Introduction to the caCore Infrastructure and the caDSR

The following documentation is part of the caCORE training program and is used with permission from the National Cancer Institute Center for Biomedical Informatics and Information Technology

Goal of documentation: To provide the AGNIS end-user with a broad overview of metadata, the caDSR, and the tools used to access metadata in the caDSR

Sections on which to focus:

- 1 Introduction
- 2 Vocabularies and Metadata
- 3.2 Cancer Data Standards Repository (caDSR
- 3.3 ISO/IEC 11179
- 4.2 Purpose of the caDSR Tools
- 4.3.1 CDE Browser
- 4.3.2 Form Builder

Updated URL:
Please note that the URL for the CDE Browser has changed and is incorrect in the documentation. The following is the correct URL for the CDE Browser:

https://cdebrowser.nci.nih.gov/CDEBrowser/

caCORE Training Workbook

Course 1000: Introduction to the caCORE Infrastructure and the caDSR

caCORE Training Website http://ncicb.nci.nih.gov/NCICB/training/cadsr_training
Help & Support ncicb@pop.nci.nih.gov (please include "caCORE Training" in the subject)

Revision History

Authors/Responsible Party	Revision Number	Page(s) Effected	Date
Jennifer Brush	Original	All	December 17, 2006
Jennifer Brush	1	All – template change	January 23, 2006
Becky Angeles	2	All – caCORE 3.1 functionality	June 5, 2006
Becky Angeles	3	All – format and content change	September 27, 2006
Jennifer Brush	4	Updated content slides	November 15, 2006
Dianne Reeves	5	Reviewed Content	November 27, 2006
Jennifer Brush	6	Updated content, inserted terms/abbreviations	November 30, 2006
Becky Angeles	7		August 24, 2007
Becky Angeles	8	Updated Content – Review Questions	July 30, 2008

Terms and Abbreviations

Administered Item: Term used to describe both a data element and its components (data element concept, value domain, object class, and property) when referring to the registration and administration of those components in the caDSR database. For example, all administered items have a name, definition, context, workflow status, version, public ID and begin/end dates.

caCORE (Cancer Common Ontologic Representation Environment): caCORE is the open source group of software products developed by the NCI CBIIT Core Infrastructure Group. By providing a common data management framework, caCORE helps streamline the informatics development throughout the cancer community. The components of caCORE support the consistency, clarity and comparability of biomedical research data and information.

caDSR (Cancer Data Standards Repository): The caDSR is a database and tool set that the NCI and its partners use to create, edit and deploy Common Data Elements.

NCI CBIIT (National Cancer institute Center for Biomedical Informatics and Information Technology): plays a lead role in bioinformatics and information technology by building many types of tools and resources that enable information to be shared along the continuum from the scientific bench to the clinical bedside and back.

Table of Contents

1 Introduction

Welcome to caCORE Training. The session is designed for caDSR Users and Metadata Consumers, including cancer Biomedical Informatics Grid (caBIG) Developers and other participants.

In order to receive credit for mastery of this content, you will need to register for this course and complete the accompanying quiz in the caBIG Learning Management System.

To register, go here: http://ncicbtraining.nci.nih.gov/TP2005/tp2000web.dll/NCICBTraining

We want these sessions to be as effective as possible in meeting your needs so we ask that you complete the short training evaluation form (available when you register for a course) to share your feedback on the overall quality of the training process and materials.

1.1 Course Details

Course Category:	caCORE
Course Number:	1000
Course Title:	Introduction to the NCI CBIIT caCORE Infrastructure and the caDSR
Course Level:	Basic
Audience:	caDSR Users and Metadata Consumers

This course is divided into 3 lessons:

1. Vocabularies and Metadata
2. caCORE Infrastructure
3. caDSR Tools

1.2 Overall Objectives for the Learner

On completion of these course materials, you will be able to:

- Define metadata and its components in caDSR
- Identify three ways Data Elements are used
- Identify the major components of caCORE
- Describe the relationship between caDSR and other caCORE components
- Identify the metadata standard on which the caDSR is based
- Describe four caDSR tools and their functionality
- Define good choices for CDE components, such as the Value Domain and Data Element Concept

2 Lesson 1: Vocabularies and Metadata

A wealth of clinical and scientific and clinical research data exists within the cancer research community. The National Cancer institute Center for Biomedical Informatics and Information Technology (NCI CBIIT) has developed an evolving core infrastructure designed to provide the ability to seamlessly translate and interchange semantically immutable information. The

caCORE infrastructure provides researchers, developers and statisticians with the ability to plug their existing data, sources and methods into a standardized sharable framework that supports information exchange, regardless of source, location or format.

2.1 Objectives for the Learner

At the completion of this lesson, you will be able to:

- Discuss the importance of standardized vocabularies
- Discuss the NCI Enterprise Vocabulary Services (EVS)
- Define Metadata
- Review Data Element fundamentals
- Discuss how and where Data Elements are used

2.2 Standardized Vocabularies

Standard vocabularies are a group of terms that are used or grouped together, usually based on a selection and approval process by a particular community or organization. Standard vocabularies are important to any application involving electronic data sharing. These vocabularies provide a structure and a ‘same-ness’ to the data being collecting.

For example, think about data being collected from a multi-institutional trial. The goal is to collect the same information at multiple places. Forms are filled out, either on paper or electronically. At some point the data ends up in a database. Eventually, when all the trial results are in, researchers and statisticians will want to access all the trial data from each institution. With minimal effort, research statisticians want each data point lined up with the associated data points. More importantly, they want to be sure that the data points they are comparing are semantically equal (e.g. they mean the same thing). In order to make that roll-up possible, not only across institutions, but eventually across trials, standard terminologies and a way to structure the ‘same-ness’ of the collected data is needed.

Standard Vocabularies provide benefits to clinical research by:

- Facilitating translational research
- Increasing accuracy and speed of data collection
- Decreasing approval time for safe and effective treatments
- Accelerating information sharing
- Accelerating discovery
- Ensuring validity of cross study comparison and analysis
- Improving evaluation efficiency, safety, and efficacy
- Integrating diverse data systems
- Improving the links between clinical research and the healthcare delivery system

Standard vocabularies are needed to distinguish data. For example:

- Alcohol (rubbing vs. drinking)
- Colon (punctuation vs. organ)
- Mole (animal vs. blemish vs. unit of measure vs. spy)
- Mouse (animal vs. computer navigation/data entry device)
- Probe (examination vs. investigation vs. instrument)

At NCI CBIIT, standard vocabularies are the foundation of the metadata and computational infrastructure.

2.2.1 Enterprise Vocabulary Services (EVS)

Enterprise Vocabulary Services (EVS) address NCI's needs for a standard/controlled vocabulary and semantics. Different biomedical domains have their own terminologies and ontological resources, each focused on concepts within their domain. A common terminology is the foundation for interoperability between systems, indexing repositories of information, and querying and integrating information sources across research areas. The Enterprise Vocabulary Services (EVS) is a set of vocabulary resources that are being used to provide a common terminology across the cancer research domain, improving communication between the people, organizations, data and data sources in it.

NCI CBIIT's clinical trials informatics efforts draw on EVS whenever domain terminology is required. This reliance on controlled vocabulary ensures that term meanings remain clear and can be related to other terms with similar concepts.

The EVS Project is a collaborative effort of the Center for Bioinformatics and the NCI Office of Communications. EVS includes the NCI Thesaurus, the NCI Metathesaurus, and other Distributed Terminology System (DTS) Vocabularies, including:

- **NCI Thesaurus:** a biomedical thesaurus created specifically to meet the needs of the NCI, is produced by the NCI EVS project. This knowledgebase contains the working vocabulary used in NCI data systems. It covers clinical, translational and basic research as well as administrative terminology.
- **GO:** GO Ontology, published by the Gene Ontology Consortium.
- **VA NDFRT:** Published by the US Veterans' Administration, the National Drug File Reference Terminology covers clinical drugs used at the VA.
- **LOINC:** Published by the Regenstrief Institute, the Logical Observation Identifier Names and Codes cover clinical laboratory terminology.
- **MGED Ontology:** The MGED (Microarray Gene Expression Data) Ontology, produced by the MGED Society, contains concepts, definitions, terms, and resources for standardized description of a microarray experiment in support of MAGE v.1. The MGED ontology is divided into the MGED Core ontology which is intended to be stable and in synch with MAGE v.1; and the MGED Extended ontology which adds further associations and classes not found in MAGE v.1.
- **MedDRA:** MedDRA (Medical Dictionary for Regulatory Activities) is an international medical terminology developed under the auspices of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).
- **SnoMed:** SnoMed is the universal health care terminology maintained by SNOMED® International, a division of the College of American Pathologists (CAP).

The EVS Project also produces the **NCI Metathesaurus**, which is based on NLM's Unified Medical Language System Metathesaurus supplemented with additional cancer-centric vocabulary. The public version of NCI Metathesaurus currently contains all public domain vocabularies from the National Library of Medicine's UMLS Metathesaurus, as well as a growing number of NCI-specific vocabularies developed by the National Cancer Institute.

EVS is composed of:

NCI Thesaurus	NCI Metathesaurus
http://nciterms.nci.nih.gov	http://ncimeta.nci.nih.gov
Stand-alone reference terminology	Relational: Links to multiple terminologies
One definition for research	One or more definitions from multiple sources
Designed for annotation and database coding to facilitate data analysis and retrieval	Designed for mapping cancer terms across terminologies throughout the cancer research community to facilitate integration

- Users drive the content of the NCI Thesaurus.
- Subject Matter Experts (SMEs) model, develop, and test the content.
- Outside reviewers critique and authorize (give credibility) to the content.

2.3 Metadata

Regardless of the application domain, any particular data variable must have associated with it a variable name or tag, a conceptualization of what the item signifies a value, and an intended interpretation of that value.

For example, an entry on a case report form may be intended to capture the patient's place of birth, and the corresponding value may be tagged electronically as Patient_placeOfBirth. But what is the intended concept? Is the data element designed to capture the country, the city, or the specific hospital where the person was born? Assuming that the intended concept is country how is the resulting value to be represented electronically? Possible representations might include the full name of the country, a standard two- or three-letter abbreviation, a standard country code, or perhaps a specific encoding unique to the application.

Metadata is "data about data," and refers to just this type of intentional information that must be made explicit in order to ensure that electronically exchanged data can be correctly interpreted. The purpose of the ISO/IEC 11179 standard is to define a framework and protocols for how such metadata can be specified, consistently maintained, and shared across diverse domains.

Note: You may see and hear the terms Common Data Element (CDE) and Data Element (DE) used interchangeably throughout the caCORE curriculum. Though these terms are synonymous, a CDE generally refers to data elements that are common (e.g. have a high level of reuse) across multiple contexts.

2.3.1 Metadata Example: Person Address Zip Code

An example of metadata can be formed by analyzing a question from a Case Report Form (CRF). “What is your zip code?” is a question taken from a sample CRF. This question can be divided into data and metadata as shown in Figure 1 below.

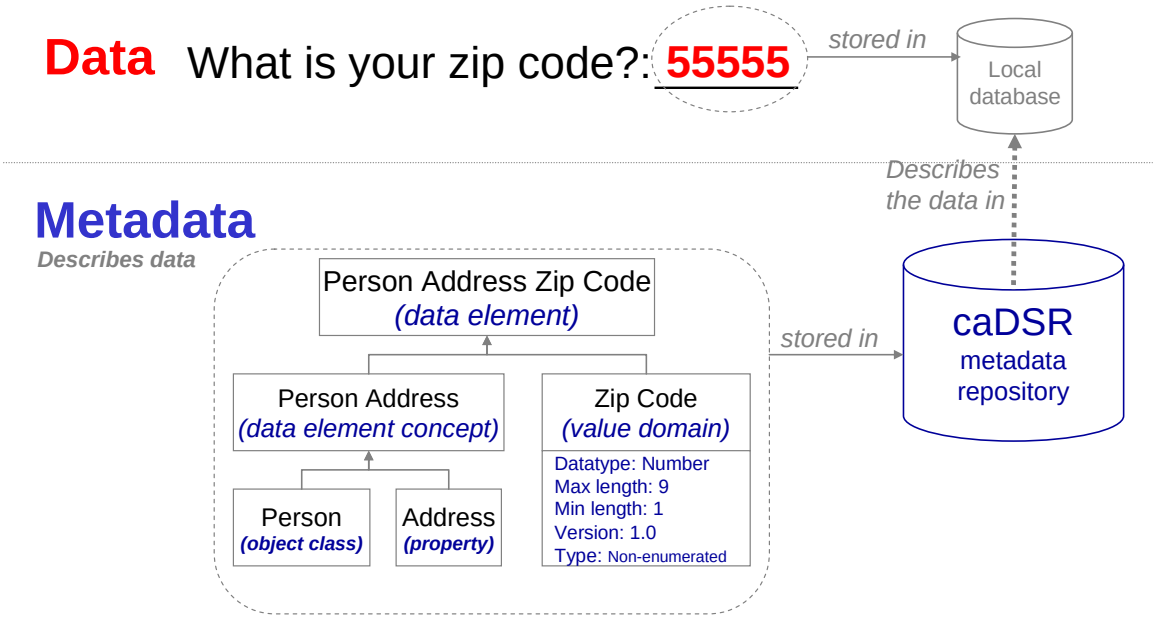

Figure 1. Metadata Example - Person Address Zip Code

The actual data, the sample answer of 55555, is stored in a local database.

The metadata is what you mean by zip code - the number representing a postal area for a person (in the USA this is a five or nine digit number). The metadata also includes how you want to represent that information - as a number, with a maximum length of 9 and a minimum length of 1.

Metadata is composed of various items. Metadata at the highest level is a Data Element (DE). A DE is composed of a Data Element Concept (DEC) and Value Domain (VD). A DEC is composed of an Object Class and Property. For the example of the CRF question above, “What is your zip code?” the metadata is defined as the following:

- Object Class = What am I describing? (Person)
- Property = What characteristic(s) am I describing? (Address)
- DEC = Object Class + Property (Person Address)
- Value Domain = How is the answer represented? (Zip Code)
- Data Element = DEC + VD (Person Address Zip Code)

The metadata describes the data in the local database and is stored in a metadata repository.

2.3.2 Data Element Fundamentals

As stated briefly above, A Data Element is the unique pairing of a Data Element Concept (DEC) and a Value Domain (VD). A DEC is composed of an Object Class and Property. Figure 2 below shows a visual representation of Data Element fundamentals.

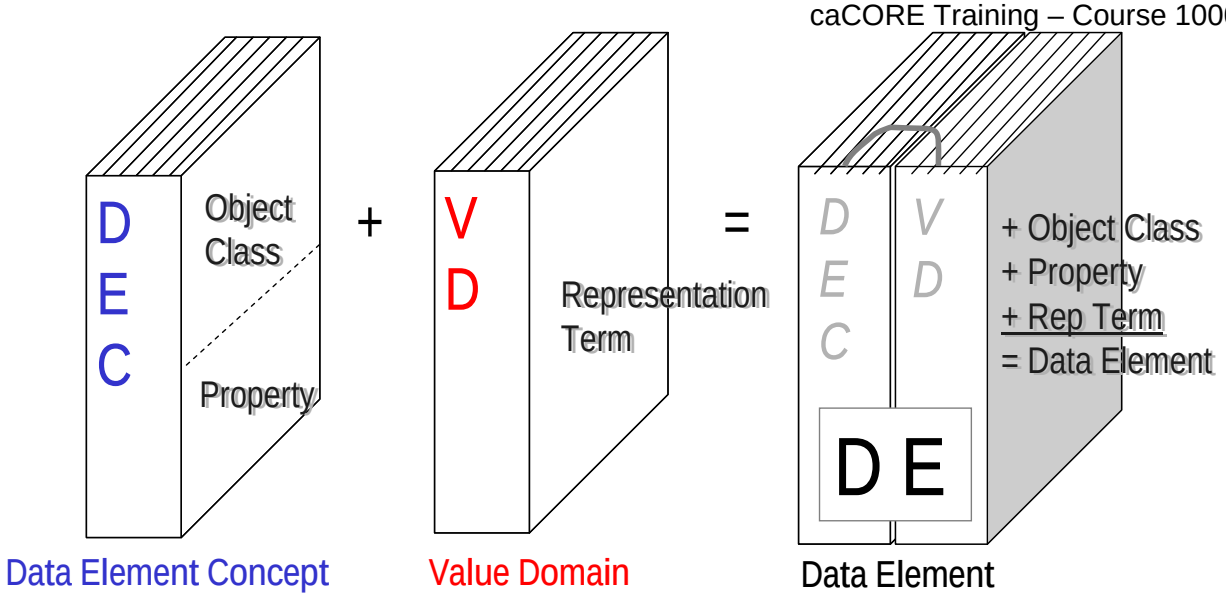

Figure 2. Data Element Fundamentals

Figure 3 below represents an example for Person Address Zip Code.

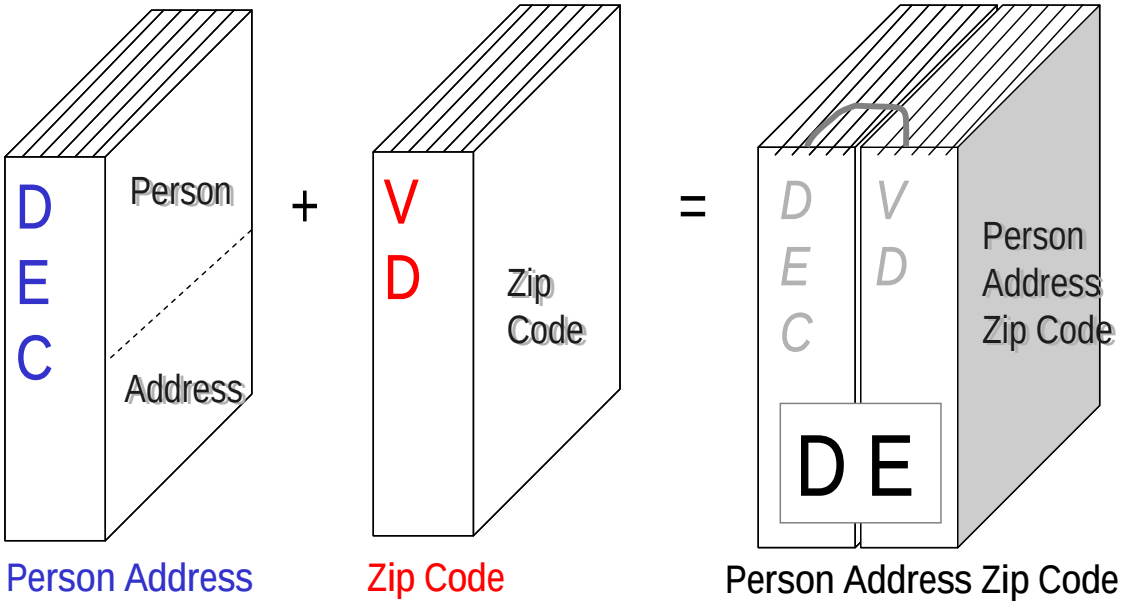

Figure 3. Data Element Fundamentals - Person Address Zip Code

Think of Data Elements and their components as a library of re-usable components. Figure 3 above shows that the DEC of “Person Address” and the VD of “Zip Code” form the DE of “Person Address Zip Code”. Any of these components, the Object Class “Person”, Property “Address” and Value Domain “Zip Code” can be used to form other Data Elements.

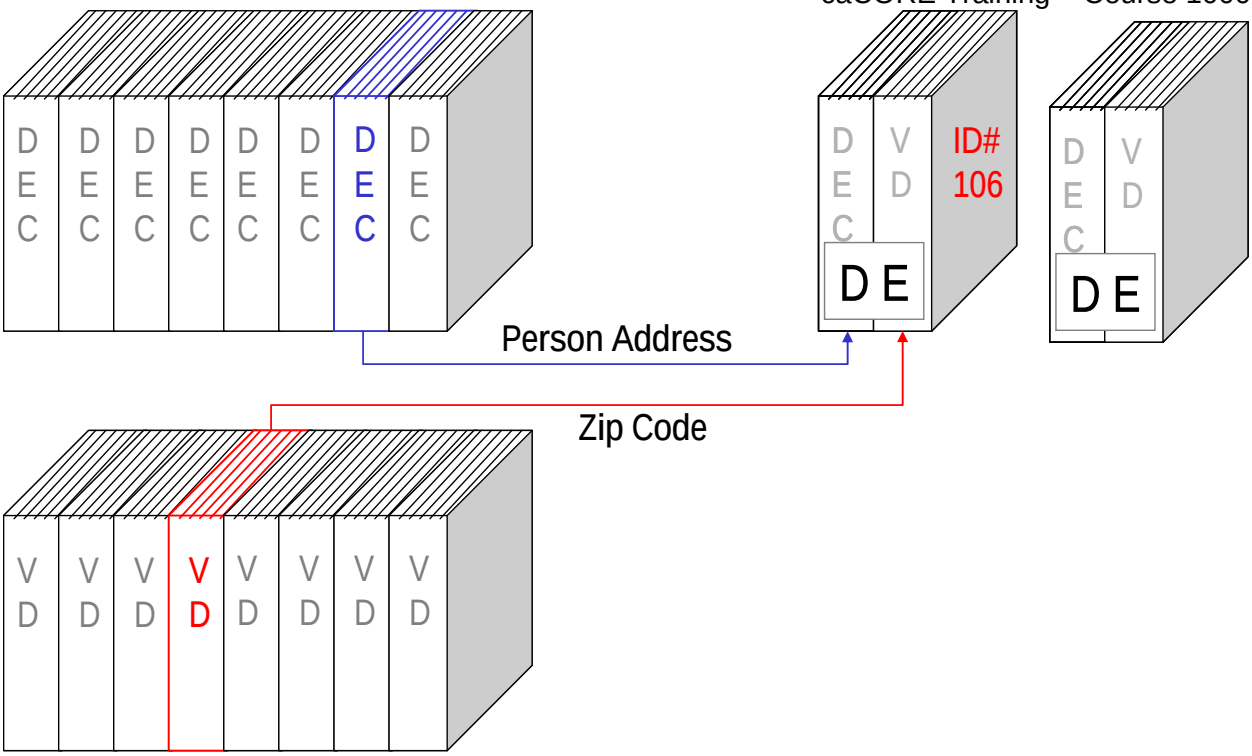

Figure 4. Libraries of Re-Usable Components – Person Address Zip Code

A goal of the caDSR is to create metadata using existing components as often as possible. By re-using the DEC of “Person Address” and combining it with a VD of “State Code”, a new DE of “Person Address State Code” is formed. This re-using of “Person Address” is displayed in Figure 5 below.

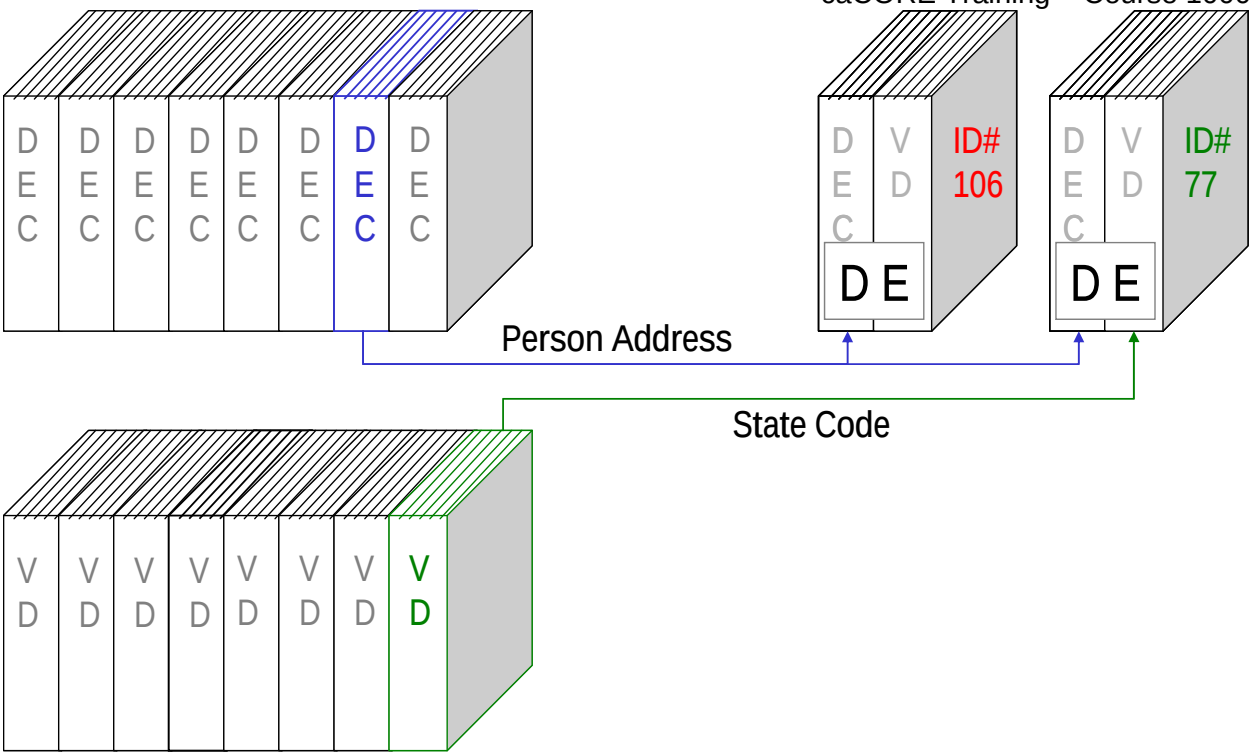

Figure 5. Libraries of Re-Usable Components - Person Address State Code

2.4 How Data Elements are Used

Data Elements are used many ways, some of which include:

- On forms (Case Report Forms or CRFs) for data collection
- In Databases to describe database field attributes and constraints
- Information/UML (Unified Modeling Language) Modeling
- APIs (Application Programming Interfaces)
- Projects
- In applications to describe UI (User Interface) widgets, validation rules, display name and format

3 Lesson 2: caCORE Infrastructure

caCORE is composed of caBIO, the caDSR and EVS. This lesson will focus on each of these components.

3.1 Objectives for the Learner

At the completion of this lesson, you will be able to:

1. Describe the Cancer Data Standards Repository (caDSR)
2. Introduce ISO/IEC 11179 Metadata Standard
3. Describe the major caCORE Components
4. Describe the caCORE Infrastructure
5. Discuss the characteristics of a caCORE-like system

3.2 Cancer Data Standards Repository (caDSR)

The caDSR is a metadata repository and registry designed to integrate caCORE infrastructure. caDSR supports the development and deployment of Data Elements used as metadata descriptors, initially for NCI-sponsored research, but now for an ever-widening group of users.

Common Data Elements (CDEs) based on EVS thesauri and standard vocabularies are developed by various NCI-sponsored organizations, including caBIG, and are centrally stored and managed at NCI CBIIT in the Cancer Data Standards Repository. The caDSR is an Oracle database. caDSR is comprised of a suite of tools for creating, sharing and deploying CDEs including a public CDE Browser that lets you search for data elements, create forms and download CDEs. All of the various tools and interfaces connect to the same central Oracle database.

3.3 ISO / IEC 11179

For metadata to be useful, it must be accessible to applications at runtime. For this reason, the NCI CBIIT developed the caDSR to store metadata, based on the ISO/IEC 11179 metamodel. This model describes a wide range of characteristics of Data Elements including definitions, permissible values, data type, unit of measure, minimum and maximum length, etc.

caDSR leverages the ISO/IEC 11179 Information Technologies: Metadata Registries (MDR) standard to harmonize, register and integrate user defined UML information models with existing and new caDSR content.

In the ISO/IEC 11179 model, a Data Element consists of two parts:

1. A **Data Element Concept** that provides the conceptual definition of the data element, the primary semantics or meaning of the Data Element.
2. A **Value Domain** that describes specific acceptable values for that data element. Value domains can be either 1) enumerated with an explicit list of permissible values, or 2) non-enumerated, restricting the values to a description, specification or rule. Attributes of the Value Domain include data characteristics such as the data type and unit of measure.

The parts of the caDSR implementation of the ISO/IEC 11179 model, Object Class, Property, and Value Domain are controlled vocabulary terms maintained by the EVS. Thus the caDSR provides a link between a data element (such as an attribute in an object model) and definitions in a controlled vocabulary. A Data Element Concept is represented by a combination of at least two EVS concepts, an Object Class and a Property, each of which may have qualifiers that are also EVS terms. Similarly, the Value Domain has at least a representation that is the form in which the value is being recorded. The representation could be ‘Currency’, ‘Number’, ‘Code’, etc. The representation is intended to convey information in addition to the datatype. If the value domain is enumerated, the list of valid values may come from EVS as well the value meaning associated with each valid value.

3.4 caCORE Components

This section contains non-essential information

NCI CBIIT provides biomedical informatics support and integration capabilities to the cancer research community. NCI CBIIT has created a core infrastructure referred to as the Cancer Common Ontological Research Environment or caCORE, a data management framework

designed for researchers who need to be able to navigate through a large number of data sources. caCORE is NCI CBIIT's platform for data management and semantic integration which is built using formal techniques from the software engineering and computer science communities. caCORE provides web services to access caDSR content programmatically. The caDSR, based on open source standards, is freely available for use by other government agencies and for download and use by interested parties. caCORE is the open-source foundation upon which the NCI CBIIT builds its research information management systems.

The main components of caCORE, created and deployed by NCI CBIIT, include Cancer Bioinformatics Infrastructure Objects (caBIO), Cancer Data Standards Repository (caDSR), Enterprise Vocabulary Services (EVS), Common Security Model (CSM) and the Common Logging Module (CLM).

Cancer Bioinformatics Infrastructure Objects (caBIO) is a set of JavaBeans with open source APIs that can be used to directly access bioinformatics data. Unified Modeling Language™ (UML) models of biomedical objects are implemented in Java as middleware connected to various cancer research databases to facilitate data integration and consistent representation.

- For additional information related to caBIO, please visit http://ncicb.nci.nih.gov/core/caBIO .

Cancer Data Standards Repository (caDSR) is a metadata registry based upon the ISO/IEC11179 standard that is used to register the descriptive information needed to render cancer research data reusable and interoperable. The caBIO, caMOD, EVS and caDSR data classes are registered in the caDSR, as are the data elements on NCI-sponsored clinical trials case report forms. caDSR provides, among other things, a semantic bridge between the data elements in registered data objects and standard vocabularies and ontologies.

- For additional information related to the caDSR, please visit http://ncicb.nci.nih.gov/core/caDSR .

Enterprise Vocabulary Services (EVS) are controlled vocabulary resources that support the life sciences domain, implemented in a description logics framework. EVS vocabularies provide the semantic 'raw material' from which data elements, classes, and objects are constructed.

- For additional information on EVS, please go to http://ncicb.nci.nih.gov/core/EVS

Common Security Model (CSM) provides a flexible solution for security and access control with three main functions: Authentication to validate and verify a user's credentials, Authorization to grant or deny access to data, methods, and objects and User Authorization Provisioning to allow an administrator to create and assign authorization roles and privileges.

Common Logging Module (CLM) provides a separate service under caCORE for Audit and Logging Capabilities. It also comes with a web based locator tool. It can be used by a client application directly without the application using any other components like CSM.

NCI CBIIT-developed caCORE components are distributed under open-source licenses that support unrestricted usage by both non-profit and commercial entities, and are downloadable from the NCI CBIIT web site.

3.5 caCORE Infrastructure

This section contains non-essential information

As previously stated, caCORE is composed of caBIO, the caDSR and EVS. Figure 6 below illustrates the infrastructure of these components.

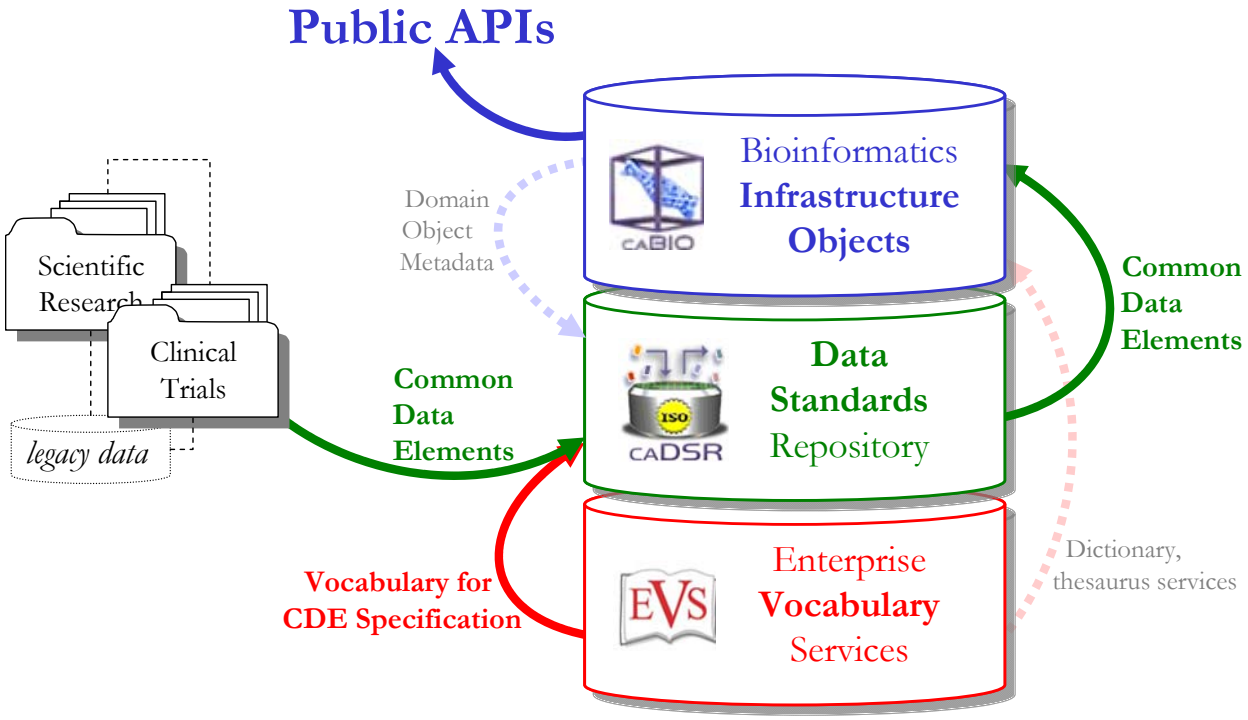

Figure 6. caCORE Components

Starting on the left side of Figure 6, data from clinical trials and scientific research is transformed into descriptive metadata and represented in the caDSR. EVS provides access to standard vocabularies and ontologies curated by external standards development groups. Vocabularies are used in naming and defining the metadata in the caDSR. caBIO Model Driven Architecture (MDA) allows users to easily generate public APIs to access caDSR.

3.6 caCORE Characteristics

This section contains non-essential information

A system is said to be “caCORE-like” when it has the following characteristics:

- Model-Driven Architecture (MDA)
- N-tier Architecture
- Controlled Vocabularies
- Registration of Metadata

Model-Driven Architecture (MDA) is an approach to software development that has been standardized by the Object Management Group (OMG). MDA leverages UML to allow developers to build platform-independent models (PIMs) which, through a series of transformations, are converted to platform-specific models (PSMs) and subsequently to executable code.

n-tier Architecture is the idea of open, standard APIs that are heavily rooted in object-orientation. The Object Oriented (OO) paradigm is tightly coupled to MDA and allows us to take advantage of all of the benefits that these two approaches provide, including information hiding, encapsulation, abstraction, inheritance, etc.

In order to implement an object-oriented system, the caCORE Software Development Kit (SDK) uses a variety of technologies among which is object-relational mapping. The ORM tool implemented by the SDK, Hibernate, provides rich functionality that can be controlled at a very granular level by the developer without having to get caught up in complex implementation details.

Controlled Vocabularies yield semantic interoperability at runtime. This is the ability to determine, at runtime, the context of data that is returned by a data or analytical service. This is critical to future grid architectures that will need to perform federated queries across multiple services and will need to determine the nature of the data on the grid. Examples of controlled vocabularies include SNOMED, MedDRA, GO Ontology and the NCI Thesaurus. All of these are served by the Enterprise Vocabulary Services (EVS).

Registration of Metadata promotes semantic interoperability and caGRID compliance. The caDSR is an ISO/IEC 11179 compliant metadata repository. The caDSR training courses are an excellent way to learn about caDSR and how to use it to create, curate and browse data elements.

4 Lesson 3: caDSR Tools

4.1 Objectives for the Learner

At the completion of this lesson, you will be able to:

1. Describe the purpose of caDSR Tools
2. Demonstrate and discuss the functionality of:
   - CDE Browser
   - Form Builder
   - CDE Curation Tool
   - caDSR Sentinel Tool
   - caDSR Admin Tool

4.2 Purpose of the caDSR Tools

caDSR Tools are designed to:

- Create, consume, distribute and promote ISO/IEC 11179 compliant metadata
- Enable semantic consistency across research domains
- Support the metadata life-cycle and governance processes

By ISO/IEC compliant, we mean that our implementation of the standard, which includes NCI CBIIT specific extensions, still maps to the mandatory and optional metadata items prescribed by ISO/IEC 11179, Part 3. There are various levels of conformance, because of these business needs-related extensions, caDSR is considered “conforming” as opposed to “strictly conforming”.

4.3 Tools Overview

The sections below will introduce the caDSR tools at a high level. Subsequent training sessions will focus on more detail and features of each application.

4.3.1 CDE Browser http://cdebrowser.nci.nih.gov

The CDE Browser should be the starting point for exploring the contents of the caDSR. The CDE Browser allows users to:

- Search for Data Elements (basic, advanced, tree)
- Compare Data Elements
- Download Data Elements (in Excel/XML)

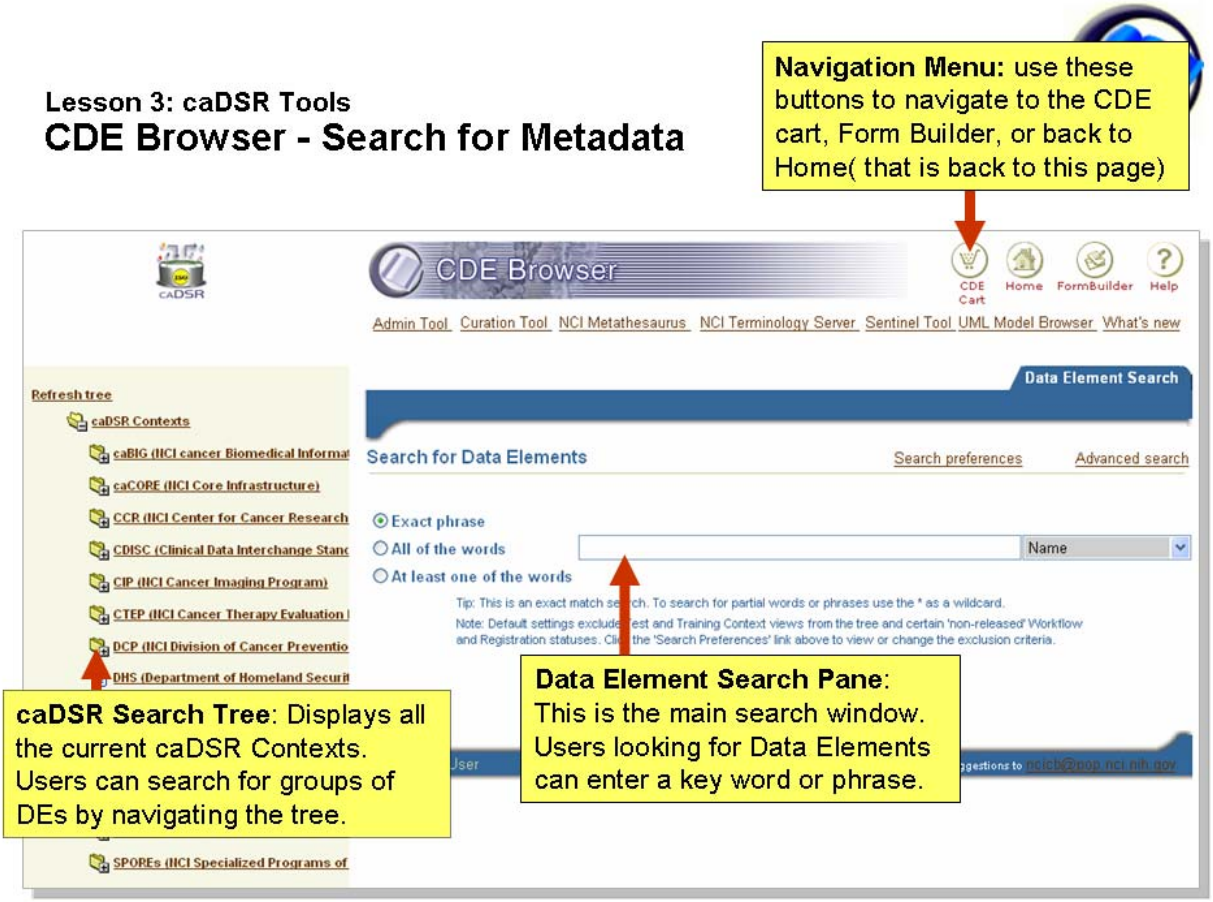

Figure 7 - CDE Browser Main Search Screen

The CDE Browser supports browsing, searching, and exporting CDEs within or across contexts, (metadata-data owning organizations). The CDE cart feature provides users the ability create a customized set of CDEs for exporting. The Form Builder feature lets users organize a customized set of CDEs from the cart as "forms". The CDE Browser also allows users to export/download data elements to XML or Excel.

4.3.2 Form Builder (link from CDE Browser)

The Form Builder allows users to:

- Create Forms using existing Data Elements
- Share Forms with user community
- Download Data Elements associated with Forms
- Print Forms

The Form Builder requires curatorial authority; a caDSR account must be assigned for accessing Browser Form Builder feature. It may be the same account assigned for accessing the Curation Tool and the Administration Tool.

Please skip this section

4.3.3 CDE Curation Tool http://cdecurate.nci.nih.gov

The Curation Tool, intended for Context Administrators, allows users to:

- Create or Edit individual Data Element Concepts (DECs), Value Domains (VDs) and Data Elements (CDEs)
- Block Edit groups of Data Element Concepts (DECs), Value Domains (VDs) and Data Elements (DEs). Components that can be block edited include:
  - Classification Scheme(s)/Classification Scheme Item(s), Data Element Concept, Value Domain, Version, Workflow Status, and Registration Status
- Search for Associated Administered Components

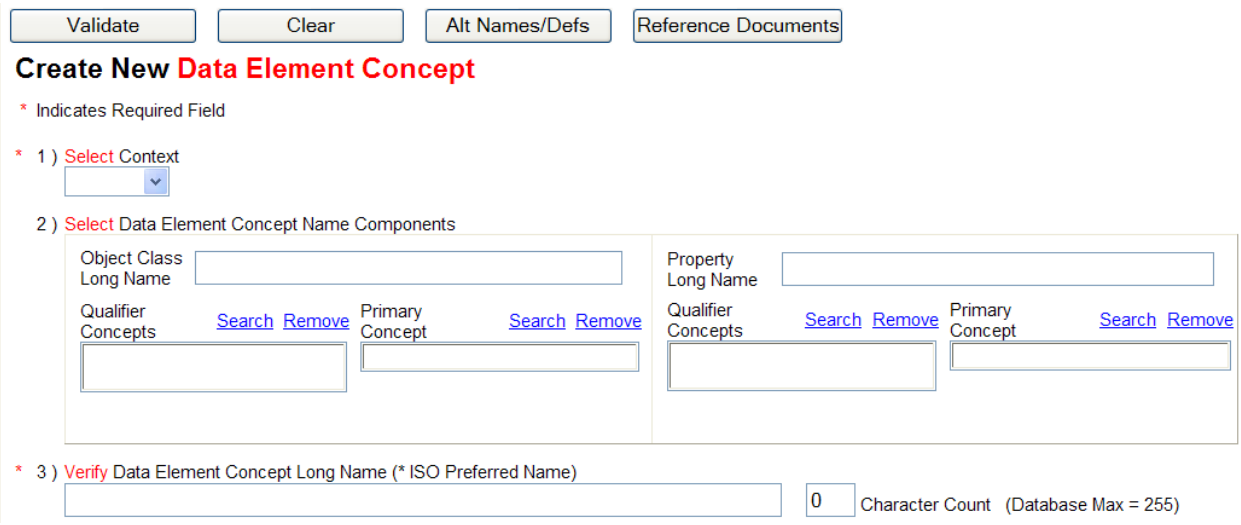

Figure 8 - Curation Tool: Create New Data Element Concept Screen

An account is required to use the Curation Tool, and the user must have curatorial authority for a specific Context. This tool's features facilitate the use of the NCI CBIIT Enterprise Vocabulary System (EVS) to create administered item names and definitions using a 'naming wizard'. This ensures ISO/IEC 11179 compliance and caDSR naming conventions are reached. The "Get Associated" feature allows the user to easily navigate related components such as finding all the DEs associated with a specific Object Class.

Please skip this section

4.3.4 Sentinel Tool http://cadsrsentinel.nci.nih.gov

The caDSR Sentinel Tool allows users to:

- Create and manage Alert Definitions for the caDSR
- Monitor all changes to Administered Components (including Data Elements, Data Element Concepts, Value Domains, Object Classes, and Properties)
- Filter the Alerts by Context, specific Forms or Templates, Classification Schemes / Classification Scheme Items, Creator and Modifier

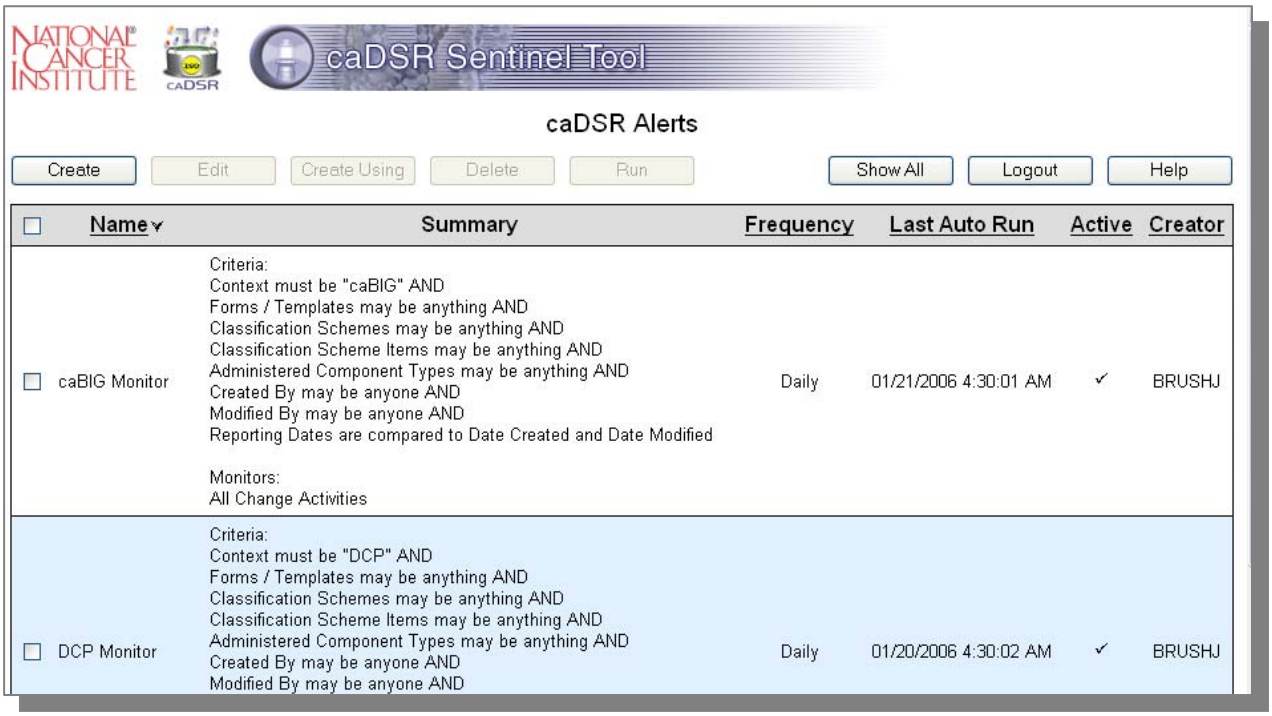

Figure 9 - Sentinel Tool: List of Active Alert Definitions

Alert Definitions are a set of rules that are periodically evaluated against the caDSR. If the conditions in those rules are met, notification is sent to the user (via email) in the form of a hyperlink to an activity report that specifies the changes that have taken place. Reports may be set to run automatically daily, weekly or monthly; on demand; sent only to the Alert owner or to a distribution list.

4.3.5 Admin Tool Please skip this section

The Admin Tool is the main administrative interface to all of the caDSR features and components. The Admin Tool allows users to:

- Create Conceptual Domains, Classification Schemes, and Protocols
- Establish the Context Specific Ontology to be used in curation: Create Object Class, Property, Representation Terms, Conceptual Domains, Classification Schemes
- Perform System Administration (limited users will have privileges)

The Admin Tool is intended for use by both Context and central caDSR administrators. This tool also provides access to EVS for creating content.

There are a number of administrative and curatorial tasks that are not supported in the CDE Curation Tool, such as creating Conceptual Domains, Classification Schemes and Protocols

used across all caDSR tools. These must be performed through the caDSR Administration Tool. An account and appropriate privileges are required to use this tool.

5 Review Questions

1. The caDSR is based on the _____ metadata standard.
   a. ISO/IEC 11179
   b. HIPPA
   c. caCORE

2. Metadata is defined as:
   a. Data larger than data
   b. Data about data
   c. Data smaller than data
   d. Data about metaphysics

3. The major components of caCORE include EVS, caDSR Data Standards and _____ Bioinformatics Objects.
   a. caBIO
   b. caBIG
   c. caDSR

4. _____ can be used on forms for data collection.
   a. Data Elements
   b. Objects
   c. caDSR

5. The caDSR Sentinel Tool is used to _____ changes to caDSR metadata.
   a. Monitor
   b. Make
   c. Curate

6. The CDE _____ has a CDE cart for users to collect a variety of CDEs from a diverse set of searches.
   a. Browser
   b. Curation Tool
   c. Admin Tool

7. The CDE _____ can be used to download metadata from the caDSR.
   a. Browser
   b. Curation Tool
   c. Admin Tool

8. The caDSR _____ is the main administrative interface to all caDSR features and components.
   a. Browser
   b. Curation Tool
   c. Admin Tool

9. caCORE is _____ and can be used by anyone.
   a. Open Source
   b. Private
   c. Password protected

10. Common Data Elements (CDEs) in the caDSR are based on EVS _____ and standard vocabularies.
   a. Thesaurus
   b. Metathesaurus
   c. caBIO

11. A Data Element Concept is the unique pairing of an Object Class and _____.
   a. Value Domain
   b. Data Element
   c. Property

12. A Data Element is the unique pairing of a _____ and a Value Domain.
   a. Data Element Concept
   b. Conceptual Domain
   c. Property

13. An easy way to search for groups (protocols or forms) of Data Elements in the CDE Browser is to use which tool feature?
   a. Basic Search
   b. Tree Search
   c. Term Search

14. For the Data Element “Patient Race Category” what is a likely candidate for the Data Element Concept?
   a. Patient
   b. Race
   c. Patient Race
   d. Category

15. For the Data Element “Patient Race Category” what is a likely candidate for the Value Domain?
   a. Patient
   b. White
   c. Patient Race
   d. Category

16. For the Data Element describing when a patient was born, what is a likely candidate for the Value Domain?
   a. Country
   b. Year
   c. Time
   d. City

17. The Data Element Concept of “Patient Ethnic Group” and the Value Domain of “Ethnic Group Category” together form a Data Element called:
    a. Patient Ethnic Group
    b. Patient’s Ethnicity
    c. Patient Ethnic Group Category
    d. Patient Ethnic Category

18. What is the most complete Data Element for the marital status category of a person?
    a. Person Marital Status
    b. Marital Status
    c. Person Marital Status Category
    d. Marital Status Category

19. The Data Element Concept of “Person Self Report Age” and the Value Domain of “Age Value” together form a Data Element called:
    a. Person Age
    b. Person Self Report Age
    c. Person Self Report Age Value
    d. Self Report Age Value

20. The Data Element “Patient Race Specify” is most likely constructed from:
    a. DEC: Patient Race VD: Specify
    b. DEC: Patient VD: Race
    c. DEC: Race VD: Patient
    d. DEC: Specify VD: Race

6 Documentation / Recommended Reading Materials

Below is a list of links to documentation used to create this session and of recommended reading materials.

caDSR Homepage:
- http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr

caDSR Training Home Page:
- http://ncicb.nci.nih.gov/NCICB/training/cadsr_training

caCORE Developer’s Guide:
- https://gforge.nci.nih.gov/docman/view.php/148/8650/caCORE%20SDK%204.0%20Developer's%20Guide_101007.pdf

caCORE Programmer’s Guide:
- ftp://ftp1.nci.nih.gov/pub/cacore/SDK/v3.2.1/caCORE_SDK_3.2.1_Programmers_Guide.pdf

caDSR Business Rules:
- http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr/business_rules/

caDSR_Users ListServ Subscription:
- https://list.nih.gov/archives/cadsr_training-l.html
- http://list.nih.gov

Send Request for caDSR Account to:

- ncicb@pop.nci.nih.gov

7 Contact Information

NCI CBIIT Application Support:		ncicb@pop.nci.nih.gov
Instructor:	Jennifer Brush	brushj@mail.nih.gov
NCI CBIIT Liaison:	Dianne Reeves	reevesd@mail.nih.gov