Development of the Informatics Infrastructure for the MCC

Hai Hu, PhD
Senior Director of Biomedical Informatics
Deputy Chief Scientific Officer
Windber Research Institute

WRNMMC • Bethesda, MD • June 23, 2014
Big Science, Team Science

Cancer Genome Discovery, Domen-Farber Cancer Institute, Boston, Massachusetts 02115 USA.
Department of Biochemistry and Molecular Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605 USA.
Department of Pathology, Brigham and Women’s Hospital, Boston, Massachusetts 02115 USA.
The Cancer Institute of New Jersey, New Brunswick, New Jersey 08903 USA.
Department of Pathology, University of Maryland School of Medicine, Baltimore, Maryland 21201 USA.
Department of Pathology, Stanford University School of Medicine, Stanford, California 94305 USA.
Department of Pathology, Weill Cornell Medical College, New York, New York 10065 USA.
Department of Veterinary Science, University of Illinois at Urbana-Champaign, Illinois 61801 USA.
Department of Medical Oncology, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854 USA.
National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892 USA.
Department of Pathology, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78284 USA.
Department of Pathology, University of Washington, Seattle, Washington 98195 USA.
The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030 USA.
The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 USA.
Department of Anatomy, Division of Genitourinary and Somatic Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030 USA.
Department of Anatomic Pathology, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030 USA.
Department of Pathology, University of Texas Medical Branch, Galveston, Texas 77555 USA.
Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas 75390 USA.
Department of Pathology, University of Virginia, Charlottesville, Virginia 22908 USA.
Big Data—Data Silos or Data Center
MCC Clinical Services and Research Programs

Clinical Departments and Services

- Dermatologic Oncology/Mohs Surgery
- Hematology Oncology
- Gynecologic Oncology
- Orthopaedic Oncology
- Pediatric Oncology
- Radiation Oncology
- Surgical Oncology

Disease Specialties

- Breast Cancer
- Gynecologic Cancers
- Pediatric Cancers
- Prostate Cancer

Patient and Family Support Services

- Support and Counseling Services

Translational Research Programs and Clinical Trials:

- Women's Malignancies Program (CBCP, GYN, and more)
- Military Population Sciences and Epidemiology Program
- Urologic Malignancies Program (CPDR, and more)
- Inflammation, Infection, Immunity, and Stroma Program (new)
Informatics Infrastructure

Adapted from Hu H. et al. 2008
Clinical Laboratory Workflow System for CBCP

- Subject enrollment
- Questionnaire completion
- Double data entry

- Specimen processing
- Tissue banking

- Specimen integration w/ lab modules

- Other utilities

- Designed by WRI
- Implemented by Cimarron
- Deployed in November, 2004
Data Warehouse for Translational Research (DW4TR)

Flexible and Extensible Data Model
- Patient-centric clinical data model
- Specimen-centric molecular data model
- Temporal data model, image data model

Federation

Internal Data
- Medical History
- Pathology
- Diagnostic
- Treatment Outcome
- Medical Images
- Specimen/Tissue Banking
- Genomic Analysis
- Proteomic Analysis

External Data
- Genomic Data
- Proteomic Data
- Disease Data

Phase I. Vision and Conceptual Design

**Vision Definition**
Meeting with MCC Executives

**Conceptual Design**
Meetings with leaders of functional units
Major inputs, processes, and outputs
Interactions between functional units
Existing systems—strengths and weaknesses
Expected infrastructure
Learn from successful NCI CCCs

Big Picture
Phase II. Detailed System Design/Use Case Development

- Meeting with technical staff of each functional unit
- Clinicopathologic data
  - Questionnaires and data forms
    - A general form for common data elements
    - Followed by disease-specific ones
  - AHLTA/EMR
- Biospecimen handling and banking
- Molecular study platforms
- Clinical Trials
- Data tracking
  - Different levels of details for different types of data
- Data Warehousing
  - What to integrate and how
  - How data will be used
- Scale of data to be managed
- System hosting and management
  - Locally hosted at MCC, or at USUHS
  - Remotely hosted, offsite
  - CLOUD
  - Military firewall
Phase III. Implementation Methods and Execution

**Evaluation of**
MCC existing systems, extensibilities, and team capabilities
Open source systems
Commercial systems
Timeline and the cost

**System implementation model**
Build—internal and external
Purchase—external
Combination

Internal development:
• Use of existing infrastructure and open source solutions
• Need to build/contract a development team

External efforts
• Will develop Request For Proposals covering requirements/use cases, and criteria for evaluation
• Invite for on-site presentation and demonstration

**Implementation in stages**
Develop a thin application layer to reflect the conceptual design of the whole system, enable expansion
Focus on one cancer and go deep
Extend to other cancer studies one by one
Summary

**Background:** Big science  
Big data

**Experiences:** System design  
Data tracking  
Data warehousing

**Development:** Big picture  
Details  
Implementation
Project Team

**WRI Biomedical Informatics**
Hai Hu, Deputy CSO and Sr. Director
Leonid Kvecher, Data Manager
Ed Elston, IT Manager
Yuanbin (Kevin) Ru, Bioinformatics Scientist

**Informatics Management Consultants, LLC**
Marc Nodell, President, and the team

**MCC**
COL Craig D. Shriver, Director
Al J. Kovatch, Jeff A. Hooke
and the team
Richard J. Mural, WRI Chief Scientific Officer
Died of cancer on May 29, 2014
Thank you all!

h.hu@wriwindber.org