Advancing Clinical and Translational Research with Informatics at the University of Kansas Medical Center

Russ Waitman, PhD
Director of Medical Informatics,
Associate Professor, Department of Biostatistics
Director, Frontiers Biomedical Informatics
Assistant Vice Chancellor, Enterprise Analytics
University of Kansas Medical Center
Kansas City, Kansas

This project is supported in part by NIH grant UL1TR000001.
Outline

- A Fishing Story
- What is Biomedical Informatics?
- What are the Clinical Translational Science Awards?
- Informatics Aims
  - Tools for storing information: REDCap
  - Tool for viewing/getting information: HERON/i2b2
    - Oversight Process
    - Information Architecture Observations
- Questions/Demo
• We have tools and expertise to manage data and convert it into information

• **REDCap** and **CRIS** – enter and manage data

• **HERON** – **fish for data** from the hospital/clinic

• **Biweekly** Frontiers Clinical Informatics **Clinics**
  – Tuesday 4-5 pm in 1028 Dykes Library.
  – Next session July 10.
You’re that fisherman: wanting to land data to answer your research hypothesis

Bennett Spring Trout Park, Lebanon Missouri
http://mdc.mo.govregions/southwest/bennett-spring
The Fish: Diagnoses, Demographics, Observations, Treatments
Why so many fish? Medical Informatics

**Current Goal:** Build Hatchery, Manage the Fishery
Second Goal: If you need help fishing, hire a guide from Medical Informatics

Photo Credit: HuntFishGuide.com
http://www.flickr.com/photos/huntfishguide/5883317106/
Prepare and Analyze data with Biostatistics

Photo Credit: S. Klathill
http://www.flickr.com/photos/sklathill/505464990/
Our shared goal: a tasty publication

Photo Credit: Steve Velo
http://www.flickr.com/photos/juniorvelo/259888572/
Nightmare: looks like a nice river, but can’t catch fish

- I’ll just enter everything in Excel....
- What if I lose or accidentally sort my spreadsheet?
- How to I let students only review de-identified data?
- Prevent the wrong people (statistician/student) from entering/changing data?
- Hospital/Clinic is making me use this Electronic Medical Record and I get nothing in return...

Little White Salmon River, Washington State, last Summer in July
Sometimes, You’re willing to enter data/buy fish:
**REDCap: Research Electronic Data Capture**

- [https://redcap.kumc.edu](https://redcap.kumc.edu)
  - It uses the same username and password as your KUMC email.
    - Non-KUMC researchers can request an affiliate account through Frontiers CTSA office
  - Check out the training materials under videos
  - Case Report Forms and Surveys

- For consultation and to move project to production: Register your project with us so we can keep track of your request.
  - [http://frontiersresearch.org/frontiers/biomedical-informatics](http://frontiersresearch.org/frontiers/biomedical-informatics)

- Check out other institutions using REDCap and possibly borrow from the master library.
  - [http://www.project-redcap.org/](http://www.project-redcap.org/)
## REDCap Case Report Form Example

### University of Kansas Medical Center

#### KU-ADC

### NACC - A1 Subject Demographics

<table>
<thead>
<tr>
<th>Field</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC Subject ID</td>
<td>1</td>
</tr>
<tr>
<td>Enrolled in NACC MDS</td>
<td>Yes, No</td>
</tr>
<tr>
<td>Primary Reason for coming to ADC</td>
<td>Participate in research study, Clinical evaluation, Other, Unknown</td>
</tr>
<tr>
<td>Principal Referral source</td>
<td>Clinician</td>
</tr>
<tr>
<td>Presumed disease Status at enrollment</td>
<td>Control, normal</td>
</tr>
<tr>
<td>Presumed Participation</td>
<td>Initial Evaluation only</td>
</tr>
<tr>
<td>ADC enrollment type</td>
<td></td>
</tr>
<tr>
<td>Subject’s month of birth</td>
<td></td>
</tr>
<tr>
<td>Subject’s year of birth</td>
<td></td>
</tr>
<tr>
<td>Subject’s sex</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Does the subject report being Hispanic/Latino ethnicity, regardless of race?</td>
<td>Yes, No, Unknown</td>
</tr>
<tr>
<td>What does subject report as his/her primary race</td>
<td></td>
</tr>
<tr>
<td>What additional race does subject report?</td>
<td></td>
</tr>
<tr>
<td>What additional race beyond what is indicated above in questions, does subject report?</td>
<td></td>
</tr>
<tr>
<td>Subject’s primary language</td>
<td></td>
</tr>
<tr>
<td>Informant’s years of education</td>
<td></td>
</tr>
<tr>
<td>Subject's Education</td>
<td></td>
</tr>
</tbody>
</table>

---

**Note:** This form is to be completed by intake interviewer per ADC scheduling records, subject interview, medical records, and proxy informant report (as needed). For additional clarification and examples, see UDS Coding Guidebook for Initial Visit Packet, Form A1. Check only one box per question.
REDCap Survey: Think SurveyMonkey
Option Two: CRIS/Velos

REDCap Disclaimer

• For clinical trials, CRIS/Velos may be a better fit
  – Multiple years of experience
  – CRIS team builds for you with biostatistics review
  – Budget for CRIS team and biostatistics explicitly

• “Investigator driven” REDCap only works if you, the Principal Investigator, takes responsibility for your data
  – Scalability: informatics provides consultation and responsibility for technical integrity; **not your dictionary or data entry**.
    • Underwritten by CTSA, but you “feed and talk to your fish”
  – Middle model where informatics can build for you in REDCap.
    • Again, you budget for our team’s time
REDCap: think Fish Tank you manage

http://www.flickr.com/photos/wiccked/185270913/lightbox/
CRIS: Vancouver Aquarium. We feed fancy fish; Biostatistics cooks tasty meal
I want to go fishing, not fill a fish tank (REDCap)
Use HERON: a managed fishery

Bonneville Hatchery: Trout, Salmon, Sturgeon, Columbia River, Oregon
Aim #2: Create a data “fishing” platform: HERON, https://heron.kumc.edu

- **Get a License:** Develop business agreements, policies, data use agreements and oversight.

- **Get a Fishing Rod and Bass Boat:** Implement open source NIH funded (i.e. i2b2 https://www.i2b2.org/) initiatives for accessing data.

- **Know what your catching:** Transform data into information using the NLM UMLS Metathesaurus as our vocabulary source.

- **Stock Different Tasty Fish:** link clinical data sources to enhance their research utility.
• Fill out System Access Agreements to sponsor students/staff
• Fill out Data Use Agreement to request data export
• **No Limit!!! IRB Protocol Not Required** to view or pull de-identified data
• Must be on campus or use VPN
• Check [http://informatics.kumc.edu/work/blog](http://informatics.kumc.edu/work/blog) for latest status
The i2b2 “Fishing Rod”: build Diabetes cohort

Drag concepts from upper left into panels on the right

Types of “fish” in folders
i2b2 : AND in Frontiers Research Registry

Dragging over the second condition
When you add a numeric concept, i2b2 asks if you want to set a constraint.
i2b2 Result: **497 patients in Cohort**

Run the Query

Query took 4 seconds

497 patient in cohort
I2b2: Explore Cohort, **Visualize Timelines**

**Timeline**

Specify Data | View Results | Plugin Help
---|---|---

Person #1774 - f - Non-Hispanic
- BMI (Calculated) [288,218 facts; 80,395 patients]
- Diabetes mellitus [376,016 facts; 31,783 patients]
- Frontiers Research Participant Registry [7,654 facts; 7,654 patients]
- GLUCOSE (2011) [1,007,666 facts; 119,524 patients]
- HEMOGLOBIN A1C (2013) [80,703 facts; 35,269 patients]

Person #1805 - f - Non-Hispanic
- BMI (Calculated) [288,218 facts; 80,395 patients]
- Diabetes mellitus [376,016 facts; 31,783 patients]
- Frontiers Research Participant Registry [7,654 facts; 7,654 patients]
- GLUCOSE (2010) [872,592 facts; 100,192 patients]
- GLUCOSE (2011) [1,007,666 facts; 119,524 patients]
- HEMOGLOBIN A1C (2013) [80,703 facts; 35,269 patients]

Person #21253 - f - Non-Hispanic
- BMI (Calculated) [288,218 facts; 80,395 patients]
- Diabetes mellitus [376,016 facts; 31,783 patients]
- Frontiers Research Participant Registry [7,654 facts; 7,654 patients]
- GLUCOSE (2010) [872,592 facts; 100,192 patients]
- GLUCOSE (2011) [1,007,666 facts; 119,524 patients]
- HEMOGLOBIN A1C (2013) [80,703 facts; 35,269 patients]

Person #13473 - f - Non-Hispanic
- BMI (Calculated) [288,218 facts; 80,395 patients]
- Diabetes mellitus [376,016 facts; 31,783 patients]
- Frontiers Research Participant Registry [7,654 facts; 7,654 patients]
- GLUCOSE (2010) [872,592 facts; 100,192 patients]
- GLUCOSE (2011) [1,007,666 facts; 119,524 patients]
- HEMOGLOBIN A1C (2013) [80,703 facts; 35,269 patients]

Person #3138 - m - Non-Hispanic
- BMI (Calculated) [200,216 facts; 60,395 patients]
- Diabetes mellitus [376,016 facts; 31,783 patients]
- Frontiers Research Participant Registry [7,654 facts; 7,654 patients]
- GLUCOSE (2010) [872,592 facts; 100,192 patients]
- GLUCOSE (2011) [1,007,666 facts; 119,524 patients]
- HEMOGLOBIN A1C (2013) [80,703 facts; 35,269 patients]

Person #25806 - f - Non-Hispanic
- BMI (Calculated) [376,016 facts; 31,783 patients]
- Frontiers Research Participant Registry [7,654 facts; 7,654 patients]
- GLUCOSE (2010) [872,592 facts; 100,192 patients]
The dream: landing the big one

http://www.oregon.com/columbia_gorge_attractions/bonneville_hatchery
Without getting bit
• The Fundamental Theorem of Biomedical Informatics:
  – A person working with an information resource is better than that same person unassisted.

– NOT!!
Background: Bill Stead

Evidence

Patient Record

Clinician

Synthesis & Decision

William Stead: http://courses.mbl.edu/mi/2009/presentations_fall/SteadV1.ppt
The demise of expert based practice is inevitable

Dan Masys, William Stead: http://courses.mbl.edu/mi/2009/presentations_fall/SteadV1.ppt
Background: Edward Shortliffe

Biomedical Informatics Applications

Background: Edward Shortliffe

Biomedical Informatics Research Areas

Machine learning
Text interpretation
Knowledge engineering

Knowledge Base

Inferencing System

Biomedical Knowledge

Biomedical Data

Knowledge Acquisition

Biomedical Research Planning & Data Analysis

Data Acquisition

Real-time acquisition
Imaging
Speech/language/text
Specialized input devices

Model Development
Information Retrieval
Diagnosis
Treatment Planning
Human Interface
Teaching
Image Generation
“It is the responsibility of those of us involved in today’s biomedical research enterprise to translate the remarkable scientific innovations we are witnessing into health gains for the nation.”
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
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.
NIH CTSAs: Home for Clinical and Translational Science

CTSA HOME

- Trial Design
- Biostatistics
- Regulatory Support
- Participant & Community Involvement
- Advanced Degree-Granting Programs
- Clinical Research Ethics
- Biomedical Informatics
- Clinical Resources

Gap!

Dan Masys: http://courses.mbl.edu/mi/2009/presentations_fall/masys.ppt
Reengineering Clinical Research

Bench → Bedside → Practice

Interdisciplinary Research
Innovator Award
Public-Private Partnerships (IAMI)
Building Blocks and Pathways
Molecular Libraries
Bioinformatics
Computational Biology
Nanomedicine

Translational Research Initiatives
Integrated Research Networks
Clinical Research Informatics
NIH Clinical Research Associates
Clinical outcomes
Harmonization
Training

Dan Masys: http://courses.mbl.edu/mi/2009/presentations_fall/masys.ppt
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).
Aim 1 Examples: this lecture, intro,

**REDCap: Research Electronic Data Capture**

- Added REDCap self service model to complement Velos Clinical Trial Management System
- For the last 2 quarters KUMC has some of the highest growth in new users nationally (> 700)
- Low barrier to entry, no extra passwords low management cost (<1 FTE)
- But, people are free to screw up their research
Aim 2: Constructing a Integrated Data Repository
Ethical and Regulatory Concerns

• Who “owns” the data? Doctor, Clinic/Hospital, Insurer, State, Researcher... perhaps the Patient?
  – Perception/reality is often the organization that paid for the system owns the data.
  – My opinion: we are custodians of the data, each role has rights and responsibilities

• Regulatory Sources:
  – Health Insurance Portability and Accountability Act (HIPAA)
  – Human Subjects Research

• Research depends on Trust which depends on Ethical Behavior and Competence

• Goals: Protect Patient Privacy (preserve Anonymity),
Will the released columns in combination with publicly available data re-identify individuals?

What if the released columns were combined with other items which “may be known”?

Sensitive columns, diagnoses or very unique individuals?

New measures to quantify re-identification risk.

Constructing a Repository: Understanding Source Systems, Example CPOE

Most Clinical Systems focus on transaction processing for workflow automation.
Constructing a Repository: Understanding Differing Data Models Used by Systems

Hierarchical databases (MUMPS), still very common in Clinical systems (VA VISTA, Epic, Meditech)

http://www.cs.pitt.edu/~chang/156/14hier.html

Hypothetical Relational Database Model

Relational databases (Oracle, Access), dominant in business and clinical systems (Cerner, McKesson)

Star Schemas: Data Warehouses

• 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.
• 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”
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.

Artificial Risk Scale (0 to 10)

- Identified Data Set
- Limited Data Set
- De-Identified Data Set

Obfuscated Counts | Actual Counts | View line-item (plug ins) | Download line-item | View notes - scrub LDS/DeID | Download notes - scrub LDS/DeID

Author: Russ Waitman, KUMC
Last modified: July 22, 2011
“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
HERON, an Integrated Data Repository, is Hard: Need Technical Expertise to Surmount Challenges

Moving to FusionIO storage memory platform (August 2011) improved performance, but after upgrading to i2b2 1.6, query times increased significantly.

Hardware/Software: SUSE Linux Enterprise Server 11 (x86_64), Oracle 10g, 70G of RAM, 12 CPUs, Fusion-IO 1.28 TB IoDrive Duo storage tier
HERON Cimarron (data through June 2012)
- Cancer Cases [9,325,899 facts; 63,779 patients]
- Demographics [18,190,532 facts; 1,915,719 patients]
  - Age [1,891,822 facts; 1,891,822 patients]
  - Ethnicity [1,915,719 facts; 1,915,719 patients]
  - Frontiers Research Participant Registry [11,637 facts; 11,637 patients]
  - Gender [1,915,719 facts; 1,915,719 patients]
  - Language [1,915,719 facts; 1,915,719 patients]
  - Marital Status [1,915,719 facts; 1,915,719 patients]
  - Place: School District [213,881 facts; 213,881 patients]
  - Place: State [1,273,103 facts; 1,273,103 patients]
  - Place: distance from KUMC [1,205,390 facts; 1,205,390 patients]
  - Race [1,915,719 facts; 1,915,719 patients]
  - Religion [1,915,719 facts; 1,915,719 patients]
  - Vital Status [2,100,385 facts; 1,915,719 patients]
    - Deceased [21,691 facts; 21,691 patients]
    - Deceased per SSA [184,666 facts; 184,666 patients]
    - Deferred
      - Living [203,073 facts; 203,073 patients]
    - Not recorded [1,690,955 facts; 1,690,955 patients]
- Diagnoses [20,461,541 facts; 617,972 patients]
- Flowsheets [490,945,281 facts]
- Laboratory Tests [75,783,302 facts; 267,529 patients]
- Medications [100,395,527 facts; 262,717 patients]
- Medications by VA Class/Clinical Dose Form (DRAFT) [89,758,747 facts; 249,754 patients]
- Procedures [10,099,551 facts; 554,882 patients]
- REDCap [15,922 facts; 123 patients]
- Specimens [31,261 facts; 3,056 patients]
- UHC DRAFT 2008 Q4 to 2012 Q1 [3,898,600 facts; 56,930 patients]
  - UHC Agency for Healthcare Research and Quality [381,163 facts; 56,861 patients]
  - UHC Core Measures [308,324 facts; 12,134 patients]
  - UHC Demographics [184,764 facts; 56,930 patients]
  - UHC Diagnosis [1,952,136 facts; 56,930 patients]
  - UHC Procedures [232,001 facts; 49,374 patients]
  - UHC Visit Details [840,212 facts; 56,930 patients]
- Visit Details [2,242,318 facts]
- Clinical Services [2,242,318 facts]

HERON: Current Contents

- >800 million facts
- 1.9 million patients but...
  - Most are just old administrative registrations
Richness of Phenotype is the Goal.
Example: Frontiers Participant Registry

Frontiers Participant Data Richness Today

- Patients with Medications Data
- Patients with Laboratory Results

*All Frontiers Participants have Diagnosis and Procedure Data.
Engagement and Review

• Dedicated Coordinator. Informatics Clinics held biweekly and one-to-one trainings and consultations offered

• Integrating HERON’s use into other research workflows
  • Finding patients for prospective trials: combining the Frontiers Participant Registry with the EMR data to find willing participants that meet study criteria.
  • Searching for samples: Biospecimen Repository combined with EMR to find tissues that meet research criteria.

• Auditing small queries
Supporting National Cancer Institute Cancer Center Designation

Incorporate Clinical, Administrative, Research Datasources

- Inpatient and outpatient electronic medical records (Epic)
- Professional Services Billing and Scheduling (GE IDX)
- KUCC Biospecimen Shared Resource Samples Database
- Hospital (KUH) Tumor Registry (NAACR format)
- Social Security Death Master File (NIST format)
- Technical Charges from hospital and clinics (UHC validated format)
- Research Data Capture (REDCap)
- Clinical Research Information System (Velos)

HERON’s current contents with Cancer Center centric data in green

- Demographics (master patient index)
- Race/Ethnicity
- Laboratory Results
- Nursing observations/vital signs
- Clinical Diagnoses (ICD9)
- Medications (dispensed, ordered, home meds, administered)
- Physician Orders
- Procedure charges (CPT)
- Outpatient Billing diagnoses (ICD9)
- Inpatient visit/provider service

- Specimen collected
- Tumor Staging and Grade
- Diagnosis and Treatment
- Survival and Progression
- Site Specific Factors (e.g. ER positive)
- Death per Social Security Administration
- MSDRG, APDRG, LOS, Readmissions
- Technical Charge Diagnoses ICD9
- Service line, AHRQ quality and JCAHO core measures
- Triple Negative Breast Cancer Registry initial pilot completed

Status as of September 19, 2012

http://informatics.kumc.edu/work/wiki/HeronProjectTimeline#March2012Planning
- contains current plan for next several monthly releases
Idealized HERON Cancer Center Research Workflow

1. See what we have
2. Define a cohort
3. Conduct Analysis
4. Resulting Plot

Cancer Survival: Obese (BMI > 30) Diabetic Breast
Realizing Return and Future Directions

• Where next?
  – Decreasing bioinformatics costs versus pricey EMR and extraction
  – **Engagement** to maximize return: Enterprise and Self-Service versus Fee for Service financial model

• Clinics/Hospital as Informatics “Lab” versus Direct Service
  – “Pathologist”: Diagnosing what’s happening with EMR and outcomes
  – “Surgeon”: apply informatics to influence clinical behavior and care
  – Service model for partnership with informatics versus standard consultant vendor model for hospitals? **Estimated win-win return**

• Expansion/Collaboration City, State, and Beyond?
  – Health care happens everywhere once you’re a decent catchment area
  – Unique assets may be less about money than people/political/cultural
  – Sharing to create environment for research? example: SHRINE project