
How Medical Informatics and HERON Can Help Your Research?

Russ Waitman
Division of Medical Informatics
Department of Biostatistics
November 17, 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**
- **Upcoming Clinical Data in HERON**
 - Supporting both the CTSA and Cancer Center
- **Open time to explore hypotheses**
 - Cohort identification for prospective trials
 - Cohorts for observational health services research.
- I will not cover REDCap, which is a good electronic data capture tool we are now providing <https://redcap.kumc.edu>

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).

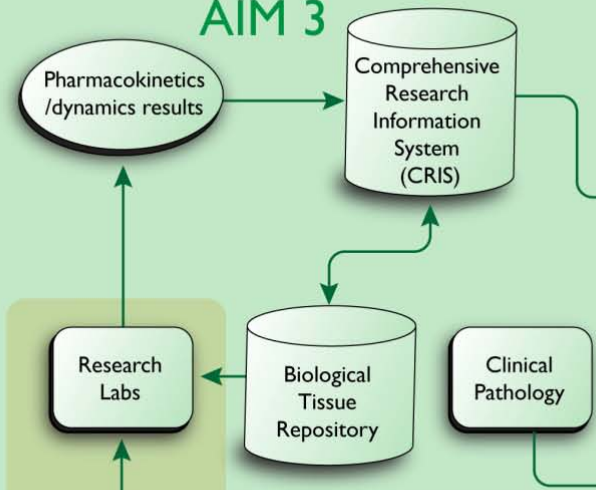
C.1: HICTR Single Signon Portal **AIM 1**

Researcher 



C.3: Advance Innovation T1

AIM 3

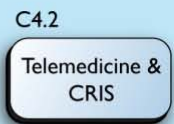


C.2: Healthcare Enterprise Repository for Ontological Narration

AIM 2



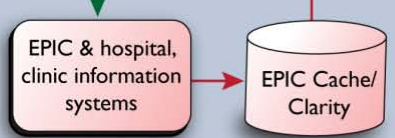
C.4.1 HIE leadership, C.4.2 clinical research in the community, C.4.4: community HIT as a research platform C.4.5 Database Resources Integration T2



State Agencies Healthcare Policy Researcher 

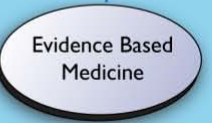
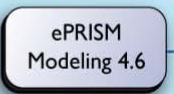
C.4.3: "wire" hospital systems for T2 research

AIM 4



Tertiary Care Clinician 

C.4.6: deploy novel decision support in new environments T2



AIM 4

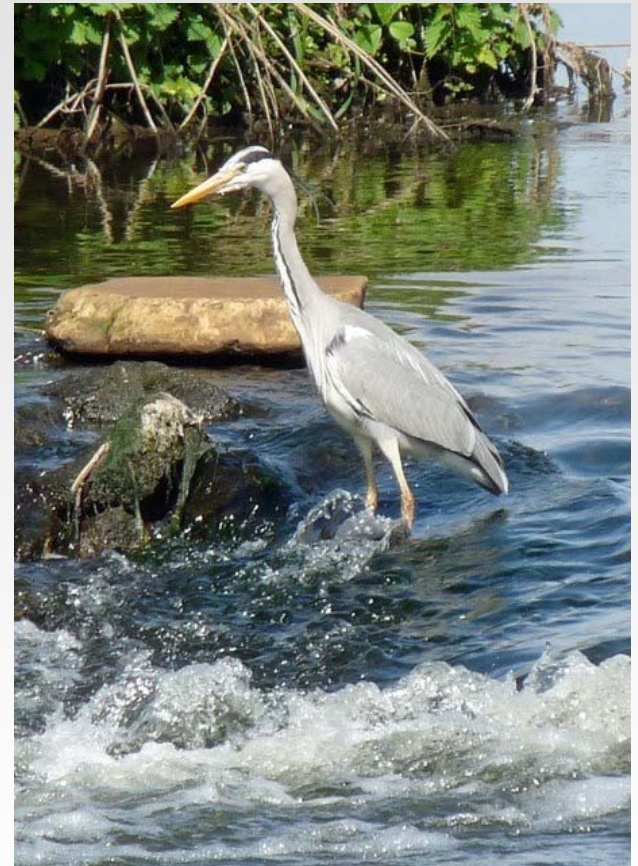
AIM 4



Community Clinician 

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 using your email username**
- **Real-time check for current human subjects training**
- **System Access Agreements, Data Use Agreements and Review Processes implemented in HERON with web pages for monitoring system use**

- **Demonstration**
 - **i2b2 and HERON tools**

HERON De-identification Decisions

- **HIPAA Safe Harbor De-identification**

- Remove 18 identifiers and **date shifting by 0-365 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)

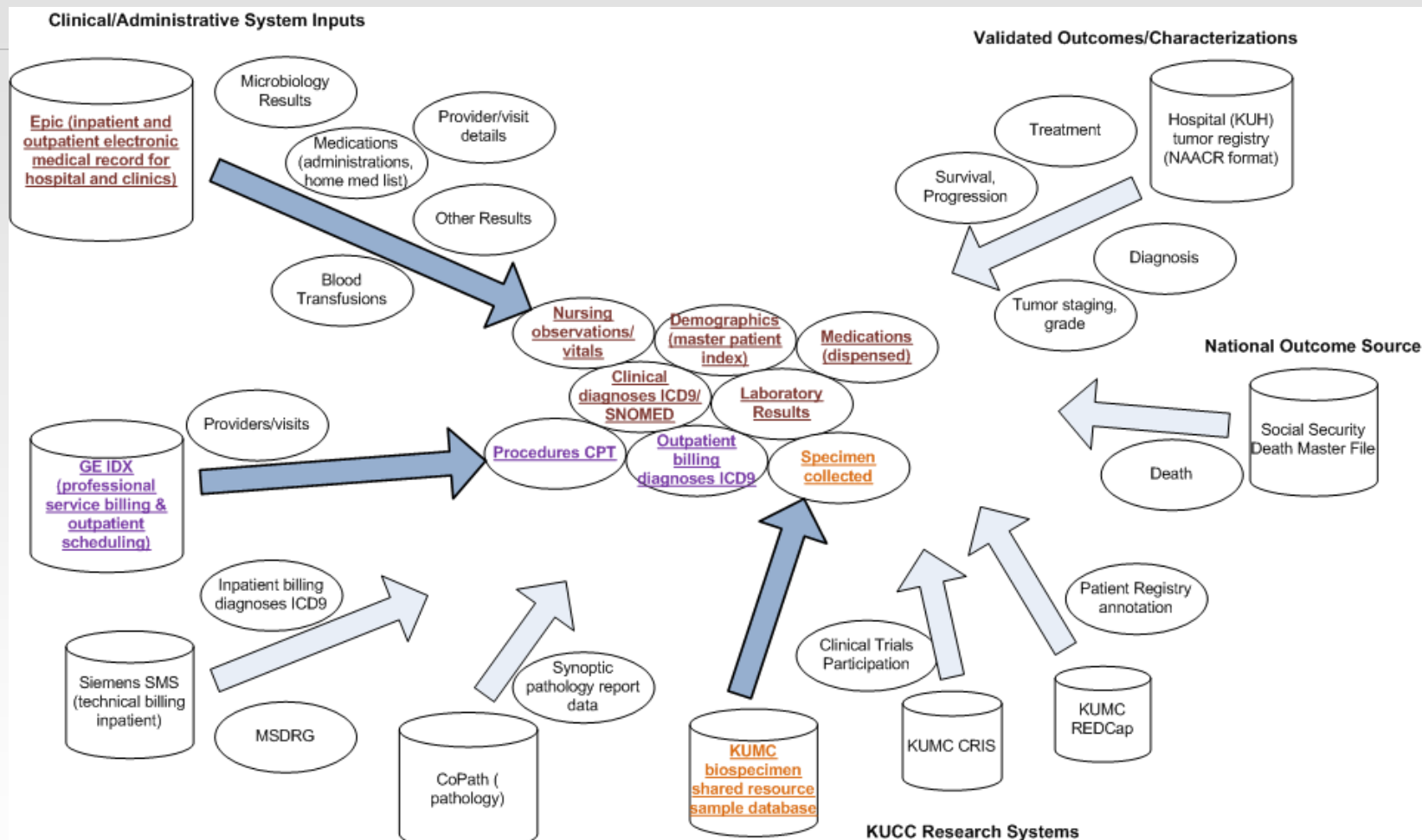
- **To be addressed:**

- For now, we won't add free text such as progress notes with text scrubbers (DeID, MITRE Identification Scrubber toolkit)

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

The screenshot displays the i2b2 Query & Analysis Tool interface. The top navigation bar includes 'Find Patients', 'Admin', 'Analysis Tools', 'Message Log', 'Raven', 'Send Feedback', 'Help', and 'Logout'. The main workspace is divided into several panels:

- Navigate Terms:** Lists various data categories such as Demographics (15,314,963 facts; 1,852,332 patients), Diagnoses (16,211,283 facts; 547,771 patients), Flowsheets (366,435,648 facts), Laboratory Tests (39,114,314 facts; 154,792 patients), Medications (23,831,200 facts; 85,720 patients), and Procedures (8,797,452 facts; 508,467 patients).
- Workplace:** Shows a folder structure with 'SHARED' and 'rwaitman'.
- Previous Queries:** Lists recent queries with their IDs and dates, such as 'N-W-0-0-0-S-A@16:11:44 [8-8-2011] [rwaitman]'.
- Query Tool:** The central area where the query is built. It shows three groups connected by 'AND' operators. Each group contains a search term: 'Neoplasms [1,349,216 facts]', 'WBC COUNT (#3009) [954]', and '001- #11 Height [426,645 facts]'. Below the groups, there are green boxes labeled 'one or more of these'.
- Query Status:** Shows the results of the query: 'Finished Query: "N-W-0-0-0-S-A@16:11:44" Patient Count - 8517 patients FINISHED [15.0 secs]'.

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.

Biospecimen Shared Resource Integration

i2b2 Query & Analysis Tool

Navigate Terms

Find Terms

- Heron Cottonwood (Contains July 2011 data)
 - Demographics [15,367,969 facts; 1,858,481 patients]
 - Diagnoses [16,413,722 facts; 551,339 patients]
 - Flowsheets [374,860,347 facts]
 - KUMC BSR Specimens [20,412 facts; 2,247 patients]
 - fluid [12,843 facts; 1,936 patients]
 - blood [12,692 facts; 1,933 patients]
 - other [150 facts; 11 patients]
 - tissue [7,569 facts; 1,522 patients]
 - ? [47 facts]
 - bone, connective tissue, skin, breast [869 facts; 232 patients]
 - digestive systems and peritoneum [1,701 facts; 238 patients]
 - genitourinary [3,733 facts; 849 patients]
 - Bladder [124 facts; 25 patients]
 - Cervix [33 facts; 10 patients]
 - Endometrium [116 facts; 56 patients]
 - Fallopian Tube [28 facts; 11 patients]
 - Kidney [656 facts; 89 patients]
 - Labia [<10 facts]
 - Myometrium [<10 facts]
 - Ovary [346 facts; 93 patients]
 - Prostate [2,293 facts; 598 patients]
 - Seminal Vesicle [<10 facts]
 - Uterus [119 facts; 23 patients]
 - Vaginal Wall [<10 facts]
 - Vulva [<10 facts]
 - lip, oral cavity, pharynx [301 facts; 45 patients]
 - respiratory and intrathoracic organs [571 facts; 104 patients]
 - unspecified sites [347 facts; 74 patients]
 - Laboratory Tests [39,974,387 facts; 157,954 patients]
 - Medications [24,366,357 facts; 88,062 patients]
 - Procedures [8,861,630 facts; 511,233 patients]

i2b2 Web Client Breast cancer biomarkers and HER2 t...
breast cancer biomarkers

nd Patients | Admin | Analysis Tools | Message Log | Raven | Send Feedback | Help | Logou

Query Tool

Query Name:

Group 2			Group 3			Group 4		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
			Prostate	Malignant neoplasm of pro		Surgical Procedures on the		

one or more of these AND one or more of these AND one or more of these

Run Query New Query 4 Groups New Group

Query Status

Finished Query: "PRO-Pro-Mal-Sur@21:00:45"
Patient Set - 262 Patients - 262 patients
Patient Count - 262 patients

FINISHED [11.0 secs]
FINISHED [11.0 secs]

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 match on SSN and Date of Birth**
 - **Should be released with August data this month**
 - **~ 200,000 of our 1.8 million patients are noted as dead in the SS DMF.**

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**

After Tumor Registry

- **Hospital Billing System**
 - Technical charges diagnoses, MSDRGs, insurance status
 - Clean up demographic inconsistencies (race, ethnicity)
- **i2b2 upgrade from 1.4 to 1.6 to reason over visits**
 - Ex: Lab and med during the same encounter
- **More good data from O2/Epic**
 - Medications (ordered, home meds, administered)
 - Exploit i2b2 1.6: units, service and primary/secondary Dx
- **Clinical Trial status and patient registry data from KUMC CRIS and REDCap**
- **CoPath for discrete path report data**

Questions and Hypothesis Generation

