> 35,000 patients,
•require Wong-Baker pain scale (84,000) -> 14468 patients,
•Body Temperature (158,000) -> 14463 patients,
•Surgical Pathology Procedures CPT (85,000) -> 12446 patients,

Finally selective serotonin 5-HT3 antagonist antiemetics -> 8517 patients

With our improved hardware (Fusionio memory cards), the cohort size is returned in 15 seconds for this 8 group query.
CTSA Aim #3: Link biological tissues to clinical phenotype and our research cores’ results

- Support Cancer Center, IAMI, and bridge to Lawrence Research
- **First focus:** Incorporate clinical pathology and biological tissue repositories with HERON and CRIS to improve cohort identification, clinical trial accrual, and improved clinical trial characterization
  - Aligned with existing enterprise objectives to improve biological tissue repository information systems
  - Clinical trial accrual identified by many as a weak point institutionally
  - Target both biological research specimens and routine clinical pathology
KUH Tumor Registry

- Validated Outcomes and Observations
  - Tumors, Nodes, Metastasis (TNM) on complete cases
  - Untapped investment: 7 cancer registrars (Tim Metcalf)
  - ~65,000 cases, data since 1950s

- North American Association of Central Cancer Registries (NAACCR) file format
  - Will build on work at other NCI designated i2b2 users
    (Group Health Cooperative in Seattle, Kimmel Cancer Center in Philadelphia have shared their code/metadata with us)
  - John Keighley providing invaluable expertise

- Later, supplement with additional treatment information not in NAACCR file
Adding Social Security Death Master File

- Have Death status on approximately 90 million people.
  - Contains Social Security Number, Name, Date of Birth, Date of Death, Place of Death
  - Monthly update file from ntis; will sync with releases

- Initial thoughts: may allow the user to pick how good a match by having separate ontological categories
  - SSN only
  - SSN plus DOB
  - SSN plus Name
  - SSN plus Name and DOB
Future Functionality: IRB and i2b2 1.6

- Moving beyond counting to line item data review
  - In August, Karen Blackwell Privacy officials agreed to allow timeline access under current system access agreement
  - Need to address some technical issues before release feature

- i2b2 version 1.5: DataMart Request Form to facilitate our Data Use Agreement

- i2b2 version 1.5: Visit enabled queries

- i2b2 Modifiers with i2b2 version 1.6
  - Will have to redo ELT to take advantage
Data re-identification risk and sensitivity for different data access possibilities with HERON/i2b2

We’ve committed to offer this manually with a DUA and IRB approval; Currently provide contact info with HICTR participant committee approval

Current Practice when giving researchers Epic access

We’ve committed to offer this with a DUA

Would be nice to offer this with a level of review between a SAA and DUA

We are currently here and require SAA, faculty or sponsorship, and HSC training

Author: Russ Waitman, KUMC
Last modified: July 22, 2011
Example: Prostate Cancer and PSA tests
Data Mart Request Form

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
What do Visits and Modifiers Offer?

- **Visits:**
  - I want to know the patient had the lab and the medication in the same episode of care.
  - Conceptually, i2b2 has had a table for the visit dimension but the software never exploited the data.

- **Modifiers:**
  - Is it a billing diagnosis or from the problem list? Is it a primary or secondary?
  - How to I represent all parts of a medication order (dose, route, frequency)?
Constrain observations to the same visit
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events_slides/i2b2_OpeningTalk_20110628_Murphy.pdf
HERON <-> REDCap Integration

- i2b2: excels at data warehousing, knowledge management, hypothesis exploration
- REDCap: pretty solid tool for storing and collecting research data and it’s very user friendly.
- Goal: if we can integrate the best of both, we will inherit the advancements in each project.
- Use cases:
  - Breast Cancer Registry in REDCap integrated with HERON which holds the biospecimens (REDCap -> HERON)
  - Fulfilling Data Request for Participant Contact Information (HERON -> REDCap)
Breast Cancer Registry

Similar to Tumor Registry: BSR Personnel create forms and enter data to improve annotation for fields that are difficult to automatically extract from Epic and other clinical systems.
Use Case for Managing Patient Contact Information for Frontiers Participant Registry Data Request Committee

Step 1: explore data to see if there are enough subjects (ideally done before they wrote the protocol...)

Step 2: Ask the CTSA Data Request Committee for permission to contact Participants

Step 3: Build Cohort Database for Investigator and DRC Auditing

Step 4: Contact Individuals and Audit: was the patient contacted? Was the information accurate? Did they qualify? Will they participate? Did they remember signing the agreement? Do they want their name removed from the registry?

Author: Russ Waitman August 30, 2011
Questions?