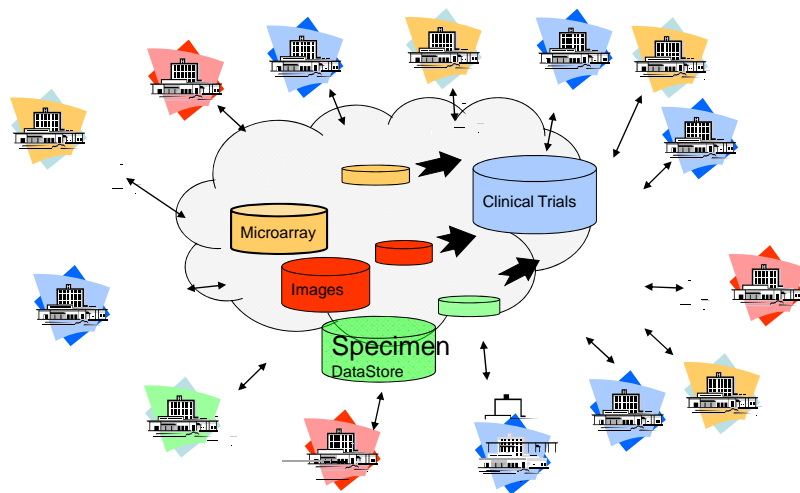




Common Data Elements (CDEs): A Perspective from  
the National Cancer Institute and the caBIG® Program  
September 24, 2010

Dianne M. Reeves, RN, MSN  
National Cancer Institute Center for Biomedical Informatics  
and Information Technology

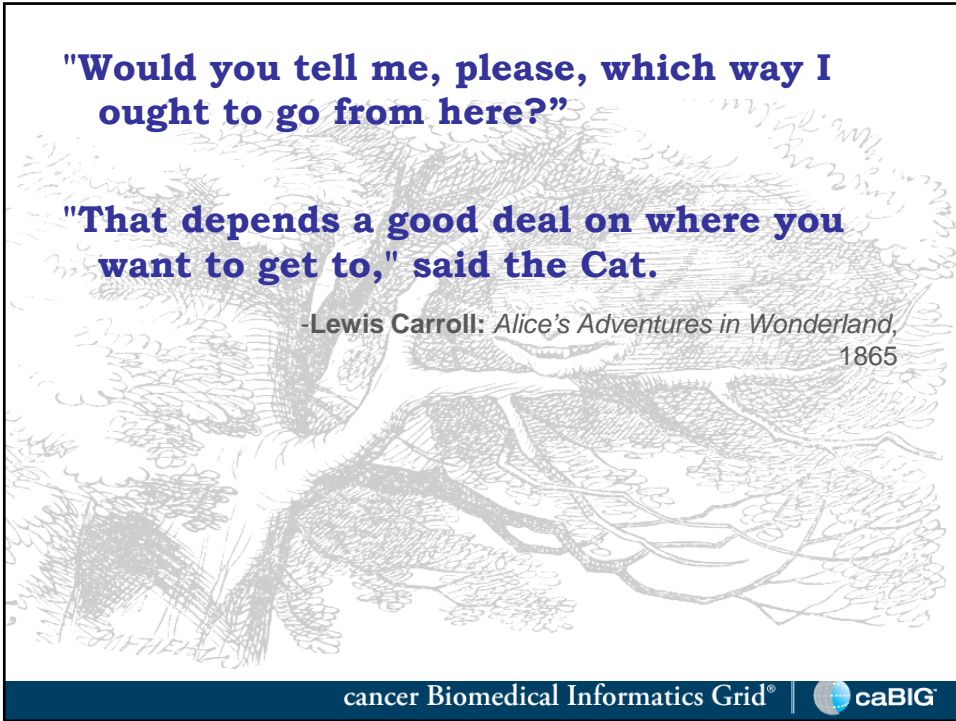
Numerous data sets generated by independent, often  
silozed systems, aggregated and used by many sites



**"Would you tell me, please, which way I ought to go from here?"**

**"That depends a good deal on where you want to get to," said the Cat.**

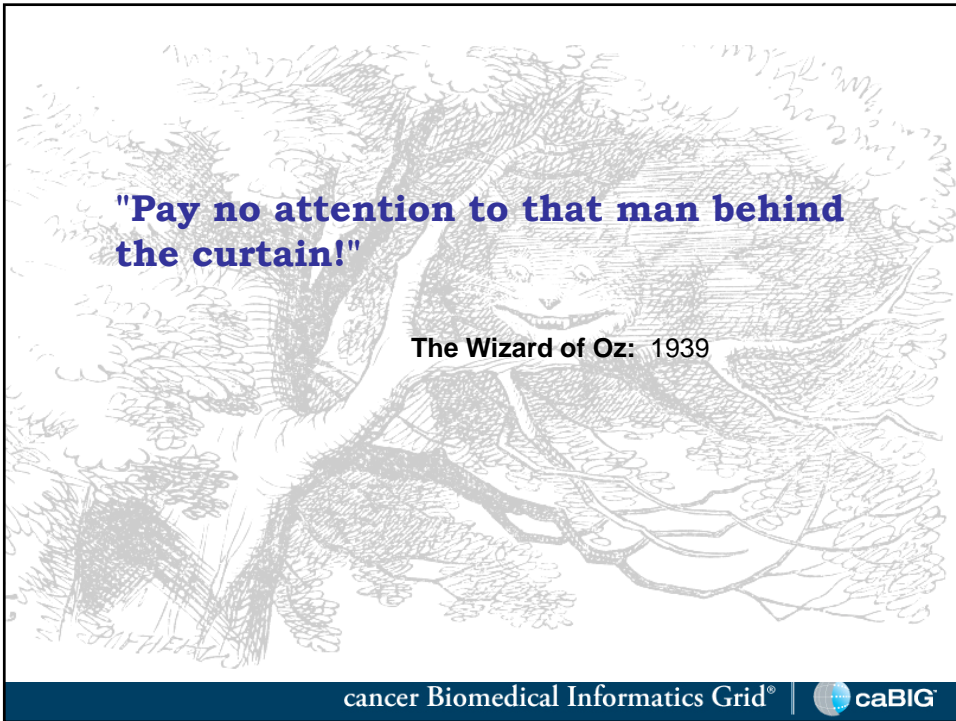
-Lewis Carroll: *Alice's Adventures in Wonderland*, 1865



cancer Biomedical Informatics Grid® | caBIG®

**"Pay no attention to that man behind the curtain!"**

The Wizard of Oz: 1939



cancer Biomedical Informatics Grid® | caBIG®

## 2010 Cancer Statistics

### Projections for the US are:

- 1,529,560 new cancer cases
- 569,490 deaths

**Cancer accounts for more deaths than heart disease in persons younger than 85 years**

**Approximately 2 – 3% of patients participate in clinical trials**

**Relieve suffering and death due to cancer by the year 2015**

## Cancer Clinical Trials

### Cancer Therapy Evaluation Program

- Largest publicly funded clinical trial organization world-wide
- 900 Active Trials
- 30,000 Participants annually
- 400 Grants/Cooperative Group agreements
- > 100 Investigational New Drugs (INDs)

All CTEP trials collect **data** using registered **data elements**

## Impacts on Data Processes

### Regulatory Requirements

### Sponsor Requirements

### Resources

### Data Specifics

- Imaging
- Genomics
- Clinical/Treatment

### Reporting Requirements

**Data is a final deliverable from a research project**

## Cancer Clinical Trials

### Cancer Clinical Trial

- Collect 200 to 800+ separate variables
- Collect variables in 20 – 45 Case Report Forms (CRFs)
- Variety of Reporting Requirements

= Large Data Sets to **Aggregate** and **Analyze**

**What happens if the data cannot Aggregate?**

## Cancer Data Standards Registry and Repository (caDSR)



Common Data Elements

cancer Biomedical Informatics Grid® |  caBIG®

### caDSR Overview



- **Enterprise-class system/database for storing metadata, with APIs, download capability**
- **Implements the ISO/IEC 11179 v2 standard, a flexible model for describing metadata**
- **Basic caDSR unit of metadata information to describe a datum - a Common Data Element or CDE**
- **Common Data Element collections in 'Forms' are also stored in caDSR**

cancer Biomedical Informatics Grid® |  caBIG®

## Data Elements

### What is a (Common) Data Element?

A single unambiguously described data point/variable

- The Question
- The Answer

Data Elements support

- Interoperability
- Aggregation of data sets
- Reuse

### caDSR Overview:

### Metadata Example: Patient Age in Years

Data

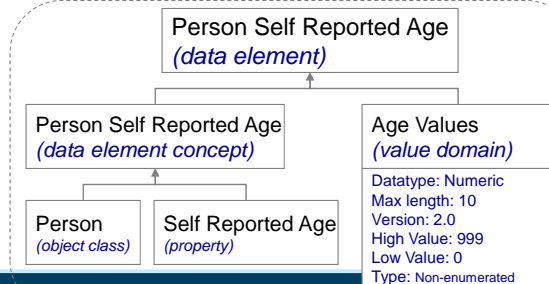
What is your age?:

32

stored in

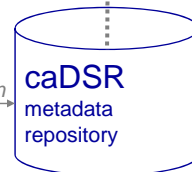


Metadata



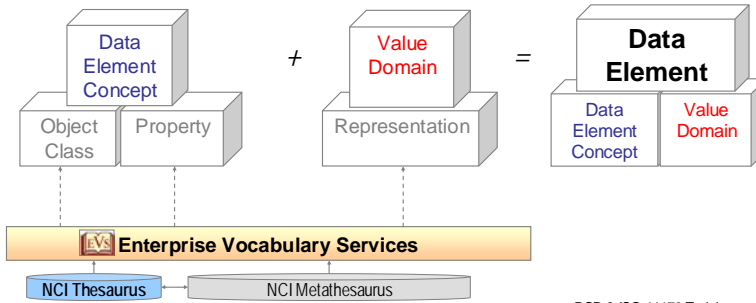
stored in

Describes  
the data in



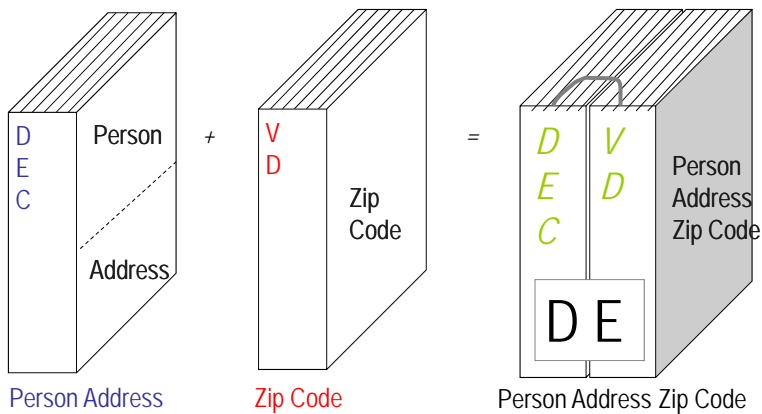
## Data Element Fundamentals Use of Controlled Vocabularies

*What is it?* + *How do you want to represent it?* = *Common Data Element*

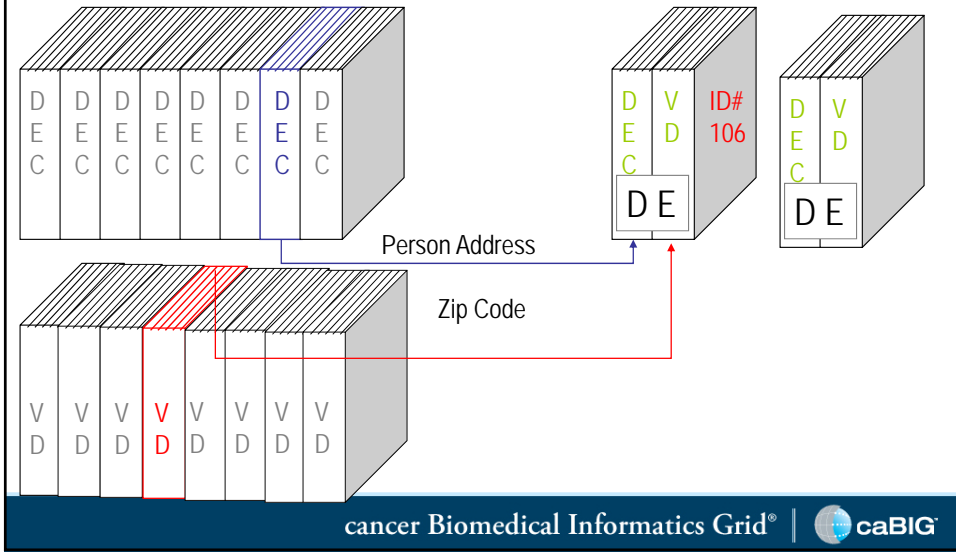


caDSR & ISO 11179 Training  
Jennifer Brush, Dianne Reeves

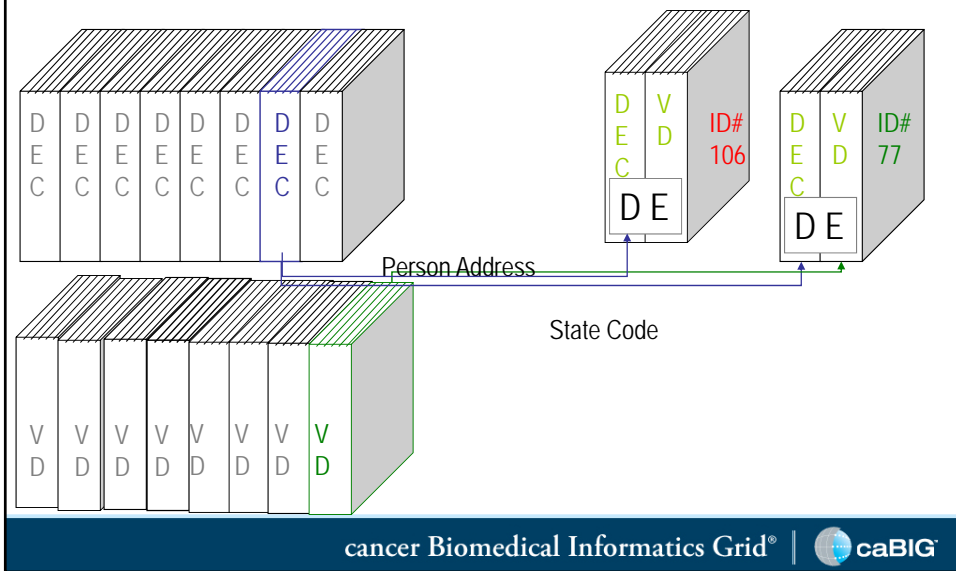
## Data Element Fundamentals



### Libraries of Re-usable Components



### Libraries of Reusable Components



## caDSR Tools



### **CDE Browser / FormBuilder**

- Search for and Download Data Elements
- Collect Data Elements onto Forms and Download Forms



### **CDE Curation Tool**

- Curate (Create and Edit) Data Element Concepts, Value Domains and Data Elements



### **Sentinel Tool**

- Create Alert Definitions to monitor changes to caDSR metadata



### **Admin Tool**

- Main administrative interface to all of the caDSR features and components

## Purpose of a Metadata Registry/Repository

**Provide a location to register variables/descriptors/data elements to describe research, applications, etc.**

**Provide a location for organizational standards, elements for widespread use**

**Support semantic interoperability**

# CDE Browser

The screenshot shows the CDE Browser interface. On the left is a tree view of 'caDSR Contexts' including items like 'caBIG (NCI) cancer Biomedical Informatics Grid' and 'caCORE (NCI Core Infrastructure)'. The main content area has a search bar with a dropdown menu set to 'Name'. Below the search bar are 'Search' and 'Clear' buttons. At the bottom, there is a footer with the text 'cancer Biomedical Informatics Grid' and the 'caBIG' logo.

# CDE Browser

The screenshot shows the search results page. At the top, there are links for 'Download Data Elements to Prior Excel', 'Download Data Elements to Excel', 'Download Data Elements as XML', and 'Download CDE Browser DTDL'. Below these is a table of search results. The table has columns: 'Long Name', 'Preferred Question Text', 'Owned By', 'Used By Context', 'Registration Status', and 'Version'. A circled dropdown menu indicates '1 - 100 of 4501'. The footer contains the 'cancer Biomedical Informatics Grid' and 'caBIG' logos.

Long Name	Preferred Question Text	Owned By	Used By Context	Registration Status	Version
<input type="checkbox"/> Address Additional Identifier Text	Additional Geographic Place Name Service	caBIG		Standard	RELEASED 2405182 1.0
<input type="checkbox"/> Address Blocks Text Identifier	Address Blocks	caBIG		Standard	RELEASED 2405421 1.0
<input type="checkbox"/> Address Canton Identifier	Address Canton	caBIG		Standard	RELEASED 2405438 1.0
<input type="checkbox"/> Address City Name	City	DCP	CCR CTEP NHLBI caBIG caCORE	Standard	RELEASED 2179601 1.0
<input type="checkbox"/> Address Country Name	Country	DCP	CCR OP caBIG caCORE	Standard	RELEASED 2179604 1.0
<input type="checkbox"/> Address County/Parish Name	Address County/Parish Sub-division Name	caBIG		Standard	RELEASED 2345409 1.0
<input type="checkbox"/> Address County/Parish Sub-Division Name	Address	caBIG		Standard	RELEASED 2405442 1.0
<input type="checkbox"/> Address Department/Division Name Text	Address	caBIG		Standard	RELEASED 2341930 1.0
<input type="checkbox"/> Address Dependent Street/Thoroughfare Name	Address Dependent Street/Thoroughfare Name	caBIG		Standard	RELEASED 2405166 1.0
<input type="checkbox"/> Address District Name	Address District Name	caBIG		Standard	RELEASED 2345413 1.0
<input type="checkbox"/> Address Door Text Identifier	Address Door	caBIG		Standard	RELEASED 2405161 1.0
<input type="checkbox"/> Address Double Dependent Locality Name	Address Double Dependent Locality	caBIG		Standard	RELEASED 2405444 1.0
<input type="checkbox"/> Address Entity Name Text	Address Entity Name	caBIG		Standard	RELEASED 2341926 1.0
<input type="checkbox"/> Address Entity Number	Address Entity Number	caBIG		Standard	RELEASED 2341928 1.0
<input type="checkbox"/> Address Farm Identifier	Address Farm	caBIG		Standard	RELEASED 2405432 1.0

## CDE Browser – Use of CRF Standards

Search Results Search within results

Results fewer than expected? Check Search Preferences

[Download Data Elements to Prior Excel](#) [Download Data Elements to Excel](#) [Download Data Elements as XML](#)  
[Download CDE Browser DTDs](#)

Sort order: (Default) Registration Status >> Workflow Status >> Long Name (Ascending)

Long Name	Preferred Question Text	Owned By	Used By Context	Registration Status	Workflow Status	Public ID	Version
<input type="checkbox"/> Address Postal Code Identifier	ZIP Code	DCP	CCR CTEP caBIG	Standard	RELEASED	2179606	2.0
<input type="checkbox"/> Centers for Disease Control and Prevention Ethnicity Unique Identifier Codes	CDC Ethnicity Code	caBIG	CCR, NIDCR	Standard	RELEASED	2200284	2.0
<input type="checkbox"/> Centers for Disease Control and Prevention Race Unique Identifier Code	CDC Race Code	caBIG	CCR, NIDCR	Standard	RELEASED	2200286	1.0
<input type="checkbox"/> Ethnic Group Category Test	Ethnicity	caBIG	CTEP DCP NINDS	Standard	RELEASED	2192217	2.0
<input type="checkbox"/> Form Completion Date	Date Completed	CCR	CTEP DCP NIDCR, SPOREs caBIG	Standard	RELEASED	2003745	5.0
<input type="checkbox"/> Person Birth Date	Patient's Date of Birth	CTEP	DCP NINDS caBIG	Standard	RELEASED	793	5.1
<input type="checkbox"/> Person Education Level Summary Type	Educational Attainment	caBIG	CCR	Standard	RELEASED	2681652	1.0
<input type="checkbox"/> Person Education Level Type	Education Level	caBIG		Standard	RELEASED	2674076	2.0
<input type="checkbox"/> Person Enrollment Address United States Integer 2000 Census Tract Code	Person Census Tract Code at Enrollment	caBIG		Standard	RELEASED	2943243	1.0
<input type="checkbox"/> Person Gender Text Type	Gender of a Person	caBIG	CTEP DCP NHLBI, NINDS	Standard	RELEASED	2200904	3.0
<input type="checkbox"/> Person Healthcare Payer Type	Payment Method	caBIG	CTEP	Standard	RELEASED	2865130	1.0
<input type="checkbox"/> Person Marital Status Category	Marital Status	CCR	caBIG	Standard	RELEASED	2188083	2.0
<input type="checkbox"/> Race Category Test	Race	caBIG	CTEP DCP	Standard	RELEASED	2192199	1.0
<input type="checkbox"/> Person Address Country Name	Country of Residence	CTEP	DCP caBIG	Qualified	RELEASED	315	4.0
<input type="checkbox"/> Person Birth Country Type	Where were you born	CTEP	caBIG	Qualified	RELEASED	2001700	5.0

1 - 16 of 16

Trusted sites 100%

cancer Biomedical Informatics Grid® caBIG

## CDE Browser – Value Meanings/Definitions

Version: 1.0

Long Name: Category

Short Name: C25372

Context: caBIG

Representation Concepts

Concept Name	Concept Code	Public ID	Definition Source	EVS Source	Primary
Category	C25372	2203876	NCI	NCI_CONCEPT_CODE	Yes

Permissible Values

PV	PV Meaning	PV Meaning Concept Codes	PV Meaning Description	PV Begin Date	PV End Date	VM Public ID	VM Version
Divorced	Divorced C51775	C51775	A person whose marriage has been legally dissolved and has not remarried.	2006-03-06		2578101	1.0
Domestic Partnership	Domestic Partner C53262	C53262	A person who is a member of an unmarried couple including same sex couples living together in longstanding relationships, that are registered or unregistered.	2006-06-12		2578103	1.0
Married	Married C51773	C51773	A person currently joined in matrimony. Classify common law marriage as married. Includes married couples living together and not living together.	2006-03-06		2578102	1.0
Never Married	Never married C51774	C51774	A person who has never been married or whose only marriages have been annulled.	2006-03-06		2576696	1.0
Separated	SEPARATED C51777	C51777	A person living apart from his/her spouse by legal arrangement.	2006-06-08		2561620	1.0
Widowed	Widowed C51776	C51776	A person who is no longer married because of the death of his/her spouse and who has not remarried.	2006-06-08		2561048	1.0

Reference Documents

Document Name	Document Type	Document Text	Context	URL
There are no reference documents for the value domain.				

User: Public User Privacy Notice Version: 4.0.1.1 Build: 2 Please send comments and suggestions to [nci@biomax.nih.gov](mailto:nci@biomax.nih.gov)

cancer Biomedical Informatics Grid® caBIG

## CDE Browser – Reuse of Elements

<input type="checkbox"/> Long Name	Preferred Question Text	Owned By	Used By Context	Registration Status	Workflow Status	Public ID	Version
<input type="checkbox"/> <a href="#">Address Postal Code Identifier</a>	ZIP Code	DCP	CCR,CTEP,caBIG	Standard	RELEASED	2179606	2.0
<input type="checkbox"/> <a href="#">Centers for Disease Control and Prevention Ethnicity Unique Identifier Codes</a>	CDC Ethnicity Code	caBIG	CCR,NIDCR	Standard	RELEASED	2200284	2.0
<input type="checkbox"/> <a href="#">Centers for Disease Control and Prevention Race Unique Identifier Code</a>	CDC Race Code	caBIG	CCR,NIDCR	Standard	RELEASED	2200286	1.0
<input type="checkbox"/> <a href="#">Ethnic Group Category Text</a>	Ethnicity	caBIG	CTEP,DCP,NINDS	Standard	RELEASED	2192217	2.0
<input type="checkbox"/> <a href="#">Form Completion Date</a>	Date Completed	CCR	CTEP,DCP,NIDCR,SPORES,caBIG	Standard	RELEASED	2003745	5.0
<input type="checkbox"/> <a href="#">Person Birth Date</a>	Patient's Date of Birth	CTEP	DCP,NINDS,caBIG	Standard	RELEASED	793	5.1
<input type="checkbox"/> <a href="#">Person Education Level Summary Type</a>	Educational Attainment	caBIG	CCR	Standard	RELEASED	2681552	1.0
<input type="checkbox"/> <a href="#">Person Education Level Type</a>	Education Level	caBIG		Standard	RELEASED	2674076	2.0
<input type="checkbox"/> <a href="#">Person Enrollment Address United States Integer:2000 Census Tract Code</a>	Person Census Tract Code at Enrollment	caBIG		Standard	RELEASED	2943243	1.0
<input type="checkbox"/> <a href="#">Person Gender Text Type</a>	Gender of a Person	caBIG	CTEP,DCP,NHLBI,NINDS	Standard	RELEASED	2200604	3.0
<input type="checkbox"/> <a href="#">Person Healthcare Payer Type</a>	Payment Method	caBIG	CTEP	Standard	RELEASED	2865130	1.0
<input type="checkbox"/> <a href="#">Person Marital Status Category</a>	Marital Status	CCR	caBIG	Standard	RELEASED	2188083	2.0
<input type="checkbox"/> <a href="#">Race Category Text</a>	Race	caBIG	CTEP,DCP	Standard	RELEASED	2192199	1.0
<input type="checkbox"/> <a href="#">Person Address Country Name</a>	Country of Residence	CTEP	DCP,caBIG	Qualified	RELEASED	315	4.0
<input type="checkbox"/> <a href="#">Person Birth Country Type</a>	Where were you born	CTEP	caBIG	Qualified	RELEASED	2001708	5.0

1 - 15 of 15

## Use of Tools and Infrastructure through caBIG®

### CDISC

**National Heart, Lung and Blood Institute**

**National Institute of Child Health and Development**

**National Institute of Dental and Craniofacial Research**

**National Institute of Neurological Disorders**

**HL7 Clinical Interoperability Council**

**Pharmaceutical Companies**

**Tumor Registries**

## National Cancer Institute Center for Bioinformatics (NCI CBIIT)

### NCICB established 2001

- To provide the **interoperable** biomedical informatics infrastructure, tools, and data that biomedical communities need for research, prevention and care.
- Bioinformatics provides researchers with the tools, information technologies, and analytical methodologies needed to **manage large volumes** of data.
- Spearheads critical public-private partnerships to develop and disseminate informatics for **managing, analyzing, and sharing** information.

cancer Biomedical Informatics Grid® |  caBIG™

## Cancer Biomedical Informatics Grid (caBIG®)

### caBIG™ established 2004

- A collection of activities to partner with biomedical technologies
- To accelerate research discoveries and improve patient outcomes by linking researchers, physicians, and patients throughout the cancer community
- Goal: To enable large scale data sharing through the common use of applications, infrastructure, and data

cancer Biomedical Informatics Grid® |  caBIG™

## caBIG®: Organization

### Domain Workspaces

- Clinical Trials Management Systems
- Integrative Cancer Research
- In Vivo Imaging
- Tissue Banks and Pathology Tools

### Cross-Cutting Workspaces

- Architecture
- Vocabularies and Common Data Elements

### Strategic Level Workspaces

- Data Sharing and Intellectual Capital
- Documentation and Training
- Strategic Planning



## caBIG®: Organization

### Domain Workspaces

- Clinical Trials Management Systems
- Integrative Cancer Research
- In Vivo Imaging
- Tissue Banks and Pathology Tools

### Cross-Cutting Workspaces

- Architecture
- Vocabularies and Common Data Elements

### Strategic Level Workspaces

- Data Sharing and Intellectual Capital
- Documentation and Training
- Strategic Planning



## NCI CBIT: Products

### Tools for researchers

- **Biology**
- **Clinical trials**
- **Genomics**
- **Laboratory support**
- **Translational research**
- **Data standards**
- **Imaging**
  - NCIA
  - AIM
- **Vocabularies**
  - Radlex



cancer Biomedical Informatics Grid® |  caBIG

## NCI CBIT: Products

### Tools for cancer researchers

- **Biology**
- **Clinical trials**
- **Genomics**
- **Laboratory support**
- **Translational research**
- **Data standards**
- **Imaging**
  - NCIA
  - AIM



cancer Biomedical Informatics Grid® |  caBIG

<https://cabig.nci.nih.gov/tools/GeneConnect#product>

National Cancer Institute | U.S. National Institutes of Health | www.cancer.gov

caBIG™ Cancer Biomedical Informatics Grid™

home » tools » geneconnect

### GeneConnect

Database IDs are linked by [Direct Annotation](#), [Inferred Annotation](#), or [Sequence Alignment](#)

**Quick Links to Page Contents**

- [Product Summary](#)
- [Demo Files & Exercises](#)
- [Documentation & Training](#)
- [Accessing the Tool](#)
- [End User Support Resources](#)
- [Related caBIG™ Tools](#)
- [Developer Institution\(s\) & Adopter Institution\(s\)](#)
- [Related Articles](#)

**Product Summary**

**Intended Audiences:** Bioinformaticians, Bench Scientists

**Area of Focus:** Genome Annotation      **Primary Workspace:** ICR

## Why Use Standards?

- **To bring order to mountains of data**
- **To permit reuse of data**
- **To enable data aggregation,**
  - Thus enabling discovery
- **To insure semantic interoperability**

## The Future

- New GOALS
  - BIGGER than caBIG
  - “BIG Health”
- Issues
  - Broaden caBIG focus
  - Broaden to other platforms
  - Broaden to other tool sets
- “Beyond” implications:
  - Much BIGGER than NCI CBIIT
  - Necessitates adoption and adaption to multiple EXTERNAL Standards and Standards Bodies
  - Original “seamless interoperability on caGrid” is not ONLY use case to satisfy

33

## The Future

Focus on Interoperability

Lower the barriers to collaborators

Evolve our architecture and tools

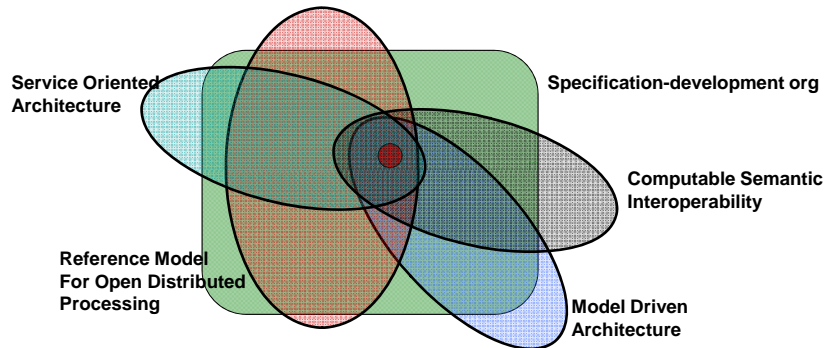
- SAEF is an architectural framework
- Supports Scalability

Current tooling is undergoing change

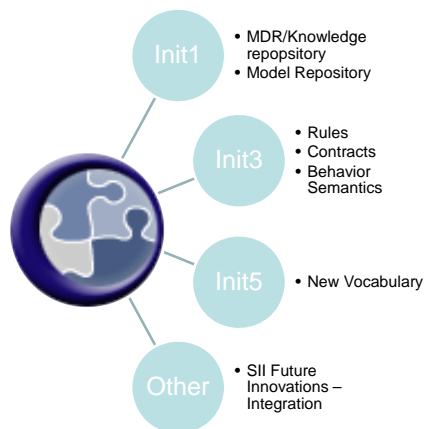
- Requirements-gathering
- Use Case documentation
- Improvement in our tools and work processes

## Contextualizing SAEF

- **SAEAF**: ...portions of a *methodology*, and a *framework* for defining an EA, a robust, durable business-oriented set of constructs that provide *extensibility, reuse, and governance*.



## Requirements Elicitation Process



## The Future










**Curation and Maintenance of Metadata Registry continues**



**New Semantic Infrastructure is projected to be available late 2011**

**Migration Strategy planning is underway**

- Protect investments of our partners
- Improve processes and tooling

.gov/NCICB/training/cadsr\_training/TrainingCurriculum

Course #	1000	1010	1020	1030	1040	1060	1070	2010	2020	2030
<b>Training Role</b>										
<b>Application Developer - Using APIs</b> 	◇	◆	◆	◆	◆			◆	◆	◆
<b>caBIG™ Developer</b> 	◇	◆	◆	◆	◆			◆	◆	◆
<b>caBIG™ Adopter</b> 	◇	◆	◆	◆						
<b>Metadata Analyst</b> 	◇	◆	◆	◆	◆					
<b>Protocol Builder</b> 	◇	◆	◆							
<b>Form Builder</b> 	◇	◆	◆						◆	◆
<b>Metadata Curator - Using Forms*</b> 	◇	◆	◆	◆	◆	◆			◆	◆
<b>Metadata Curator - Using UML Models*</b> 	◇	◆	◆	◆	◆		◆	◆	◆	◆
<b>Context Administrator</b> 	◇	◆	◆	◆	◆				◆	
<b>Subject Matter Expert/ Content Expert</b> 	◇	◆	◆	◆						
<b>VCDE Reviewer</b> 	◇	◆	◆	◆						

\* These roles involve a Preceptorship. The role of Preceptor is explained in detail [here](#) . The skills and competency checklist for a metadata curator preceptorship can be found [here](#) .

- ◇ Course has NO Pre-requisites
- ◆ Course HAS Pre-requisites
- ◆ Suggested for this role, but not required.
- Basic Course
- Intermediate Course

**Now it's Your Turn!**

**Thank you!**