

caBIG

Vocabularies & Common Data Elements

Heather Kincaid

IT Manager

COMPASS Coordinating Center

Fred Hutchinson Cancer Research Center

hkincaid@fhcrc.org

February 18, 2004

Early Detection Research Network (EDRN)

- 5-Year collaboration supported by NCI
- Goal: Identify, evaluate, and validate promising biomarkers to support the early detection of cancer
- Comprised of:
 - 18 Biomarker Laboratories
 - 9 Clinical and Epidemiology Centers
 - 3 Biomarker Validation Laboratories
 - Data Management and Coordinating Center

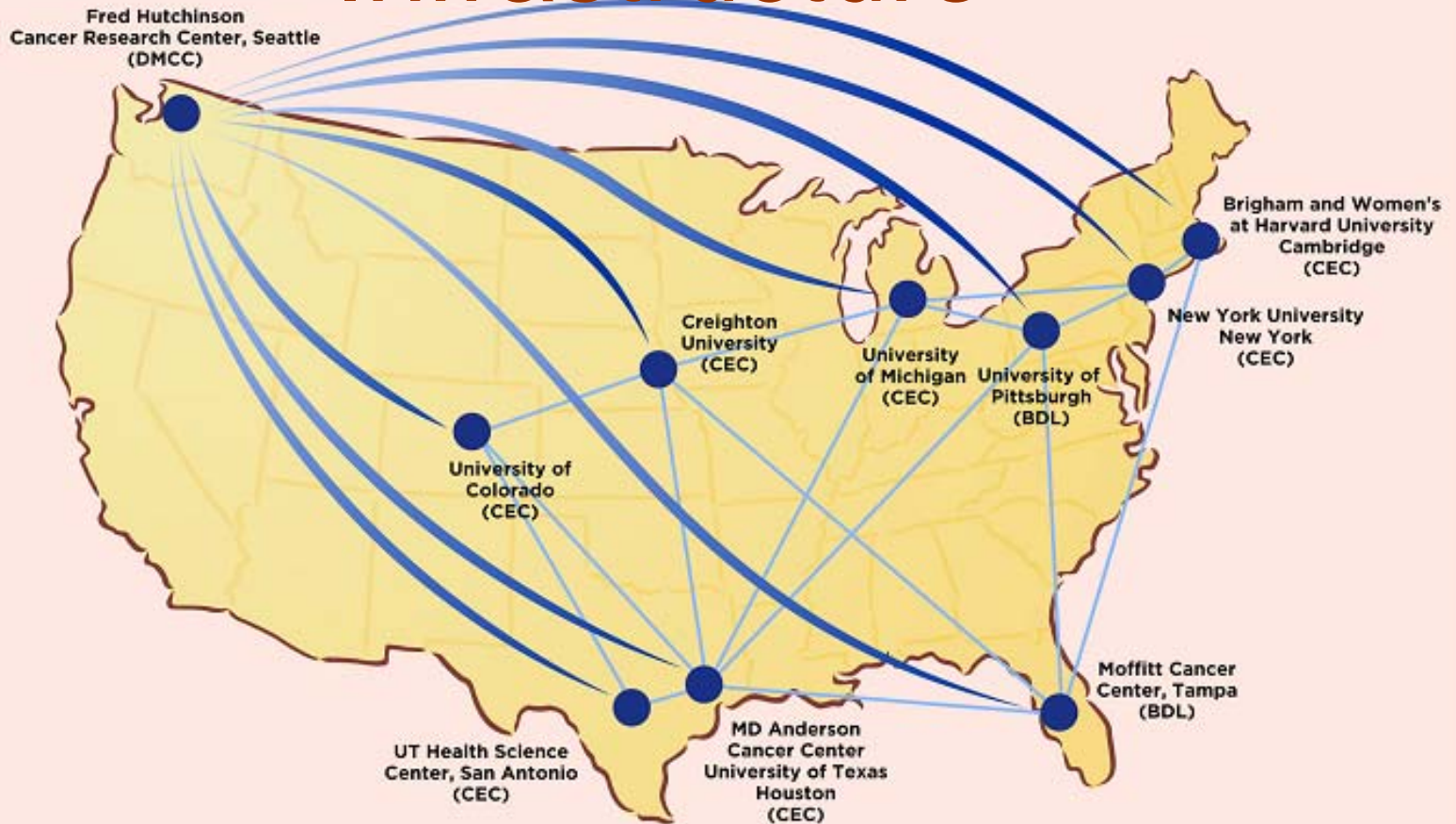
Informatics Approach

- *Cross-disciplinary* team of biomedical and computer science researchers
- *Common Data Elements* to standardize data definitions for databases and forms
- *Informatics infrastructure* that allows for capture and exchange of information across EDRN centers
 - Leverage JPL/NASA's *experience* and software in developing IT infrastructures to support planetary science
 - Use *existing EDRN databases* without requiring changes
 - Develop a *common IRB protocol template*
- Common portals to access data (secure, validation, public, etc) as a single entry point

EDRN Resource Network Exchange (ERNE)

- Virtual Specimen Repository
- Informatics infrastructure created for EDRN
- Existing sites specimen databases maintained locally
- Uses EDRN Common Data Elements (CDEs)
- Maps institutions local data definitions to EDRN CDEs
- Secure and Confidential
- Secure Dynamic Portal

Deployed EDRN Informatics Infrastructure



EDRN Informatics Tools

- EDRN Secure Website - CDE Tools
 - CDE Repository
 - Form Tools
 - Mapping Tools
- EDRN Resource Network Exchange (ERNE)
 - An *infrastructure* for sharing data resources across EDRN
 - Supports *real time* (on demand) *distribution* of data to users
 - First release - *Specimen sharing tool*
 - EDRN CDE Mapping Tool
- Validation Infrastructure (VSIMS)
 - Provide common infrastructure across validation studies
 - Online Forms - Data Driven from CDE Repository

The screenshot shows the EDRN Secure Website interface. At the top, there is a navigation bar with links for SEARCH, FEEDBACK, SITE MAP, EDRN PUBLIC, LOGIN, and Logout. Below this is a 'Search Criteria' section with a search box and a 'Search' button. The search results are displayed in a table with columns for CDE Name, Baseline, Demographics, and Gender Code. The table contains several rows of data, including CDE names like 'CDE Classification All' and 'CDE Physical All'. A 'CDE Name Baseline Demographics Gender Code' table is also visible, showing columns for CDE Name, Baseline, Demographics, and Gender Code. The table contains several rows of data, including CDE names like 'CDE Classification All' and 'CDE Physical All'. The table also includes a 'Source' column and a 'Note' column.

The screenshot shows the EDRN Resource Network Exchange (ERNE) interface. At the top, there is a navigation bar with links for SEARCH, FEEDBACK, SITE MAP, EDRN PUBLIC, LOGIN, and Logout. Below this is a 'Search Criteria' section with a search box and a 'Search' button. The search results are displayed in a table with columns for CDE Name, Baseline, Demographics, and Gender Code. The table contains several rows of data, including CDE names like 'CDE Classification All' and 'CDE Physical All'. A 'CDE Name Baseline Demographics Gender Code' table is also visible, showing columns for CDE Name, Baseline, Demographics, and Gender Code. The table contains several rows of data, including CDE names like 'CDE Classification All' and 'CDE Physical All'. The table also includes a 'Source' column and a 'Note' column.

The screenshot shows the Validation Infrastructure (VSIMS) interface. At the top, there is a navigation bar with links for SEARCH, FEEDBACK, SITE MAP, EDRN PUBLIC, LOGIN, and Logout. Below this is a 'Search Criteria' section with a search box and a 'Search' button. The search results are displayed in a table with columns for CDE Name, Baseline, Demographics, and Gender Code. The table contains several rows of data, including CDE names like 'CDE Classification All' and 'CDE Physical All'. A 'CDE Name Baseline Demographics Gender Code' table is also visible, showing columns for CDE Name, Baseline, Demographics, and Gender Code. The table contains several rows of data, including CDE names like 'CDE Classification All' and 'CDE Physical All'. The table also includes a 'Source' column and a 'Note' column.

Semantic Architecture

- Many institutions have existing specimen repositories with locally defined data models
- EDRN Common Data Elements (CDEs)
- ISO/IEC 11179

Data Model Mapping

- Communicating EDRN CDEs
- EDRN CDE Mapping Tool
- EDRN CDE Repository

CDE Tools

- CDE Attributes

- Form Tool

Early Detection Research Network
 An infrastructure for supporting collaborative research on molecular, genetic and other biomarkers in early cancer detection and risk assessment
 NATIONAL CANCER INSTITUTE
 DIVISION OF CANCER PREVENTION

SEARCH FEEDBACK SITE MAP EDRN PUBLIC LISTSERV Relevant Sites:

Search Criteria:

CDE Document Text: All

Search Method: Matches on all words (AND)

CDE Category: All

CDE Classification: All

CDE Phase: All

CDE Version: All

CDE Form: All

Permissible Value: All

Specimen Sharing Project: All

Current Version: Yes

CDE Name: BASELINE_DEMOGRAPHICS_GENDER_CODE

Document Text: Gender (What is your gender?)

Category: Baseline

Classification: Demographics

Version: 1.0

Data Type: Integer

Permissible Value: List ID	List Value	Source
1	Male	
2	Female	
9	Unknown/refused	


Note: Permissible Values listed above are specified for all related forms.

Forms Applied: [Core Baseline CDE \(1.31 - Approved\)](#)
[Feasible Form \(1.2 - Development\)](#)
[hematologic \(1.0 - Development\)](#)
[MSA Baseline \(1.0 - Approved\)](#)

Specimen Sharing Project: Yes

Early Detection Research Network
 An infrastructure for supporting collaborative research on molecular, genetic and other biomarkers in early cancer detection and risk assessment
 NATIONAL CANCER INSTITUTE
 DIVISION OF CANCER PREVENTION

SEARCH FEEDBACK SITE MAP EDRN PUBLIC LISTSERV Relevant Sites:

Note: Click on the  icon to see the details of selected CDE(s) of interest.

EDRN Core Baseline CDE

EDRN Site ID

EDRN Protocol ID

EDRN Participant ID

Was consent obtained prior to data collection?

No Yes

IRB waived requirement for consent

Was consent obtained prior to data collection? (If No, explain)

Date of baseline data collection

Date of Participant Enrollment

Demographics

Gender (What is your gender?)

Male Female

Unknown/refused


Height [in inches] (What is your total current height in inches?)

Hispanic or Latino (Are you Hispanic or Latino?)

No Yes

CDE Tools

- CDE Mapping

Note: Click on the  icon to see the details of selected CDE(s) of interest.

Map ID: 1059 Last Modified Date: 2002/07/08




Site and Protocol Information

58 The Biomarker Knowledge System Informatics Pilot Project Supplement To The Biomarker Development Laboratory at Moffitt (Bedlam) 89 H. Lee Moffitt Cancer Center (BDL)

[Protocol Database and Site Technology Survey](#)

Mapping Results		Comments
	EDRN CDE 	EDRN2_Sput_Epi
Table Name		M_Sput_Subject
CDE Name	BASELINE_DEMOGRAPHICS_GENDER_CODE	SEX
CDE Version	1.0	
CDE Data Type	Integer	Character
CDE Document Text	Gender (What is your gender?)	Gender
Permissible Value	1 Male 2 Female 9 Unknown/refused	F M U
CDE Comments		
Mapping Solution		Case M.Cdem_SEX F --> E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 2 M --> E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 1 Else E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 9 EndCase Else E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 9 EndCase
Mapping Comments		This is being done in Product Server
Mapping Status	Completed	Completed

- Mapping Evaluation

	EDRN CDE 	EDRN2_Sput_Epi
CDE Category	Baseline	
CDE Classification	Demographics	
Table Name		M_Sput_Subject
CDE Name	BASELINE_DEMOGRAPHICS_GENDER_CODE	SEX
CDE Version	1.0	
CDE Data Type	Integer	Character
CDE Document Text	Gender (What is your gender?)	Gender
Permissible Value	1 Male 2 Female 9 Unknown/refused	F M U
CDE Comments		
Mapping Solution Proposed		
Mapping Solution		Case M.Cdem_SEX F --> E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 2 M --> E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 1 Else E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 9 EndCase Else E.BASELINE_DEMOGRAPHICS-GENDER_CODE = 9 EndCase
Others		
Mapping Comments		This is being done in Product Server

Distributed Specimen Sharing infrastructure

- “Grids” heterogeneous centers together
- Real-time access to distributed cancer centers

EDRN Home Page

Server Status

DMCC Up
 Creighton University Up
 Moffitt Cancer Center Up
 University of Colorado Up
 University of Michigan Up
 University of Pittsburgh Up
 UT, San Antonio Up

EDRN Resource Network Exchange
 - Informatics Pilot Project

Early Detection Research Network

Search For: (Listed specimens are available at sites selected. Field with * is required)

Specimen Source: Blood Participant Cancer Status: Participant With Cancer

Update Options

PLEASE CHOOSE ALL THAT APPLY

Characteristics of Specimen:

Specimen Stored: DNA Final Storage: All

Demographics:

Gender: All Male Female

Hispanic/Latino Origin: All Hispanic/Latino Not Hispanic/Latino

Race: All White Black or African-American American Indian or Alaska Native Asian Native Hawaiian/Other Pacific Islander

History of Regular Smoking: All Yes No

Characteristics of Individuals with Cancer:

Cancer Sites:

All Bladder Bone Brain
 Breast Cervix Colon Esophagus
 Head/Neck Kidney Leukemia Liver
 Lung Lymphoma Ovary Pancreas
 Prostate Rectum Skin Stomach
 Thyroid Uterus

Histology Classification:

Search Results - Microsoft Internet Explorer

EDRN Home Page

EDRN Resource Network Exchange
 - Informatics Pilot Project

Options for another Search:

Search Results:

Specimen Source: Blood
 Specimen Stored: DNA
 Participant Cancer Status: Participants With Cancer

Protocol ID	Site ID	Site Name	# of Samples	# of Ports	Contact
65	80	Creighton Univ.	288	39	public@creighton.edu
64	73	Univ. of Colorado	3	3	wilbur.franklin@uchic.edu
38	67	Univ. of Michigan	12	3	dbrenner@umich.edu

Data Summary (Creighton University, Protocol ID: 65) Details

ODE Category	Number of Samples	Number of Participants
Gender		
Male	207	14
Female	81	25
Race		
White	274	36
Unknown	14	3
Smoking History		
Smoked Regularly	49	14
Not Smoked Regularly	20	10
Unknown	219	15

Data Summary (University of Colorado, Protocol ID: 64) Details

ODE Category	Number of Samples	Number of Participants
Gender		
Male	3	3
Race		
White	3	3
Smoking History		
Smoked Regularly	3	3

Data Summary (University of Michigan, Protocol ID: 38) Details

ODE Category	Number of Samples	Number of Participants
Gender		
Male	0	2
Female	4	1
Race		
White	2	1
Black or African-American	4	1
Unknown	6	1
Smoking History		
Not Smoked Regularly	0	2
Unknown	4	1

If you have any questions or suggestions, please send a feedback message to [DMCC](#).

[Members](#) | [Committees](#) | [Collaborative Groups](#) | [Resources](#) | [Informatics](#) | [Network](#) | [Publications](#) | [Home](#)

Download: [Number Counter](#) | [Change Password](#) | [Registration](#) | [Privacy Policy](#) | [Upgrade Process](#) | [Security](#)

This site is maintained by COMPASS, Fred Hutchinson Cancer Research Center.

COMPASS For comments and suggestions regarding this website, Email info@edrn.org

EDRN Informatics Tools: Validation Studies

- Forms Entry
- Data Transfer
- Eligibility Confirmation
- Specimen Tracking
- Real-time Data Reporting
- Issue Tracking
- Portal Interface

VSIMS
Validation Study Information Management System

An infrastructure for supporting
validation studies in the EDRN
Sponsored by National Cancer Institute



Protocol: MSA (ID: 108) User: Qing Xiao (Site ID: 128)

[Logout](#) [Home](#) [Back](#) [Print](#) [VSIMS](#)

Staff Name: Qing Xiao

MSA Role: None

▣ **Early Detection Research Network Validation Study** (Baseline Entry)

- MSA Consent-1.0
- MSA Eligibility-1.0 [Ineligible List \(14 \)](#)
- MSA Baseline-1.21
- MSA Specimen Collection-1.0
- MSA Cystoscopy Procedure and Results (G2 & G3 only)-1.0
- MSA TURBT Procedure and Results (G3 only)-1.0
- MSA Other Procedure and Results (G3 only)-1.0
- MSA Upper Tract Imaging (G3 only)-1.0
- MSA Urinalysis and/or Heme Results (G1 & G2 only)-1.2
- MSA Culture and Sensitivity (CNS) Results (G2 Only)-1.0
- MSA Urine Cytology Results-1.0

[Submit](#)

[Reset](#)

[Confirm Eligibility](#)

[Follow-up Form](#)

[Main](#)

Secure site maintained by COMPASS, Fred Hutchinson Cancer Research Center ©. Last updated on 1/12/04

Contact: edrdmco@frc.org

Acknowledgements

- Brigham and Womens
- Cancer Biomarkers Group, NCI
- Creighton University
- Fred Hutchinson Cancer Research Center
- H. Lee Moffitt Cancer Center
- MD Anderson
- New York University
- NCI Center for Bioinformatics
- NASA Jet Propulsion Laboratory
- University of Texas, San Antonio
- University of Colorado
- University of Michigan/Dartmouth University (Great Lakes New England Consortium)
- University of Pittsburgh

Informatics Goals

- Science-driven approach and requirements
- Development of a core infrastructure to enable seamless data sharing between EDRN informatics centers
- Centers “plugged” into the informatics infrastructure
- Data collection and sharing environments for
 - Specimens
 - Validation Studies
 - Biomarkers, Genomics and Proteomics
- Science informatics tools