AMIA Secondary Use of Health Data – Research WG  
Thursday, June 14th, 2007

Michael J. Becich, MD PhD (becich@pitt.edu)  
Chairman, Dept of Biomedical Informatics – http://www.dbmi.pitt.edu

Funded Efforts at Pitt Relevant to Secondary Data Use:
CDC Mesothelioma Virtual Tissue Bank – http://www.mesotissue.org
NCI Shared Pathology Informatics Network (SPIN) – http://spin.nci.nih.gov
NCI Cooperative Prostate Cancer Tissue Resource (CPCTR) – http://www.prostatetissues.org
NCI SPOREs in Lung and Head & Neck – http://spore.nci.nih.gov
NCI Cancer Center Support Grant – http://ccsg.nci.nih.gov
NCRR CTSA Program – http://www.ctsi.pitt.edu/

NHLBI Production Assistance for Cellular Therapies - http://www.pactgroup.net/
Pennsylvania Cancer Alliance Bioinformatics Consortium – http://pcabc.upmc.edu
Modern clinical research requires complex, interconnected systems

- Clinical research occurs in complex systems of organizations, people and information systems that require detailed descriptions of
  - definitions of data (what)
  - the process of care (how)

- So that differences in outcomes can be analyzed and compared

NOTE: Credit for this and the next two slides are to Doug Fridsma, MD PhD (U Pitt and BRIDG project innovator)
Modern clinical research requires complex, interconnected systems

- But clinical research must be able to interoperate with the larger research enterprise
  - Other clinical and research applications
  - Other clinical and research systems
  - Other clinical and basic departments

- Standardization and the ability to “hand off” data and research results is critical to innovation
Modern clinical research requires complex, interconnected systems. Ultimately, successful clinical research requires complex connections to exchange information within and between organizations.
Researchers Need Answers To These Questions

- What kind of questions are asked by investigators:
  - Which genes are regulated in disease? At different stages?
  - Disease classification – assign new diseases classes
  - Diagnostics – assign patient samples to known disease classes
  - What is the correlation between disease progression & -omics?
  - Does gene expression correlate with outcome?
  - Reveal disease related alterations in pathways (systems biology)
  - Monitor effects of experimental therapies (theranostics)

- Generally investigators focus on one or two of these questions.
  - Where is the opportunity? Recycling data to general research population
  - Strategy? Allow whole scale de-identified clinical annotation with -omics analysis and offer to research community
Clinical Environment
(Identified Data)

Tissue Bank and Cancer Registry
(“Honest Broker Services”)

Research Environment
(De-Identified Data)

Data Integration Grid
Supporting Research
Honest Broker System

Goals and Objectives

• Enhance Collaborative Translational Research Efforts
• Monitor Requests for Information & Data Use Practices
  – Quality Assurance and Process Improvement
  – Preparation for Research
  – IRB-Approved Research
  – Other Uses
• Develop Web-Based Request Tracking Tool
• Centrally Approve Requests and Review of Supporting Documentation
• Centralize Training and Management (HIPAA, Human Subjects, Security and Privacy)
15 Cancer Centers, Universities and Research Institutes with Cancer Research Facilitated by the University of Pitt’s Cancer Informatics Grid

Slide 9
Department of Biomedical Informatics
Proven Infrastructure for Collaboration

- Portfolio of active collaborations:
  - Pennsylvania Cancer Alliance Bioinformatics Consortium (PCABC) – Abramson Cancer Center (U Penn), Fox Chase Cancer Ctr, Kimmel Cancer Ctr (TJU), Penn State Cancer Center (Hershey) and Wistar Institute
  - Shared Pathology Informatics Network (SPIN) – Harvard, UCLA, Univ Indiana
  - Cooperative Prostate Cancer Tissue Resource (CPCTR) – GWU, Howard Univ, Med College of Wisconsin, NYU and VA
  - Patient Safety Database – Kaiser Permanente, Henry Ford, Univ Iowa and West Penn Allegheny
  - Overcoming Barriers to Clinical Trials - MGH, U Colo, Wash U, UC Davis, OSU
  - Early Detection Research Network – Numerous sites; data sharing model
- Tracking IRB and consents for protocol (clinical trial) and non-protocol
- De-identification software for ‘honest brokering’ of clinical data
- Honest broker program administered by Network Cancer Registry Lead
- Network security, storage, archiving and support
Information Model for Bioinformatics

Support of the Pancreas SPORE

Firewall

Clinical Environment (Identified Data)
- Clinical Trials Info Management System (CTIMS)
- Anatomic Pathology - Lab Info System (cells & tissue)
- Clinical Information Sources
- Clinical Path - Lab Info System (serum & blood)
- Clinical Blood and Tissue Source Info
- Cerner/MARS - Clinician Nursing Data
- One way interface

Tissue Bank and Cancer Registry (Honest Broker Environment)
- Tissue Bank Information System (Clinical, Pathology & Outcomes)
- Cancer Registry
- Clinical Annotation - Tissue Bank, Informatics & Outcomes Data
- One way interface
- De-identification Software
- De-identified Data Only

Research Environment (De-identified Data)
- Organ Specific Research Database
- Research Information Sources
- cDNA Microarrays
- Tissue Microarrays
- Proteomics
- High Throughput Facilities Microarrays, Mass Spectroscopy, etc...
- Genomic and Proteomic Bioinformatics Analysis Tools and Services
- Overall Architecture & Integration

Virtual Tissue Microarrays & PGEC Data Mining Tools (System Interfaces)
- Research Data Server
- Analysis and Visualization Servers
- Intranet

Overall Architecture & Integration Plan for Bioinformatics
Applications Supported in Service to the Cancer Center

- Clinical Applications Supported (and managed by our team):
  - 18 Hospitals across Western PA and 9 Cancer Centers/Regional Centers
  - 18 Hospitals Anatomic Pathology Lab Information Systems (Cerner’s CoPath)
  - 12 Hospitals Clinical Pathology Lab Information Systems (Misys FlexiLab and Cerner LIMS)
  - 12 Hospitals Cancer Registry System (IMPATH); OP/Phys Offices planned
  - Clinical Trials Management Application (4 academic hospitals, cancer centers).
  - MARS – Medical Archival System (all data sources including all ADT, Laboratory, Radiology, Pharmacy, H&P, Consult and Operative Notes)
  - UPCI and UPMC Cancer Centers websites (for patient information and scheduling) and 15 other web properties (add link)
  - LIMS for genomics (Affymetrix); GEDA, PGED, VTMA all web based
  - Tissue Banking Information and Inventory System for 4 academic hospitals and cancer centers
  - Organ Specific Databases (datawarehouse)
  - Whole Slide Imaging, Teleconferencing and Telepathology applications
IPS Server and Relinking Architecture

IPS Client

Request

IPS Server

Text Processor And Statistical Engine

Model #1
Model #2
Model #3
Model #4

Data Set #1
Data Set #2

DE-ID Processing
Encrypted Linkage File To Reside On IPS Server

Query to MARS to retrieve document types of interest

MARS Data Repository

go back
Clinical Environment (Identified Data)

Tissue Bank and Cancer Registry (Honest Broker Environment)

Research Environment (De-identified Data)

Virtual Tissue Microarrays & PGEC Data Mining Tools (System Interfaces)

Overall Architecture & Integration

Plan for Bioinformatics

Genomic and Proteomic Bioinformatics Analysis Tools and Services

High Throughput Facilities Microarrays, Mass Spectroscopy, etc...

Molecular Oncology, Molecular Virology, Immunology, Cancer Epidemiology

De-identified Data Only

Tissue & Data Integration Support

Intranet

Clinical Trials, Clinical Labs, Anatomic Pathology & Medical Records

Anatomic Pathology - Lab Info System (cells & tissue)
Clinical and Translational Science Awards (Consortium)

The following CTSA grant awardees have agreed to post part of their application for public viewing. Click on the institution’s name to view this information.

Columbia University Health Sciences (New York, NY)
Duke University (Durham, NC)
Mayo Clinic College of Medicine (Rochester, MN)
Oregon Health & Science University (Portland, OR)
Rockefeller University (New York, NY)
University of California, Davis (Davis, CA)
University of California, San Francisco (San Francisco, CA)
University of Pennsylvania (Philadelphia, PA)
University of Pittsburgh (Pittsburgh, PA)
University of Rochester (Rochester, NY)
University of Texas Health Science Center at Houston (Houston, TX)
Yale University (New Haven, CT)

NIH has created a national consortium that will transform how clinical and translational research is conducted, enabling researchers to develop new treatments faster and deliver them to patients more efficiently and quickly. Led by the National Center for Research Resources (NCRR), this new consortium, funded through Clinical and Translational Science Awards (CTSA), begins with 12 academic health centers located throughout the nation. An additional 52 AHCs are receiving planning grants to help them prepare applications to join the consortium. When fully implemented in 2012, about 60 institutions will be linked together to energize the discipline of clinical and translational science. The new program draws on NIH’s earlier initiatives to re-engineer the clinical research enterprise, one of the key objectives of the NIH Roadmap for Medical Research. The NIH News Release and the CTSA Questions and Answers offer additional information on the consortium.

UMLS Based Parser – Pathology Reports and Electronic Medical Records are in text blocks. Need to extract data into chunks.

Autocoder – Once data is “chunked” then you need to extract discrete elements (e.g., tumor location, size, grading, staging, etc…) and codify the data for rapid storage and retrieval.
PATIENT HISTORY:  
The patient is a **AGE-year-old male with a clinical history of prostate cancer.

OSS **PATH-NUMBER<1>, **DATE<11/12/00>, **PLACE .

FINAL DIAGNOSIS:

**VIEW CONCEPTS**

PART 1: PROSTATE, LEFT LOBE, NEEDLE BIOPSY (OSS# **PATH-NUMBER<1>, **DATE<11/12/00>, **PLACE)
A. MODERATELY DIFFERENTIATED PROSTATE ADENOCARCINOMA. GLEASON'S PATTERN 3+3, SCORE = 6, INVOLVING ONE OF FIVE CORES, LESS THAN 2% OF TISSUE SUBMITTED (see comment).
B. NO PERINEURAL INVASION SEEN.

**VIEW CONCEPTS**

PART 2: PROSTATE, RIGHT LOBE, NEEDLE BIOPSY (OSS# **PATH-NUMBER<1>, **DATE<11/12/00>, **PLACE)
A. MODERATELY DIFFERENTIATED PROSTATE ADENOCARCINOMA, GLEASON'S PATTERN 3+4, SCORE = 7, INVOLVING ONE OF FIVE CORES, LESS THAN 5% OF TISSUE SUBMITTED.
B. NO PERINEURAL INVASION SEEN.

kmr

**INITIALS** kmr **NAME, M.D.** Fellow/Chief Resident: **NAME, MBBS. M.D.** Resident: **NAME, M.D.** **NAME MMM VVV: **NAME, M.D.**DATE<0/2/00> 10:27

OUTSIDE ACCESSION #:
2 SLIDES LABELED **PATH-NUMBER<1>
OUTSIDE **NAME RECEIVED: Y
CONSULT MATERIAL DESCRIPTION:
Received for consultation from **NAME, M.D., are two (2) consult slides labeled **PATH-NUMBER<1> from **PLACE, **ADDRESS,
Data Sources

- Multiple Hospitals, Multiple Systems:
  - AP LIS: Merged 18 systems to 1
  - CP LIS: Merged 18 system to 2
  - Tissue Banking: Health Sciences Tissue Bank (ongoing consolidation)
  - Cancer Registry: 14 systems to 1
  - Clinical Trials Management: 1

Researchers