caDSR - Overview of Use and Collaborations

NCI Cancer Data Standards Registry and Repository (caDSR) is a centralized resource for services and web-based information technology tools to meet NCI needs for documenting and sharing unambiguous descriptions of data. These resources are used to develop standards-based IT descriptions for trials and research data shared by NCI and the larger international biomedical community, so data can be effectively and programmatically exchanged, aggregated, analyzed, used and reused for secondary research.

Because of the size and complexity of the caDSR database, the NCI provides the hosting environment for the caDSR content infrastructure. As such, caDSR data descriptions, software, and services are made freely available where possible, to all organizations who want to use them.

While this approach promotes broader adoption, it also hinders full tracking of caDSR use. Similarly, restrictions on identifying and surveying users also hinders the ability to create a complete assessment of how caDSR is being used and by whom. However, we believe that like EVS, the caDSR adopters and patterns of use are fairly clear.

This overview of use of and collaborations with the Cancer Data Standards Registry and Repository attempts to identify the primary adopters and define and interpret the patterns of use of caDSR and its associated tools, based on the information and statistics available to us.

This document was initially prepared on October 8, 2010 and is updated on an ongoing basis. It is made up of the following sections:

1. Summary
   - Introduction
   - caDSR Infrastructure and Toolset Usage
   - Use of caDSR by NCI and Organizations
   - caDSR Use by Other NIH Organizations
   - External Adoption of caDSR

2. caDSR Infrastructure and Toolset Usage
   - Overview
     - NCI caDSR CDE Browser
     - NCI caDSR Form Builder
     - NCI caDSR Curation Tool
     - NCI cgMDR Excel Addin, Template and Bulk Loader
     - NCI caDSR UML Model Browser
     - NCI caDSR Sentinel Tool
     - NCI caDSR Database Server, Domain Model and Freestyle APIs
     - NCI caDSR Semantic Integration Workbench (SIW) and UML Loader

4. NIH Usage of caDSR
   - National Heart Lung and Blood Institute (NHLBI)
   - National Institute of Dental and Craniofacial Research (NIDCR)
   - National Institute of Child Health and Development (NICHD)
   - National Institute of Neurological Diseases and Stroke (NINDS)

5. External Uses of caDSR
   - Biopathology Center, National Childrens’ Hospital
   - PhenX Toolkit
   - GIATE
   - American Heart Association (AHA) and American College of Cardiology Foundation (ACCF)
   - Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences
   - Union of Light Ion Centers in Europe (ULICE)
   - Winthrop P. Rockefeller Cancer Institute, Cancer Control Core at the University of Arkansas for Medical Sciences; Breast Mammography
   - Oncology Patient Enrolment Network (OPEN) - Sponsor-CTEP/NCI
   - Duke Translational Medicine Institute
   - Tiwan Cancer Registry
   - Children's Oncology Group (COG) and National Childhood Cancers Foundation
   - University of Miami Ontology Modeling
   - Center for International Blood and Marrow Transplant Research (CIBMTR)
   - The Human Studies Database (HSDB)
   - Towards Unambiguous formal descriptions of cancer therapy experiments
   - Ontology based queries for caGrid Infrastructure
   - Information Management Services, Inc. (IMS) BIS, NAACCR and DCCPS/caSEER program
   - The CTSA Health Ontology Mappe
   - A Growable Network Information System (AGNIS)
   - cancerGrid, cgMDR UK
1 - Summary

NCI Cancer Data Standards Registry and Repository (caDSR) is a centralized resource for services and web-based information technology tools to meet NCI needs for documenting and sharing unambiguous descriptions of data. These resources are used to develop standards-based IT descriptions for trials and research data shared by NCI and the larger international biomedical community.

The caDSR - Overview of Use and Collaborations document attempts to identify the primary adopters of caDSR, and to define and interpret the patterns of use of caDSR and its associated tools, based on the information and statistics available to us.

This Executive Summary is designed to provide a summation of the more detailed information contained in the full document. This summary contains a few stories of caDSR adoption by several representative organizations and the clinical research models they are using, as well as an overview of the usage and statistics of caDSR and its associated toolset.

This Executive Summary contains the following sections:

- Introduction
- caDSR Infrastructure and Toolset Usage
  - CDE Browser - Usage Summary
  - Form Builder - Usage Summary
  - CDE Curation Tool - Usage Summary
  - API Access to caDSR - Usage Summary
  - Semantic Integration Workbench and UML Model Loader - Usage Summary
- Use of caDSR by NCI and Organizations
  - Cancer Therapy Evaluation Program (CTEP)
  - Cancer Centralized Clinical Database (C3D)
  - Division of Cancer Prevention (DCP)
  - Specialized Programs for Research Excellence (SPOREs)
  - National Marrow Donor Program (NMDP)
- caDSR Use by Other NIH Organizations
  - National Institute of Dental and Craniofacial Research (NIDCR)
  - National Heart Lung and Blood Institute (NHLBI)
  - National Institute of Neurological Diseases and Stroke (NINDS)
- External Adoption of caDSR
  - Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences
  - Union of Light Ion Centers in Europe (ULICE)
  - Center for International Blood and Marrow Transplant Research (CIBMTR)
  - Information Management Services, Inc. (IMS) BIS, NAACCR and DCCPS/caSEER program
  - A Growable Network Information System (AGNIS)
  - cancerGrid and cgMDR (United Kingdom)

Introduction

Since 2000, the NCI's Cancer Data Standards Registry (caDSR) and its associated toolset have become the primary means by which cancer researchers are sharing descriptions of databases, and the means through which clinical trials data managers are describing the data collected in clinical and research studies. Specifically, caDSR is being used:

- By NCI and its partners in NIH, and other federal agencies, and by U.S. and international biomedical organizations, academia, and cancer researchers, for documenting and publishing human and computer interpretable representations of standard datasets, including those for cancer registries. Examples include Cancer Therapy Evaluation Program (CTEP), National Marrow Donor Program (NMDP), North American Association of Central Cancer Registries (NAACCR) and Surveillance and Epidemiology and End Results (SEER), International Agency for Research on Cancer (IARC) and Taiwan Cancer Registry as well as those used in clinical trials, such as Cancer Centralized Clinical Database (C3D).

- To document, harmonize, access, and publish data descriptions bound to standard value sets and biomedical terminologies published by EVS, such as Biomedical Research Integrated Domain Group (BRIDG) 3.0.2;
- To identify reusable data descriptions to create new research software that collect data that can be aggregated in support of new basic research, translational research, and clinical research, clinical care, epidemiology, public health administration, and public information, such as with caTissue, caIntegrator, the Human Studies Database (HSDB), TCGA, Eastern Cooperative Oncology Group and Union of Light Ion Centers in Europe (ULICE);

caDSR provides the foundational layer for representing the precise meaning and representation of data, and through linkage to the Enterprise Vocabulary System (EVS) terminology and the ISO/IEC 11179 standard, the ability to translate that meaning throughout the informatics infrastructure and across various research domains.

caDSR Infrastructure and Toolset Usage

The caDSR software was designed by Oracle and the Census Bureau in the late 1990's as a centralized repository for ISO/IEC 11179 content. Due to its dependency on an Oracle database and an application server for the Admin Tool and database, it is not well suited to installation at small cancer centers who are resource-constrained, or lack the Oracle skills to get the suite of products installed, customized, and working. To solve this problem, the NCI hosts content for those who want to use the infrastructure. This allows the NCI to help manage and track the data elements within the caDSR and the communities and applications that use them.

As such, one measure of caDSR usage is the number of data elements recorded for the purpose of specifying the detailed descriptions of data, and enabling sharing and reuse of these descriptions among data managers and software developers. caDSR usage can also be measured by the number of communities and applications that have registered their data elements and models in caDSR and are retrieving them via browser downloads and use of caDSR APIs.

As caDSR functions as the repository for the metadata being loaded and used, the caDSR tools function as the interface points for users and applications to access the information contained in the repository.

The sections that follow provide overview and usage information for the most commonly used of the caDSR tools.

More detailed information and statistics on all of the caDSR Tools can be found in Chapter 2 - NCI & Informatics Infrastructure - caDSR.

CDE Browser - Usage Summary

CDE Browser is the starting point for exploring the details of all data elements in development or registered in the caDSR. This tool was designed to support browsing, searching, and exporting CDEs in XML or Excel format. In addition, the CDE Browser works in conjunction with the Form Builder, to allow users to search for and select groups of CDEs to then place onto forms.

Usage Summary

Since statistics tracking started in 2005:

- An average of 50-60 unique users visit the CDE Browser each business day.
- An average of just over 41 pages viewed per visit for 2014.
- Top sites outside US include Canada, Brazil, Germany, China, Taiwan, and Great Britain.
- Top sites outside DC area include Raleigh-Durham, Minneapolis, Pittsburgh, Rochester, Charlottesville, Memphis, Portland, and Buffalo.
- NCI Cancer Therapy Evaluation Program (CTEP) currently has over 11,000 released and draft CDEs.
- The NCI Cancer Clinical Research (CCR) center has 3,491 selected CDEs that have been downloaded and imported into the Cancer Centralized Clinical Database (C3D). C3D has used these common descriptors to build over 900k instances of Case Report Forms (CRFs), and using those, have collected over 150M data points (responses).

Form Builder - Usage Summary

Form Builder allows users to organize select CDEs into forms that replicate the content of Case Report Forms (CRFs). Using the CDE Browser, you can search for CDEs, place them in a shopping cart, and then retrieve those CDEs through Form Builder, to insert them as questions on a form. As you place the CDEs on the form, the tool uses the stored metadata to automatically provide default question text and value domain information.

The NCI Cancer Therapy Evaluation Program (CTEP) is currently sponsoring over 600 trials that are using CDEs and CRFs for reporting trial results. In addition, commercial vendors Medidata (RAVE), Westat (for CTEP) and Eastern Oncology Centers Group-ACRIN (ECOG-ACRIN) retrieve CRFs from Form Builder (using the caDSR API) and import them to customize their data collection systems.

Usage Summary

The NCI has tracked Form Builder usage since January of 2009, though the tool has been available since 2002. Since 2009:

- On an average day there are 6 curators using Form Builder.
- Since Jan 2011 there have been 780 unique visitors.
- There are 3,352 Forms defined in caDSR, 2,936 "Released" owned by 7 NCI Divisions and NIH Institutes. These are organized by group, disease and type of trial, that describe the minimum datasets to be collected for various types of research.
  - CTEP has 2,613 CRFs in Form Builder.
  - There are 407 Protocols defined by CTEP in caDSR, into which many of the CTEP CRFs are grouped.

CDE Curation Tool - Usage Summary

The CDE Curation Tool supports the creation and editing of all of the primary semantic descriptions and value domain restrictions for Data Elements used by the community on case report forms, in data repositories or application software. It can be used to find content without logging in, but it is primarily intended for use by content curators. This tool facilitates the use of the CBII Enterprise Vocabulary Services (EVS) to create administered item names, definitions and value lists, helping to ensure reuse of existing content when matching ISO/IEC 11179 components are discovered, as well as promoting consistent use of caDSR naming conventions.
Usage Summary

Since statistics tracking started in 2005:

- An average of 22 unique users visit the Curation Tool each day.
- There have been a total of 4,558 unique visits to the Curation Tool site.
- Curation Tool has served over 1.6M pages.

One point to make is that the number of new CDEs year-over-year is trending downward. We believe this is due to several factors: active maintenance to retire unused content, and the increased reuse of existing CDEs, which is the primary purpose of caDSR. This increased reuse is made possible by curator focus on harmonization activities, supported by algorithm checks to find existing content that matches a curator's needs. A decrease in the number of new CDEs may also be due in part to the fact that caBIG® models are no longer required to be registered in caDSR. This means that a minimal number of models are being loaded, and only those CDEs deemed not already defined are being added to the repository. While limiting the number of new CDEs, this helps streamline the database and the content increase reuse.

See Appendix A - Usage Statistics for caDSR Resources for more detailed statistics on CDE creation, reuse, and harmonization.

API Access to caDSR - Usage Summary

All caDSR content is available through various application programming interfaces (APIs) as well as web service interfaces. For example, AGNIS uses the API to access Forms. Medidata (RAVE), Westat (for CTEP) and Eastern Oncology Centers Group-ACRIN (ECOG-ACRIN) retrieve Case Report Forms (CRFs) using the caDSR REST API, and import them to customize their data collection systems.

caDSR provides two API interfaces for access to the caDSR database:

- **caDSR API** - caDSR API allows users to access caDSR content programmatically using a caDSR Java API or a REST interface. There is also a web browser interface to this API called "Domain Model browser" that retrieves content as HTML or XML documents.
- **Freestyle API** - Freestyle API provides access to content both via a web browser and application programming interface. The Freestyle API actually uses the caDSR Domain API to simplify object retrieval.

**Usage Summary**

During the period October 2009 through October 2010:

- The caDSR API site had 628 unique visitors.
- The caDSR API served over 25M pages to users, resulting in an estimated 1.2M logical documents being served. The differential between the number of pages and the number of documents is due to the design of the caDSR Domain Model. Because of the design, multiple pages (between 6-20) must be served in order to get one logical document.
- The Freestyle API site had 328 unique visitor.
- The Freestyle API site served over 9,600 documents to users.

See Appendix A - Usage Statistics for caDSR Resources for a list of the top URLs that access the caDSR API.

Semantic Integration Workbench and UML Model Loader - Usage Summary

The Semantic Integration Workbench (SIW) assists users in adding semantic metadata to a UML model by tagging a class diagram representing a domain model with matching concepts from the NCI Thesaurus, or by attaching existing caDSR CDEs or Value Domains to attributes in the model. These annotations ensure reuse of CDEs or other metadata elements that already reside in caDSR. The UML Loader transforms the file and decomposes it into ISO/IEC 11179 descriptive metadata, then compares to existing content and creates or reuses metadata for the model in caDSR.

**Usage Summary**

As of March 2015:

- The UML Model Browser was retired October 2015 and CDEs related to models can still be retrieved using the CDE Browser.
- The SIW Loader is being used to batch load over CDEs from models and spreadsheets.
- caDSR contains metadata for 198 UML models (including multiple versions for some).
- The models loaded in caDSR use 26,360 CDEs, with many of the CDEs reused across models.
  - For example, if the CDEs were evenly distributed across all models, there would be 133 CDEs per model.
  - However one finds that BRIDG v2.1 has 1,669 CDEs, caAERS v1.1 (Adverse Events Reporting System) has 463, caBio 4.3 has 465, caTissue Suite 2.0 has 987, caNanoLab has 524 and C3PR v2.0 (Patient Registry) has 180.
- See Appendix C - Models Registered in caDSR for complete list of models with CDEs in caDSR.

Use of caDSR by NCI and Organizations

caDSR is being used by a variety of NCI divisions offices and centers, and the NCTN. The sections that follow contain brief information about a few of those organizations, and how those organizations are using caDSR to further their mission.

More detailed information and statistics for the NCI and organizations using caDSR can be found in Chapter 3 - NCI and other organization Usage of caDSR.

Cancer Therapy Evaluation Program (CTEP)

The NCI Cancer Therapy Evaluation Program (CTEP) coordinates the largest, publicly funded oncology clinical trials organization in the world. CTEP's staff of physicians, scientists, pharmacists, nurses, and regulatory specialists, assisted by government contractors, work diligently to assure the safe, efficient, and ethical conduct of this complex research enterprise.
Several years ago, CTEP mandated that variables (Data Elements) and Case Report Forms (CRFs) used in NCI-sponsored trials be registered in caDSR. Using a series of unified workflow processes, all new and amended trials go through a review process that verifies that the elements are added to the registry both as single elements and reused on forms to represent a full clinical trial. There are more than 9800 data elements arranged on more than 2500 clinical trial-centric forms in the registry.

The Eastern Cooperative Oncology Group-ACRIN (ECOG-ACRIN) is one of the groups coordinated by CTEP and is one of the largest clinical cancer research organizations in the United States. ECOG-ACRIN conducts clinical trials in all types of adult cancers. These clinical trials use the caDSR to describe CRFs by reusing CTEP data elements for sharing detailed descriptions of the data to be collected in clinical trials in both a programmatically and human-understandable way. ECOG-ACRIN uses the caDSR APIs to retrieve computer descriptions of CRFs to customize local clinical data management systems, ensuring that data is collected in precisely the same way across studies.

**Usage Summary**

CTEP statistics include the following:

- CDEs: Over 11,708 (both released and draft status) (as of Mar 2015)
- CRFs: 3,618 (as of Mar 2015)
- Active Trials: Over 900
- Study Participants Enrolled Annually: 30,000
- Grants and Cooperative Agreements: 400
- Investigational New Drugs (INDs): 100

ECOG statistics include the following:

- Active Clinical Trials: Over 90
- Annual Patient Accrual: 6,000
- Patients in Follow-Up: Over 20,000

**Cancer Centralized Clinical Database (C3D)**

C3D is a clinical trials management system, for which CRFs are used to collect data. The CDEs used on those forms are described in caDSR and represented as computable metadata that is shared electronically to support the clinical center information systems required to collect data consistently and accurately across trials. This increases data accuracy, reduces clinician training, and increases the possibility of aggregating data across studies.

In conjunction with the Center for Cancer Research (CCR), C3D provides Clinical Data Management System (CDMS) support for over 20 cancer centers, for which trial CRFs are described by the same CDEs. This reuse increases the potential for data aggregation across independent groups.

**Usage Summary**

The statistics listed that follow are from August 2010, and represent usage of caDSR content in the C3D application used by CCR to support these cancer centers.

- Sites: 200
- Users: 1,124
- Studies: 326
- Patients: 13,514
- CRFs: 910,071
- CDEs used in C3D studies: 3,491
- Individual data points (Responses): 150,341,554
- Lab CRFs from caDSR download CDEs: 604,128
- Data points from batch loaded CRFs: 124,480,823

**Division of Cancer Prevention (DCP)**

The DCP is the primary unit of the NCI devoted to cancer prevention research. DCP provides funding and administrative support to clinical and laboratory researchers, multidisciplinary teams, and collaborative, research-based networks. DCP also requires that all trials sponsored by the organization use variables that are described by data elements in caDSR.

Over the past decade, the DCP has been instrumental in developing protocol forms and form templates within caDSR. DCP has also authored a significant number of the data standards vetted and approved by the community over the past few years. These include standards for the capture and reporting of Address components, Person name, Person height, and Person weight. These data standards are now widely reused across the NCI clinical research enterprise.

**Usage Summary (as of Mar 2015)**

- CDEs: 3,900+ - organized into categories such as Adverse Events Reporting, Protocol Deviation Notification, and types of conditions such as:
  - 193 CDEs for Acocolbifene - High Risk for Breast Cancer
  - 181 CDEs for Budesonide Lung Nodules
  - 153 CDEs for Bladder Cancer
  - 155 CDEs for SAMe - Hepatits C cirrhosis
- CRFs and Form Templates: 133

**Specialized Programs for Research Excellence (SPOREs)**
Specialized Programs of Research Excellence (SPOREs) are funded through specialized center grants (P50s) that promote interdisciplinary research and move basic research findings from the laboratory to clinical settings, involving both cancer patients and populations at risk of cancer. The outcome of interdisciplinary research is a bidirectional approach to translational research, moving laboratory discoveries to clinical settings or clinical observations to the laboratory environment.

Since 2003, SPOREs studies predominantly for head, neck, and lung cancer have been added to the caDSR. There are studies and CRFs for two Iloprost trials in the registry, as well as studies for the University of Colorado (GO studies for advanced lung cancer).

Of special note is the fact that the University of Colorado now has refined its protocol templates to the point that builds of new studies require essentially no creation of new elements. Use of existing content not only allows for the rapid building of forms, but also enables and increases the ability to aggregate the collected data.

Usage Summary

As of October 2010:

- CDE: 704
  - Including 91 CDEs designated by the Lung SPORE at Emory University Winship Cancer Institute to collect data for lung cancer patients.
- CRFs: 29

National Marrow Donor Program (NMDP)

The NMDP is a member of the Center for International Blood and Marrow Transplant Research (CIBMTR). NMDP curators, trained by preceptor NCI CBIIT staff, are using the caDSR Form Builder, the cgMDR Microsoft Excel Add-in and Query Service Manager (QSM) for semantic annotation, along with manual curation tools, to register all of the necessary data collection forms in caDSR, and make them available via the caDSR web-based tools and programming interfaces.

This organization began with a requirement to add nearly 100 CRFs to the registry. At the present time, nearly all of the high volume forms have been completed, with CRFs designed for more rarely seen malignancies slated for future curation.

To harmonize the data across these forms and render the CDEs programmatically accessible via caDSR APIs, NMDP has chosen caDSR to record their reusable data elements, allowing them to develop sharable forms through Form Builder.

NMDP utilizes AGNIS, a network-based electronic forms system based on FormsNet to collect donor information. By recording their reusable data elements in caDSR, the NMDP team can then use Form Builder to build collections of CDEs to match their existing varied and unharmonized FormsNet forms. The use of caDSR and careful planning by the NMDP curation team is allowing NMDP to harmonize fields across all donor program data collection forms and efforts (caBIG® 2010 Annual Meeting Poster 82).

Usage Summary

- CDEs: 1,535 curated or identified for reuse.
  - An anticipated total of 14,000 CDEs are planned for use in their Hematopoietic Stem Cell Transplant (HSCT) program.
- CRFs: Nearly 100 with more specifically designed forms slated for the future.

caDSR Use by Other NIH Organizations

The caDSR is also being used by organizations that reside within NIH but outside of the NCI. These organizations provide excellent examples of the ability to reuse and re-purpose caDSR content for applications outside of cancer research. A few of these institutions are listed in this section, along with brief descriptions of their use of caDSR.

For a more complete list along with more detailed information on the programs and their use of caDSR, see Chapter 4 - NIH Usage of caDSR.

National Institute of Dental and Craniofacial Research (NIDCR)

The NIDCR was one of the first non-cancer adopters of the caDSR as an attractive place for the registration of data elements. Since early 2005, a number of curation teams across the US have been trained by NCI CBIIT staff in the best-practices for data element creation and maintenance.

To date, NIDCR has registered more than 1400 data elements in caDSR, notably with the addition of valuable and unique dental-specific content. This organization is one of the groups driving the development of batch curation and registration functionality for the caDSR. This functionality allows curators to add or edit large groups of CDEs at once. The NIDCR curators created a set of content that has since been loaded using the caDSR Bulk Loader. We continue to use organizations like NIDCR to capture the need for enhancements to caDSR and emerging business requirements from the research community.

Usage Summary

As of November 2011:

- CDE: 2,234 organized into a Practice Based Research network collection of UML Models (PEARL), as well as by area of research i.e. Orthodontics, Oral Surgery, Periodontics, etc.
- CRFs: 95

National Heart Lung and Blood Institute (NHLBI)
The NHLBI began partnering with the NCI in 2005 to add data elements into caDSR that reflect the variables used for a Family Blood Pressure Program (FBPP) study. Working with NCI CBIIT curators, NHLBI achieved reuse of some existing caDSR elements, along with registration of a significant amount of new cardiology-specific content over the ensuing year. The NHLBI context is also the context into which the NMDP elements are being added.

There are currently more than 4000 NHLBI data elements registered in caDSR. A number of these CDEs are being used by oncology groups, demonstrating the ability to reuse variables across the research community, regardless of the domain of expertise.

### Usage Summary

As of November 2011:

- **CDE:** 4,118 organized into Potential Relapse, Family Blood Pressure Program, Labs and Therapy and National Marrow Donor Program which has further organized the CDEs into 92 Protocols.
- **CRFs:** 110

### National Institute of Neurological Diseases and Stroke (NINDS)

NINDS has developed a set of data elements for use in clinical trials over the past several years. In the past year, members of KAI Research, a NINDS Clinical Research Organization, recommended to NINDS that their core data elements may be candidates for addition to caDSR. Based on this recommendation, a comparison analysis between NINDS content and NCI content was conducted. The result concluded that more than 80% of the core research variables were common across the two groups. This again demonstrates the potential reuse of content regardless of the area of expertise.

A small team of NINDS curators has been trained by NCI CBIIT, and were slated to begin curation of an identified core set of NINDS variables in August of 2010. After the core set is loaded, next steps will be determined.

### Usage Summary

As of November 2011:

- **CDE:** 310 organized into General CDEs, Parkinsons CDEs and Stoke CDEs.
- **CRFs:** 1

NINDS will reuse 123 CDEs (out of a total of 310 CDEs) found on 21 NINDS forms that they use for the collection of neurological clinical research data. NINDS has chosen to use classifications to organize their CDEs instead of CRFs.

### External Adoption of caDSR

caDSR provides open access to all of its content via publicly accessible browsers. CBIIT Semantic Infrastructure representatives have spoken at a number of external conferences about the caDSR approach to data management and services as a means of sharing data. This exposure has drawn the attention of a number of groups outside of NCI as well as international parties looking to find data descriptions for clinical trial and research datasets. Some of these groups and their adoption of caDSR are outlined briefly in the sections that follow.

For a more complete list along with more detailed information on the external organizations and their use of caDSR, see Chapter 5 - External Adoption of caDSR.

### Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences

The Winthrop P. Rockefeller Cancer Institute has decided to implement standard operating procedures requiring that all investigator-led trials use CDEs from caDSR to conduct their research.

The first project created 12 CRFs in the program’s Clinical Data Management System (CDMS), using 72 CDEs, 38 of which were already in caDSR (50% reuse). They created the remaining 33 CDEs using the caDSR CDE Curation Tool. They have installed openMDR and plan to use this for ongoing collaborative research studies (caBIG® 2010 annual meeting Poster 80). OpenMDR is an OSU-caDSR ISO/IEC 11179 compatible registry that can pull CDEs from caDSR for use in a local repository.

The institute is also using caDSR CDEs and LimeSurvey, an open source tool, to recreate questionnaires for breast cancer mammography.

The center recreates or reuses the survey CDEs in caDSR, then imports the CDEs into LimeSurvey. This ensures that the data are collected with high-quality and common standard data descriptions across studies. As a result of doing this work, the existing survey data was analyzed, corrected and standardized and can be reused for other studies (caBIG® 2010 Annual Meeting poster 81).

### Union of Light Ion Centers in Europe (ULICE)

Specifically this section discusses the the WP7 Common Database and Grid Infrastructures project, for improving and catalysing access to research information for the broad European community.

ULICE is a 4-year project set up by 20 leading European research organizations, including 2 leading European industrial partners (Siemens and IBA), to respond to the need for greater access to hadron-therapy facilities for particle therapy research. Project coordinator is the Italian Research Infrastructure Facility CNAO (Milan).

WP7 is concerned in part with the compilation of a sufficient body of evidence to support future evidence-based evaluations of CPT – identifying new, preferred treatment protocols for specific cancers and ensuring that patients receive the best treatment.
By establishing a registry using key shared terminology and caDSR-type data elements to collect data, WP7 will collect information on every patient treated. However, such a registry will deliver the evidence only if the data being gathered is comparable or interoperable. It is necessary to be able to reliably combine data recorded in different contexts, in different clinics, and in different health care systems.

Given the variety of (changing) practices and situations, it is unrealistic to expect that a single quality standard might be imposed for data collection (Parkin and Bray 2009).

To mitigate this problem, the ULICE/WP7 registry has been set up to allow all data collected in the registry to be properly described, with detailed explanations of protocol for each clinical observation, and detailed definitions for each of the possible values recorded. These detailed explanations of the data will also be collected using caDSR-style common data elements.

The ULICE project is in Year 1 of a 48 month project. It is aligning data elements, wherever possible, with caDSR. Currently 50% of the data elements to describe ULICE studies are caDSR-based CDEs, as are 25% of the Baseline CDEs and 70% of the Followup CDEs.

For more information about ULICE, see the ULICE web site [1].

**Center for International Blood and Marrow Transplant Research (CIBMTR)**

The CIBMTR is collecting data on 45 forms containing over 5,000 data points. The program uses clustering techniques to analyze the component parts of CDEs to discover similar data points. These will be used to help create an HSCT-specific (Human Stem Cell Transplant) database model in the future.

**Usage Summary**

As of November 2011:

- CDE: N/A
- CRFs: N/A

**Information Management Services, Inc. (IMS) BIS, NAACCR and DCCPS/caSEER program**

The IMS and its associated programs provide NIH, as well as pharmaceutical, academic, and other research organizations with biomedical computing support.

IMS supports the development of CDEs for caDSR for the Person Age standard and the North American Association of Central Cancer Registries (NAACCR) CDEs. The program developed data models for seven data services for DCCPS/caSEER Program for sharing through caGrid and has registered the CDEs in caDSR. The program’s databases contain over 6 million records that can be searched and summarized to produce tables, graphs, and geographic maps for caSEER. Applications developed by IMS are designed to extract and publish specimen data using CDEs and a Common Business Model.

**Usage Summary**

As of November 2011:

- CDE: 385 (NAACCR version 11.3) organized into categories such as Demographic, Hospital Specific, Text Diagnosis, Treatment 1st Course, etc
- CRFs: 3 including the NIH 2005 Questionaire (a caGrid service) and the Fagerstrom Test for Nicotine Dependence and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial

**A Growable Network Information System (AGNIS)**

AGNIS began as an idea to transmit hematopoietic stem cell transplant data between organizations, but developed into a standards-based communications model.

AGNIS works by acting as a translator between two centers. When organizations attempt to communicate their data to one another, they must first massage the data into a format that the other organization can recognize. The translation mechanism actually uses standardized CDEs that are curated and stored in caDSR. Due to the reusable nature of data elements in caDSR, everyone benefits as more terms are created.

The AGNIS network uses the CDE Browser, and Curation Tool, as well as the caDSR APIs to create and access standard CDEs.

**Usage Summary**

As of November 2011:

- CDE: N/A
- CRFs: N/A

**cancerGrid and cgMDR (United Kingdom)**

CancerGrid is an initiative involving scientists at the Universities of Oxford and Cambridge, working together to reduce the cost of clinical research, and to increase its value through effective data sharing.

Along with a meta-model (or model template) for clinical studies, and work on the classification and registration of clinical trials, the project has produced an ISO/11179-compliant metadata registry, a semantically-aware trials design and management system, and a toolkit for clinical data transformation and integration.

The metadata registry software has been adopted for use within the US initiative, and is used by a number of organizations within the UK. Development continues at Oxford, with clinical collaboration in Cambridge and London, and technical collaboration with the CBIIT SI team in the US.
The cancerGrid Metadata Registry (cgMDR) was built to be compatible with caDSR, including extensions to ISO 11179 reference concepts as the grounding for semantics, and a new version of this registry has been developed to include CRFs, much like the caDSR CRFs, that are built from and based on CDEs. The Oxford computing lab has developed an import that can take CDEs downloaded from caDSR into an XML file, and then import them into local cgMDR installations. By default, the downloaded cgMDR repository already includes the "standard" caDSR CDEs as well as BRIDG CDEs, so that organizations who want to utilize local installations of a data standard registry can do so, allowing them to more easily reuse and harmonize their data with NCI standards.

Usage Summary

As of November 2011:

- CDE: N/A
- CRFs: N/A

2 - caDSR Infrastructure and Toolset Usage

caDSR provides the foundational layer for representing the precise meaning and representation of data, and through linkage to EVS terminology and ISO/IEC 11179 standard, the ability to translate the meaning throughout the informatics infrastructure and across various research domains. It is comprised of a central database and a set of web based tools and software interfaces (APIs).

This section includes the following topics.

- Overview
- NCI caDSR CDE Browser
  - CDE Browser Description and Usage
  - CDE Collections as Prepackaged File Downloads
- NCI caDSR Form Builder
- NCI caDSR Curation Tool
- NCI cgMDR Excel Addin, Template and Bulk Loader
- NCI caDSR UML Model Browser
- NCI caDSR Sentinel Tool
- NCI caDSR Database Server, Domain Model and Freestyle APIs
- NCI caDSR Semantic Integration Workbench (SIW) and UML Loader

Overview

Until 2010, all data, (30K+ "Released" data elements, 44k+ total), and information models (198 including multiple versions of some models) for all systems and applications in caBIG® were required to have the meaning coded with controlled terminology, and recorded in the caDSR via UML Models. The high number of data elements in caDSR represent the many and varied ways in which similar data are collected and stored in databases throughout the community and provides a means by which data transformation programs can be written to enable data to be combined, or to design new applications that can reuse existing data.

By exposing these details in a common, structured way using ISO/IEC 11179, caDSR tools enable the construction of data collection instruments and software heretofore not possible or feasible because the details were hidden away in application programs and paper documentation. In addition to the recording details of the data elements used in applications, the registry also records the meaning in reusable semantic components that provide the basis for detecting conceptual equivalence across data elements and thus improves the potential for data aggregation.

Applications, such as caIntegrator and caB2B, access caDSR via API at runtime to display descriptive information about data on data collection forms or to provide drop down lists for populating fields. Others import CDEs to use to customize data collection forms in local software applications such as a Clinical Data Management System (CDMS), C3D, LimeSurvey and others.

Adoption/adaption includes development of an open source XML-based version of caDSR by the University of Oxford Computing Lab, cgMDR, to support the cancerGrid's clinical trails and other clinical and population studies in the UK, caDSR CDEs were downloaded, decomposed and imported into this registry to form the basis of new content development, reusing many of the NCI's CDEs in UK trials. This XML-based registry was taken by Ohio State University and enhanced to work with caGrid, and provides the basis for annotating services with caDSR CDEs during UML Model design, generating caGrid compatible services for deployment on the grid. OSU has also used this registry, named openMDR, to support its Clinical and Translational Science Award (CTSA) program and another consortium involved in collecting human studies, HSDB.

caDSR usage can be measured by the large number of communities and applications that have registered their data elements and models in caDSR. It enables sharing of these descriptions among data managers and software developers through CDE Browser downloads and the use of caDSR APIs, including an HTTP API that can be used to browse and download content in HTML or XML.

The caDSR software was designed by Oracle and the Census Bureau in the late 1990’s as a centralized repository for ISO/IEC 11179 content, and due to its dependency on an Oracle database and an application server for the Admin Tool and database, is not well suited to installing small cancer centers who are resource-constrained or lack the Oracle skills to get the suite of products installed, customized and working. Instead, the NCI hosts content for those who want to use the infrastructure. Others who want their own registry now also have the cgMDR/openMDR option.

NCI caDSR CDE Browser

CDE Browser Description and Usage

The CDE Browser is the starting point for exploring the details of the data elements described and registered in the caDSR. This tool supports browsing, searching, and exporting CDEs in XML or Excel format, within or across end user contexts as follows.
• Search for data elements by NCI Thesaurus Concept, Value Domain Permissible Value, Classifications or simple text searches in the names and definitions of the elements in caDSR.
• Search for classes of information in caDSR and find reusable CDEs. For example, advanced search panel can be used to search for the Property "Email Address" and retrieve the 43 that describe different types of Email addresses such as "Organizational Unit Email Address Text", "Person Email Address Text", "Clinical Trial Participant Email Address Text" and "Investigator Email Address Text". The reuse of "Email Address" is recorded in the caDSR so that developers can detect the similarity between data elements and enable them to discover and potentially combine data from different sources.
• Using the Shopping Cart, create a customized set of CDEs for exporting in XML or Excel format or sharing with other applications such as Form Builder.
• The CDE Browser publishes its DTD to export/download data elements in XML.
• Access the Tool at [https://cdebrowser.nci.nih.gov/CDEBrowser/](https://cdebrowser.nci.nih.gov/CDEBrowser/)

**USAGE:** An average of 47 unique users visit the CDE Browser each business day, with over 18,204 unique visitors since we started tracking statistics in 2005. The number of pages served during this time is 1,510,814. Approximately 11 pages were viewed per 133,149 visits. CTEP currently has over 10,000 released and draft CDEs. See Statistical Appendix for more details.

The NCI Cancer Clinical Research (CCR) center has 3,491 selected CDEs that are downloaded and imported into the Cancer Centralized Clinical Database (C3D), with 935,573 instances of CRFs built from these common descriptors and over 150M data points (responses) collected.

The following chart shows the number of unique visitors, per quarter, to the CDE Browser from the 1st quarter of 2009 through the 3rd quarter of 2011.

The average number of visits began to rise in 2009, potentially due to the caBIG® program reaching maturity and that all data services had to have a UML model registered in caDSR, the CDE Browser being the primary means by which users could review and download the elements in their models after loading. The level of unique visits has stabilized so far in 2011 to a level equivalent to early 2009.

### CDE Collections as Prepackaged File Downloads

Several collections of CDEs have been downloaded in Excel and XML format from the CDE Browser and pre-packaged for downloading from NCI’s Wiki. Pre-packaged downloads include all the CDEs that are in a "released" workflow status, those that are caBIG Standards and CDEs in TCGA, NACCAR, SEER and BRIDG. Collections are posted to the caDSR Wiki on the "caDSR Hosted Data Standards, Downloads, and Transformation Utilities" page.

### NCI caDSR Form Builder

Form Builder allows users to share data element descriptions across multiple forms to ensure that data will be comparable. It helps users organize CDEs that replicate the content of Case Report Forms. These forms can include modules grouping CDEs together that can be copied from one form to another. Forms can include complex behavior such as skip patterns where the answer to a question determines the next question to be asked. Using the CDE Browser, you can search for CDEs, place them in a shopping cart, and from there insert CDEs as Questions on a Form. As you place the CDEs on the Form, the tool uses the stored metadata to provide default question text and value domain information automatically. If the value domain information is presented as an enumerated list of values, you can perform basic functions to organize the list.

Key capabilities of Form Builder include:

• Define skip patterns between questions based on question responses
• Define repeating groups
• Define default values for questions in repeating groups
• Publish a Form in the caBIG® Context's Form Catalog
• Subscribe to Sentinel Reports that are triggered by changes to CDEs on the Form
• Classify the Form in one or more caDSR Classification Schemes
• Classify the CDEs on a form in one or more caDSR Classification Schemes
• Download the Form to Excel
• View and print from a Printer Friendly version of the Form
• Edit, Save or Download the CDE Shopping Cart
• Edit, Save a Forms shopping cart for storing collections of form to export to other systems through the Object Cart API

Access the Tool at https://formbuilder.nci.nih.gov

**USAGE:** As of October 2011, an average day there are 6 curators using Form Builder creating or modifying forms. There are 3,288 forms in caDSR, 466 Protocols (a 14% increase of 2010), and over 2,900 forms in the CTEP context, caDSR's largest single user (a 12% increase over 2010). CTEP Case Report Forms (CRFs) are grouped to describe minimum datasets to be collected for various types of cancer organized by disease and type of trial. NCI CTEP uses CDEs and CRFs for reporting trial results for CTEP Sponsored trials, of which there are currently 611 trials.

Commercial vendors Medidata (RAVE), Westat (for CTEP) and Eastern Oncology Centers Group (ECOG) retrieve CRFs using the caDSR API and import them to customize their data collection systems. Refer to the section titled NCI caDSR Database Server/Domain Model and Freestyle APIs for API usage.

ACRIN is involved in registering standard data elements for imaging and using them to create forms in Form Builder.

The following chart shows the number of unique visitors, per month, to Form Builder from the 1st quarter of 2009 to the 3rd quarter of 2011.

---

**NCI caDSR Curation Tool**

The CDE Curation Tool supports creation and editing of all the primary semantic descriptions for Data Elements used by the community in data repositories or application software. It is intended for use by Context Administrators. This tool's features facilitate the use of the CBIIT Enterprise Vocabulary Services (EVS) to create administered item names and definitions, helping to ensure ISO/IEC 11179 compliance and also use of caDSR naming conventions.

Public users can browse for CDEs, reusable ISO/IEC 11179 value domains (VD), data element concepts (DEC), Object Classes and Properties using identifiers, search strings, Classification Schemes or EVS concepts. The tool leverages the ISO 11179 metamodel to "get associated" items.

Access the Tool at https://cdecurate.nci.nih.gov

**USAGE:** Since 2005, on an average day 22 unique users use the Curation Tool, with 4,558 unique visits to the site, and over 1.6M pages served. New CDEs can be created using this tool, the Admin Tool or the UML Model Loader. With the detail of caBIG® models no longer required to be registered in caDSR, that combined with curator harmonization activities designed to ensure and increase reuse of existing CDEs, the number of new CDEs year-over-year is trending down.

The following chart shows the number of unique visitors, per month, to the CDE Curation Tool from January 2009 to September of 2011.
NCI cgMDR Excel Addin, Template and Bulk Loader

The term “cgMDR” actually refers to a group of components designed to work together to help users get large lists of administered components bulk loaded into caDSR. The initial audience for these products was the National Marrow Donor Program (NMDP) but interest in its use has spread beyond that group.

cgMDR actually stands for the “CancerGrid Metadata Registry” and specifically refers to a downloadable localized database based on ISO/IEC 11179 where you can store and administer a personalized set of data elements and their components. This database and its interface were created by the CancerGrid team at Oxford University Computing Lab in the United Kingdom. The NCI GForge project archive is on this wiki.

The CBIIT cgMDR installation also includes a group of components, including two add-ins and two Excel 2007 spreadsheets that work in conjunction with cgMDR as well as with other data repositories. These additional features help provide a complete solution to assist you in creating a list of personalized data elements that can then be bulk loaded into caDSR without the need for creating a UML Model or by individual manual curation of each element.

**USAGE:** National Marrow Donor Program (NMDP), University of Michigan, ACRIN for batch loading CDEs and Value Domains into caDSR.

NCI caDSR UML Model Browser

The UML Model Browser supports web browsing and searching data described by UML Models transformed and loaded in the caDSR repository via the UML Loader. This allows users to find administered items that are part of registered UML models for data services on caGrid. The UML Model Browser supports browsing, searching and exporting the classes, attributes and relationships between classes of a UML domain model. Within the framework of a UML Model CDEs are mapped to the UML attribute level. Search results display the Package Names, Attributes and Java primitive types. The CDEs used for semantic resolution are presented as links to the CDE Browser. CDEs in UML Models can be downloaded from the UML Model Browser.

Access the Tool at [http://umlmodelbrowser.nci.nih.gov](http://umlmodelbrowser.nci.nih.gov)

**USAGE:** Since tracking began in March 2006, on an average day six visitors come to the site with 2,765 unique visitors through October 2010. 34,522 web pages were viewed in this four year period. There are 198 models loaded into caDSR describing the data for each of these applications using standard ISO /IEC 11179 descriptions. The descriptions of this data in caDSR ensure that the meaning of the data is unambiguously represented for both human and computer interpretation. The UML Model API is used by caGrid and caB2B to explore registered models programmatically. The caGrid Portal exposes the semantic metadata in its portal, which was accessed by 5,163 unique visitors between October 2009 and October 2010. According to ISO 8000, a Data Quality standard, registration of the details of these data and application models ensures the model owner both owns the data and meets guidelines for Data Quality.

The chart shows the number of unique visitors, per quarter, to the UML Model Browser from the 2nd quarter of 2006 to September of 2011.
UML Model Browser usage peaked with loading of models by the caBIG® community during 2008 and 2009, and has leveled-off slightly lower than when the tool was first introduced.

The tool is useful for viewing the elements in a specific model, and we anticipate that with more emphasis on reusing content from existing models, the average number of visitors using the model browser will remain the same or increase slightly.

NCI caDSR Sentinel Tool

The caDSR Sentinel Tool was first introduced in 2005 to allow users to create and manage Alert Definitions for the caDSR. 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 by email, with a hyperlink to a report that specifies the changes that have taken place. A script that kicks off this tool runs nightly, but the reports can also be run through the user interface.

The caDSR Sentinel Tool provides the capability to:

- Monitor all changes to Administered Items including Data Elements, Data Element Concepts and Value Domains and Case Report Forms
- Filter report content by Context, Specific forms or templates, Classification Scheme, Class Scheme Item, Creator and Modifier
- Trigger report generation using Workflow Status, Registration Status and Version
- Set reports to automatically be generated daily, weekly, monthly or on demand
- Create a report distribution list which may optionally include a process URL to send the report in XML format to software for evaluation

Access the Tool at https://cadsrsentinel.nci.nih.gov

**USAGE:** On an average day, 3 users visit the caDSR Sentinel tool site with over 264 unique visitors between October 2009 and October 2010, and 8,354 page views in that time.

NCI caDSR Database Server, Domain Model and Freestyle APIs

All caDSR content is available through various application programming and web service interfaces.

**caDSR API** allows users to access caDSR content by using a web browser to navigate the caDSR domain model and returns results in HTML or XML. A caDSR Java API provides a set of methods that can be used to retrieve content as XML documents. According to NCI statistics from Wusage, from October 2009 to October 2010, 628 unique visitors came to the site and accessed over 25M documents. The number of pages is slightly inflated due to the design of the caDSR Domain Model, as multiple pages between 6-20 pages must be served in order to get one logical document. Our estimate is that over 1.2M logical documents were retrieved.

Access the HTTP caDSR Domain Class Browser at http://cadsrapi.nci.nih.gov/cadsrapi40/

**Freestyle API** is another caDSR interface that provides access to content both via a web browser and application programming interface. From October 2009 to October 2010, 328 unique visitors came to the Freestyle API site and accessed over 9.6k documents. The Freestyle API uses the caDSR Domain API to simplify object retrieval.

Access the Tool at http://freestyle.nci.nih.gov

The chart that follows shows the comparative number of caDSR pages delivered through the caDSR API for various organizations during the period from 1st quarter 2009 to the 3rd quarter 2011.
The next chart shows the number of unique visitors to the Freestyle API during the period from 4th quarter 2006 through the 4th quarter 2010.

**usage:** The caDSR API usage information in the statistics appendix lists the top URLs that access the caDSR API. AGNIS uses the API to access Forms.

**nci cadsr semantic integration workbench (siw) and uml loader**

The Semantic Integration Workbench (SIW) a tool that assists users in adding consistent metadata to a UML model represented as an XML file, or verifying consistency with existing caDSR content by tagging a domain model with matching concepts from the NCI Thesaurus, or attaching existing caDSR CDEs or Value Domains to attributes in the model. These annotations ensure reuse of CDEs or other metadata elements that have been previously recorded in caDSR. The UML Loader transforms the file and decomposes it into ISO/IEC 11179 descriptive metadata.

**usage:** As of July 2011, there are 7 new models loaded in 2011, 2 in the queue to be loaded and 2 on hold. 198 caBIG® and NCI UML models representing all versions of these applications' data have been transformed into ISO/IEC 11179 descriptive metadata in caDSR.

These models use 26,360 CDEs, with many of the CDEs reused across models. For example, if the CDEs were evenly distributed across all models, each model would use only 133 CDEs. However when searching for CDEs in each model, one finds that BRIDG v2.1 has 1,669 CDEs, caAERS v1.1 (Adverse Events Reporting System) has 463, caBIO 4.3 has 465, caTissue Suite 2.0 has 997, caNanoLab has 524 and C3PR v2.0 (Patient Registry) has 180. This includes several versions of BRIDG, caAERS and C3PR.

See Appendix C - Models Registered in caDSR for a complete list of models with CDEs in caDSR.
The chart that follows shows the comparative number of models processed through the SIW and loaded to caDSR via the UML Model Loader, by year, from 2005 through 2010.

4 - NIH Usage of caDSR

caDSR has been adopted for use by several NIH organizations that exist outside of the NCI. This use of caDSR demonstrates the capability of the repository and its associated tools to apply to data gathering, harmonization, and reuse by fields outside of cancer research. The projects outlined here include:

- National Heart Lung and Blood Institute (NHLBI)
- National Institute of Dental and Craniofacial Research (NIDCR)
- National Institute of Child Health and Development (NICHD)
- National Institute of Neurological Diseases and Stroke (NINDS)

National Heart Lung and Blood Institute (NHLBI)

Project Sponsor: Dr. Jenny Larson, Dr. Douglas Rizzo
Project Manager: —

The National Heart, Lung and Blood Institute (NHLBI) began partnering with the NCI to add data elements to the caDSR in 2005 to reflect the variables used for a Family Blood Pressure Program (FBPP) study.

Working with NCI CBIIT curators, NHLBI achieved reuse of some existing elements, with registration of a significant amount of cardiology-specific content over the ensuing year. This is also the context into which the NMDP elements are being added.

**USAGE:** There are currently more than 4000 data elements registered in caDSR; a number of these are being used by oncology groups, demonstrating the ability to reuse variables across the research community regardless of the domain of expertise.

National Institute of Dental and Craniofacial Research (NIDCR)

Project Manager: Alice Birnbaum, Director of Biostatistics, Axio Research, LLC

The National Institute of Dental and Craniofacial Research (NIDCR) was one of the first non-cancer adopters of the caDSR as an attractive place for the registration of data elements. Since early 2005, a number of curation teams across the US have been trained by NCI CBIIT staff in the best practices for data element creation and maintenance.

**USAGE:** To date, more than 1400 data elements have been registered in the caDSR, notably with the addition of valuable and unique dental-specific content. This organization is one of the groups driving the development of batch curation and registration functionality for the caDSR; the curators created a set of content that has been loaded using the caDSR Bulk Loader. We continue to capture the need for enhancements and emerging business requirements.
National Institute of Child Health and Development (NICHD)

Project Sponsor: Dr. Steven Hirschfeld
Project Manager: —

The National Institute of Child Health and Development is one of the newest non-oncology users of the caDSR. Beginning in November, 2009, a model to represent a Newborn Assessment was annotated and loaded into the caDSR. The resulting 107 data elements were reviewed and released for widespread community review.

USAGE: Currently a small team is being trained by NCI CBIIT staff to act as curators for new content that is being planned for addition to the caDSR. As of 2011, 114 CDEs have been created for use in this research domain, 7 have been reused from other Contexts.

National Institute of Neurological Diseases and Stroke (NINDS)

Project Sponsor: Joanne Odenkirchen, NINDS
Project Manager: Staci Grinnon, Yun Lu, KAI Research

The National Institute of Neurological Diseases and Stroke (NINDS) has developed a set of data elements for use in clinical trials over the past several years. In the past year members of an NINDS CRO, KAI Research, recommended to NINDS that their core data elements may be candidates for addition to the caDSR. Based on their evaluation of the caDSR, a comparison analysis between NINDS content and NCI content was conducted. The result was in excess of 80% of core research variables across the two groups, again demonstrating the potential reuse of content regardless of the area of expertise. A small team of curators has been trained by NCI CBIIT, and curation of an identified core set of NINDS variables was slated to begin in August, 2010. After the core set is loaded, next steps will be determined.

USAGE: NINDS has reused and created 310+ CDEs found on 21 NINDS CRFs (123 reused CDEs) used to collect neurological clinical research data.

5 - External Uses of caDSR

NCI caDSR provides open access to all its content via publicly accessible browsers, and representatives have spoken at a number of external conferences on the caDSR approach to data management and services as a means for sharing data. This publicity has drawn the attention of a number of external groups and international parties looking to find data descriptions for clinical trial and research datasets. As such, these organizations have adopted caDSR as a means to accomplish this important goal.

Several of the groups who have adopted caDSR are highlighted in the next sections, and include the following:

- Biopathology Center, National Childrens' Hospital
- PhenX Toolkit
- GIATE
- American Heart Association (AHA) and American College of Cardiology Foundation (ACCF)
- Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences
- Union of Light Ion Centers in Europe (ULICE)
- Winthrop P. Rockefeller Cancer Institute, Cancer Control Core at the University of Arkansas for Medical Sciences; Breast Mammography
- Oncology Patient Enrollment Network (OPEN) - Sponsor-GTEP/NCI
- Duke Translational Medicine Institute
- Tiwan Cancer Registry
- Children's Oncology Group (COG) and National Childhood Cancers Foundation
- University of Miami Ontology Modeling
- Center for International Blood and Marrow Transplant Research (CIBMTR)
- The Human Studies Database (HSDB)
- Towards Unambiguous formal descriptions of cancer therapy experiments
- Ontology based queries for caGrid Infrastructure
- Information Management Services, Inc. (IMS) BIS, NAACCR and DCCPS/caSEER program
- The CTSA Health Ontology Mappe
- A Growable Network Information System (AGNIS)
- cancerGrid, cgMDR UK
- CDC, PHIN Reportable Conditions and Common Core Data Elements

Biopathology Center, National Childrens' Hospital

Project Sponsor: Dave Billiter

Cooperative Group Specimen Banks are using caDSR content via caDSR API in their Group Banking Committee tooling to inspect and harmonize across different specimen banks. They use the caDSR API along with the UML Model Browser and CDE Browser to identify CDEs that need to be harmonized.

PhenX Toolkit

Website

PhenX is a three year project to prioritize phenotypic and exposure measures for Genome-wide Association Studies (GWAS) and other large-scale genomic research efforts. A phenotype is an organism's observable characteristics or traits: such as its morphology, development, biochemical or physiological properties, behavior, and products of behavior (such as a bird's nest). Phenotypes result from the expression of an organism's genes as well as the influence of environmental factors and the interactions between the two.
Leaders of the scientific community, including liaisons from the Institutes and Centers of the National Institutes of Health, are driving PhenX, assessing and prioritizing a broad range of domains relevant to genomics research and public health. Domains include common complex diseases and conditions; lifestyle factors and anthropometrics; and environmental exposures.

The PhenX Toolkit provides standard measures related to complex diseases, phenotypic traits, and environmental exposures. Use of PhenX measures allow for the combining of data from a variety of studies, making it easier for investigators to expand a study design beyond the primary research focus. The data elements used to create these standard measures are based on 660 CDEs from caDSR.

GIATE

Project Sponsor: The Antibody Society

Guidelines and Information About Therapy Experiments (GIATE) is a data standard for recording data about therapy experiments. It is being promoted by the Antibody Society, as a data standard for antibody therapy experiments but it has become apparent that the scope of materials used in antibodies as therapeutic agents, means that the data standard can equally be used to describe a wide range of therapeutics.

GIATE categorizes the information about therapy experiments into three classes, that is, information about the target, therapy agent and models used to test the therapy. Nested within these classes are additional classes describing different aspects and properties of the parent class. Therefore, a logical representation of GIATE would be a tree, with three main branches, which stems into smaller branches and leaves. The contents of the GIATE tree are downloaded from the caDSR, each concept contains a reference to the caDSR data element.

American Heart Association (AHA) and American College of Cardiology Foundation (ACCF)

Project Sponsor: AHA, Duke

The AHA and ACCF have chosen the caDSR define and disseminate clinical data standards - sets of standardized data elements and corresponding definitions - to collect data relevant to cardiovascular conditions.

The excerpt that follows is from ACCF/AHA 2011 Key Data Elements and Definitions of a Base Cardiovascular Vocabulary for Electronic Health Records.

Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences

Project Sponsor: Umit Topaloglu, PhD

The institute has decided to implement standard operating procedures (SOPs) requiring that all investigator led trials use CDEs from caDSR to conduct their research. The first project created 12 CRFs in the program’s Clinical Data Management System (CDMS), using 72 CDEs, 38 of which were already in caDSR (50% reused). They created the remaining 33 CDEs using the Curation Tool. They have installed openMDR (OSU caDSR ISO/IEC 11179 compatible registry that can pull CDEs from caDSR for a local repository). The program plans to use this for ongoing collaborative research studies (caBIG® 2010 annual meeting Poster 80).

Union of Light Ion Centers in Europe (ULICE)

Project Sponsor: EU, WP Lead Dr. Bleddyn Jones, Manchester University, Oxford Computing Laboratory, Oxford UK

Informatics PIs: Professor Jim Davies, DPhil. Steve Harris, PhD

Project: WP7 Common database and grid infrastructures for improving and catalysing access to RI for the broad European community.

WP7 is concerned in part with the compilation of a sufficient body of evidence to support future evidence-based evaluations of CPT – identifying new, preferred treatment protocols for specific cancers and ensuring that patients receive the best treatment.

ULICE is a 4-year project set up by 20 leading European research organizations, including 2 leading European industrial partners (Siemens and IBA), to respond to the need for greater access to hadron-therapy facilities for particle therapy research. Project coordinator is the Italian Research Infrastructure Facility CNAO (Milan).

By establishing a registry using key shared terminology and caDSR-type data elements to collect data, WP7 will collect information on every patient treated. However, such a registry will deliver the evidence only if the data gathered is comparable or interoperable. It is necessary to be able to reliably combine data recorded in different contexts, in different clinics, and in different health care systems.

Given the variety of (changing) practices and situations, it is unrealistic to expect that a single quality standard might be imposed for data collection (Bray and Parkin 2009) (Parkin and Bray 2009), so the registry has been set up to allow all data collected in the registry to be properly described, with detailed explanations of protocol for each clinical observation, and detailed definitions for each of the possible values recorded. These detailed explanations of the
data will also be collected using caDSR-style common data elements. The ULICE project is in Year 1 of a 48 month project; it is aligning data elements, wherever possible, with caDSR. Currently 50% of the data elements to describe ULICE studies are caDSR based CDEs, as are 25% of the Baseline CDEs and 70% of the Followup CDEs.

Winthrop P. Rockefeller Cancer Institute, Cancer Control Core at the University of Arkansas for Medical Sciences; Breast Mammography

Project Sponsor: Umit Topaloglu, PhD

**USAGE:** The cancer institute is using caDSR CDEs and LimeSurvey, an open source tool, to recreate questionnaires for breast cancer mammography. The center recreates or reuses the survey CDEs in caDSR, then imports the CDEs into LimeSurvey. This ensures that the data are collected with high quality and common, standard data descriptions across studies. As a result of doing this work, the existing survey data was analyzed, corrected and standardized and can be reused for other studies (caBIG® 2010 Annual Meeting poster 81).

Oncology Patient Enrolment Network (OPEN) - Sponsor-CTEP/NCI

Project Manager: Mike Montello (Project officer)
Bioinformatics: Ravi Rajaram

The OPEN System has adopted the caCORE Application Programming Interface (API) and the Form Builder application, in addition to integrating and adapting the Common Security Model (CSM) components, as part of its architecture. OPEN uses the caDSR API to download the case report form (CRF) metadata from the caDSR. In addition, OPEN has adapted the CSM in such a way that the authorization feature of the CSM is used while authenticating the user against the National Cancer Institute – Cancer Therapy Evaluation Program (NCI CTEP) Identity and Access Management (IAM) system. The CSM user provisioning tool (UPT) is being used to provide instance and attribute level security.

Duke Translational Medicine Institute

Bioinformatics Sponsors: Dr. Jeffrey Ferranti, Patricia Gunter
M. Nahm, PhD, Associate Director, DTMi Biomedical Informatics
H. Shang, Director of Business Information Services, Duke Health Technology Solutions
Rob Califf, M.D., Vice Chancellor of Clinical Research, Duke University Health Systems
Asif Ahmad, CIO, Duke Health Technology Solutions
Dwight Smith, Director, Information Technology Application Development, Duke Health Technology Solutions

Dr. Ferranti at DTMi recognized a business problem in the lack of standardized clinical terminology across the enterprise, preventing meaningful sharing and reuse of the data – semantic interoperability.

Following NCI's lead for using controlled terminology as the basis for metadata descriptions (caDSR), the DIMI solution is to build a metadata registry that would become a foundation for future data management and data quality efforts. Currently Duke is considering caDSR open source as a possible solution versus building anew.

Silos of operational databases and data entry applications exist in Patient registration, Patient billing, Clinical systems – surgical, emergency, labs, and Disease registries in over 150 clinics with over 3 million HL7 transactions per day. To promote consistent semantics, DTMi would leverage controlled terminologies – LOINC, SNOMED-CT, ICD-9, CPT, and others, and the ISO/IEC 11179 standard for metadata registries and HL7 messaging standards. Like many institutions, DTMi data was collected on forms. The current repository collected these forms as a blob of data. DTMi used its 11179 registry to create structured descriptions of the data on these forms and used the metadata descriptions as a means by which to extract and transform the data into a data warehouse. The major difference between the NCI's caDSR and Duke's metadata registry is the terminology basis. NCI has used the NCI Thesaurus as the basis for describing data; DTMi is using concepts from the HL7 RIM as the basis for the metadata descriptions.

Taiwan Cancer Registry

ISO/IEC 11179 has been used to describe the cancer data set for the Taiwanese cancer registry.

A paper entitled "Annotating Taiwan Cancer Registry to caDSR for International Interoperability" in "Future Wireless Networks and Information Systems", 2012 explains that,

"It is very difficult to exchange and integrate the data among different cancer institutions, in order to discover useful biomedical information for the cancer community. We developed a cancer Biomedical Informatics GridTM (caBIG®) Silver level compliant cancer registry database system, called the gridTCR (grid of the Taiwan Cancer Registry), which integrates the demographic data, clinical history, pathology data, and clinical outcome data including treatment, recurrence and vital status from cancer registry databases in Taiwan. In this manuscript, we will be developing the common data elements using vocabulary standards, ontology and semantic modeling methodology. The Taiwan Cancer Common Data Element Project (TCCDEP) developed 40 data elements to annotate the cancer registry data collected. The aim of this project is to create a core set of data elements for annotating the cancer registry data and achieve the interoperability over the caBIG community. We describe the process required to develop the model, the caDSR CDEs, and the results of the modeling effort. We address difficulties we encountered and modifications for solution. Currently, the Taiwan cancer registry CDEs are released and available in CDE browser for reusing. Furthermore, we will extend our CDEs to daily clinical practice and trials, along with how the methods were used to fully implemented in hospitals and cancer research centers in Taiwan."

Children's Oncology Group (COG) and National Childhood Childhood Cancers Foundation
John Deardurff, Nationwide Childrens Hospital

The COG and National Childhood Cancers Foundation are using caDSR APIs to retrieve CDEs for development of studies using common data descriptors.

University of Miami Ontology Modeling

Yevs R. Jean-Mary

Using UML models registered in caDSR and related NCIt concepts, the University of Miami uses semantic web technologies including OWL to develop executable queries against NCIt concepts that can return instances of caGrid Data, (caBIG® 2010 annual meeting poster 87).

Center for International Blood and Marrow Transplant Research (CIBMTR)

Kirt Schapner, Robinette Aley, Barb Kramer, Dr. Douglass Rizzo

This project groups similar data elements using ISO/IEC 11179 structures.

The Center for International Blood and Marrow Transplant Research (CIBMTR) is collecting data on 45 forms over 5,000 data points. The program uses clustering techniques to analyze the parts of CDEs (Data Element Concept, Object Class and Property and Value Domain concept associations) to discover similar data points. These will be used to help create an HSCT-specific database model in the future (caBIG® 2010 annual meeting poster 95).

The Human Studies Database (HSDB)

UCSF (Leader): Ida Sim (Project Leader), Simona Carini, Rob Wynden

The project goals are to:

- Enhance national clinical and translational research capability: As a computable database with detailed scientific information about past and ongoing human studies, HSDB will be a critical infrastructure for large-scale analysis and reuse of human studies data.
- Ensure standardized computable descriptions of human studies through use of ontologies and controlled clinical vocabularies. The semantic foundation for HSDB is the Ontology of Clinical Research, with mappings to and integration with other relevant models and vocabularies (for example, BRIDG, SNOMED). By adopting ISO/IEC 11179 as the framework for representing the meaning of data, this semantic foundation is independent of the particular technology or architecture used for implementing data sharing.
- Provide a set of tools for institutions to contribute their human studies data to HSDB and to share this data over caGrid. We are focusing first on federating descriptions of study designs before federating study results. Definition and reuse of openMDR and caDSR CDEs will help to ensure consistency in the data collected across these study designs.

Current project members are primarily from Clinical and Translational Science Award (CTSA) institutions, the HSDB project is open and other participating institutions are:

- Duke University: Meredith Nahm, Swati Chakraborty
- Columbia University: Suzanne Bakken
- Johns Hopkins: Harold P. Lehmann
- Mayo Clinic: Chris Chute
- The Rockefeller University: Ed Barbour, Shamim A. Mollah, Knut M. Wittkowski
- Stanford: Samson Tu
- UC Davis: DaVera Gabriel, Hien Nguyen
- University of Colorado: Jessica Bondy
- University of Manchester, UK: Alan Rector
- UT Health Science Center San Antonio: Brad H. Pollock
- UT Southwestern: Herbert K. Hagler, Richard H. Scheuermann
- University of Washington: Jim Brinkley, Todd Detweiller
- Washington University St. Louis: Rakesh Nagarajan, Jahangheer Shaik

This consortium has adopted the caDSR-compatible openMDR as the basis for registering semantic descriptions of data in the HSDB database.

The figure that follows shows the usage process of openMDR by the Human Studies Database (HSDB).
Towards Unambiguous formal descriptions of cancer therapy experiments

Poster (and a paper reference TBD), authors: Alejandra González-Beltrán, PhD et al, Wolfson Institute of Biomedical Research and Department of Computer Science and Cancer Institute, University College London

cadSR CDEs and new CDEs based on ISO/IEC 11179 structure were developed for the Guidelines for Information about Therapy Experiments (GIATE). The CDEs are annotated with concepts from the NCI Thesaurus. Though GAITE was developed independently of NCI Thesaurus, by creating the structured CDEs, using OWL and following a semantic web and linked data approach, an unambiguous and formal description of GIATE was developed from the CDE structures and used to compare GAITE to the ontologies in NCI Thesaurus. "This matching will facilitate interoperability between GIATE compliant knowledge bases and caBIG® services." (caBIG® 2010 annual meeting poster 78)

Ontology based queries for caGrid Infrastructure

Poster, authors: Alejandra González-Beltrán, PhD, Ben Tagger, Eng.D., Anthony Finkelstein, B.Eng, M.Sc., Ph.D., C.Eng, FIET, CTP, FBCS et al, Wolfson Institute of Biomedical Research and Department of Computer Science and Cancer Institute, University College London

Using OWL as a formal language for representing knowledge, this team was able to leverage cadSR mappings between UML Models and NCI Thesaurus concepts and ontology which serve as a conceptual unified view of the data services. This project developed a service for caGrid that provides methods to take cadSR registered models and turn them into OWL. The project has developed a query process in light of OWL2 profiles (caBIG® Annual Meeting poster 79).

Information Management Services, Inc. (IMS) BIS, NAACCR and DCCPS/caSEER program

The program provides NIH, pharmaceutical, academic and other research organizations with biomedical computing support including:

- Web and Application Development
- caBIG® tool adoption and enhancement support
- Repository Management Systems
- RISMA Compliant Computer Center Hosting
- Biomedical Research Data Management and Study Research Support
- Statistical Analysis and Reporting

IMS supports the development of CDEs for cadSR for the Person Age standard and the NAACCR CDEs. The program developed data models for seven data services for DCCPS/caSEER Program for sharing through the grid and has registered the CDEs in the cadSR. The program’s databases contain over 6 million records that can be searched and summarized to produce tables, graphs, and geographic maps for caSEER. Applications developed by IM are designed to extract and publish specimen data using caBIG® CDEs and the caBIG® Common Business Model.
In 2007, Marsha Reichman and Ken Gerlach (CDC) raised the issue about interoperability with NAACCR. As a result, this initiative was funded by PS&CC and NAACCR.

The CTSA Health Ontology Mapper

This application allows users to take local data and map it to ontology codes using caDSR and LexEVS. The application queries against the NCI LexEVS api, as well as caDSR and caGRID. Further information can be found on the Health Ontology Mapper website.

A Growable Network Information System (AGNIS)

MDACC: Dr. Roy Jones, Charles Martinez

AGNIS, or A Growable Network Information System, began as an idea to transmit hematopoietic stem cell transplant data between organizations but developed into a standards-based communications model. AGNIS was originally intended to be built on an existing messaging system. However, analysis of the National Institutes of Health (NIH) caBIG® project produced a strong case for using caBIG® tools, as caBIG® is becoming a recognized leader in creating standards for grid computing and data definition. The mission to connect the cancer community to accelerate research discoveries and improve patient outcomes fits the purpose of AGNIS, which is the implementation of clinical data exchange across the HSCT community to decrease the time it takes for patient follow-up data to be available for research.

AGNIS works by acting as a translator between two centers. When organizations attempt to communicate their data to one another, they must first massage the data into a format that the other organization can recognize. The translation mechanism is actually standardized common data elements (CDEs) curated and stored on the NIH's publicly accessible caDSR. Due to the reusable nature of elements in caDSR, everyone benefits as more terms are created.

**USAGE:** The AGNIS network uses the CDE Browser, Curation Tool and caDSR APIs to create and access standard CDEs.

cancerGrid, cgMDR UK

Project Sponsors: Peter McCallum, Jim Davies, Steve Harris

Website

cancerGrid is an initiative involving scientists at the Universities of Oxford and Cambridge, working together to reduce the cost of clinical research, and to increase its value through effective data sharing. It is building upon the success of a four-year project funded by the UK Medical Research Council, to address a wider range of scientific goals, with support from Microsoft Research.

The cancergrid team have developed standards and tools for the automatic production of the systems needed to support clinical studies and translational research. Their vision:

- The researcher creates a model of their study or dataset, based upon standard templates, using a simple study designer tool;
- The software artifacts needed to run the study or interact with the dataset, such as forms and services, are then produced automatically from the model.

Along with a meta-model (or model template) for clinical studies, based upon the CONSORT statement, and work on the classification and registration of clinical trials, the project has produced an ISO11179-compliant metadata registry, a semantically-aware trials design/management system, and a toolkit for clinical data transformation and integration.

These models and software applications have been tested through initial deployment on a small number of clinical studies in Oxford and Cambridge. The metadata registry software has been adopted for use within the US caBIG initiative, and is being evaluated by a number of organisations within the UK. Development continues at Oxford, with clinical collaboration in Cambridge and London, and technical collaboration with the caBIG team in the US.

**USAGE**: cgMDR was built to be compatible with caDSR, including extensions to ISO 11179 to reference concepts as the grounding for semantics. They have developed an import that can take caDSR CDEs downloaded into an XML file and import them into local cgMDR installations. They have included the "Standard" CDEs in their registry as well as including BRIDG CDEs in their registry download so that local installations may more easily reuse /harmonize with NCI standards as much as possible.

CDC, PHIN Reportable Conditions and Common Core Data Elements

Project Sponsors: Michael Pray, Sundak Gunesan, Catherine Staes

Public health data elements and value sets described in over 90 CSTE position statements which are the foundation building blocks of Public Health Case Reporting (PHCR), ELR and Case Notification. A project is underway collaboratively distribute these data elements and value sets using the existing tools and applications at the National Cancer Institute through the caDSR CDE browser and CDC Vocabulary Server (VADS).

6 - Related Software Engineering Research Highlights

In addition to the adoption of caDSR in cancer and other research, the architecture and application of caDSR is also being viewed and reviewed as an engineering example for construction of collaborative software in research circles'. This chapter outlines some of the examples of this type of use.
"Quality Evaluation of Cancer Study Common Data Elements Using the UMLS Semantic Network"

Authors: Guoqian Jiang, Harold R. Solbrig, Christopher G. Chute

This paper discusses the use of UMLS SemNet to discover disjointedness of semantic network types in caSDR Object Class (OC) and Property (Prop) as a QA mechanism for validating the OC and property in the Data Element Concept. This could be the basis for a new collaborative curation platform. These kinds of rules could be incorporated as OWL reasoning for automated validation as well.

The article also discusses using UMLS SemNet as an alternate classification scheme for browsing content in caDSR.

Additional Examples to be Provided

Page not found for multiexcerpt macro.

The page: 7 - Training Statistics for NCI caCORE and NCI caDSR Tools was not found. Please check/update the page name used in the 'multiexcerpt' include macro.

Appendix A - Usage Statistics for caDSR Resources

This appendix contains information and usage statistics for some of the key caDSR resources. Sections in this appendix include:

- Data Element Descriptions
- Case Report Forms and Surveys
- UML Models
- caDSR Tools and Browser
  - CDE Browser (for the month of June 2011)
  - Curation Tool (as of Feb 2011)
  - Form Builder (as of Feb 2011)
  - UML Model Browser (as of Feb 2011)
  - caDSR API services (as of June 2011)
  - Freestyle services (as of Feb 2011)

Data Element Descriptions

50,818* Community CDEs across all Contexts as of Dec. 2015

*excludes retired CDEs

<table>
<thead>
<tr>
<th>Number of CDEs used</th>
<th>Group Name</th>
<th>Program Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>527</td>
<td>Albert Enistein Cancer Center (AECC)</td>
<td>Cancer Center</td>
</tr>
<tr>
<td>113</td>
<td>Buffet Cancer Center</td>
<td>Cancer Center</td>
</tr>
<tr>
<td>675</td>
<td>Lombardi Cancer Center</td>
<td>Cancer Center</td>
</tr>
<tr>
<td>455</td>
<td>Norton Cancer Center</td>
<td>Cancer Center</td>
</tr>
<tr>
<td>585</td>
<td>OHSU</td>
<td>Cancer Center</td>
</tr>
<tr>
<td>308</td>
<td>USC/NCCC</td>
<td>Cancer Center</td>
</tr>
<tr>
<td>1,096</td>
<td>DCI Duke Cancer Center</td>
<td>Cancer Centers</td>
</tr>
<tr>
<td>199</td>
<td>Adult Brain Tumor Consortium</td>
<td>Consortium</td>
</tr>
<tr>
<td>565</td>
<td>PBTC Pediatric BrainTumor Consortium</td>
<td>Consortium</td>
</tr>
<tr>
<td>202</td>
<td>BOLD</td>
<td>NCI</td>
</tr>
<tr>
<td>3,785</td>
<td>C3D</td>
<td>NCI</td>
</tr>
<tr>
<td>1,900</td>
<td>CCR</td>
<td>NCI</td>
</tr>
<tr>
<td>156</td>
<td>CIP</td>
<td>NCI</td>
</tr>
<tr>
<td>11,631</td>
<td>CTEP</td>
<td>NCI</td>
</tr>
<tr>
<td>4,175</td>
<td>DCP</td>
<td>NCI</td>
</tr>
</tbody>
</table>
As seen in the first chart below, the general trend in total numbers of CDEs has been increasing since the project’s inception in 2002, with 100 to thousands of draft new CDEs added with each new adopter’s project (NCI, NIH, External), and some retired over time.

While the total number of elements in the registry is increasing, the focus in 2010 has been on harmonization and releasing CDEs that are used, vetted, or both by the community and on identifying and retiring unused content as depicted in the second table below. Retiring unused content will make it easier for new users to find good reusable content.

**Chart showing trends in total numbers of CDEs**

![Chart showing trends in total numbers of CDEs](chart_image)

**Chart showing effect of harmonization and retirement of unused content**

![Chart showing effect of harmonization and retirement of unused content](chart_image)
Case Report Forms and Surveys

3,328 Forms created by the Community

- 2,923 CRFs in CTEP used in 611 trials
- 152 in CCR used in 322 Active Trials
- 102 in NHLB®
- 99 in caBig®
- 92 in NIDCR
- 40 in DCP
- 29 in SPOREs
- 910,071 CRFs created in C3D from caDSR CDEs
- 1 in NICHD
- 1 in NHINDS
- 1 in REDCap (Demography)

UML Models

144 Models registered by the Community

- 130 with a Released Status
- NICHD (1)
- Newborn Examinations
- NCI Population Sciences and Cancer Control (5)
- NHIS2-005, caSEER, HINTS2005, SNP500Cancer, GridEnabled Measures

The figure that follows contains a comparison chart showing the number of models loaded to caDSR by year, from 2005 through 2010.
caDSR Tools and Browser

The statistics shown in the sections that follow cover the caDSR tools most commonly used by the community.

CDE Browser (for the month of June 2011)

544 unique visitors to the site

Average number of unique visitors per day:

CDE Browser was linked to by 50 distinct web sites including:

- https://library.openclinica.com/
- https://caintegrator2.nci.nih.gov/
- https://www.phenxtoolkit.org/
- http://crsrestserver.wustl.edu:8080/

Unique visitors past 3 months: Apr 550, May 554, June 544

The top 10 non-NIH domains accessing the CDE Browser were:

- itsg.sdc-moses.com,
- portale-x.east.saic.com,
- vpn.nmdp.org,
- 69.24.154.5 (Nationwide Children S Hospital),
- 141.106.128.110 (Medical College of Wisconsin),
- host226.kai-research.com,
- chi-portal.chi.ohio-state.edu,
- mail.scenpro.com,
- 65-113-146-98.dia.static.qwest.net,
- c-67-99-175-226.roswellpark.org

Curation Tool (as of Feb 2011)

- 129 unique users visited the site this month.
- Curation Tool was linked to by 119 distinct web sites.
- Unique visitors past 3 months: Nov 108, Dec 111, Jan 115
  (Compared with 2010 Nov 539, Dec 430, Jan 490)
- Trend Year/Year - 19% Decrease, Average of 117 unique visitors for past 3 months (144 same period in 2009-10).

The top non-NIH Domains using Curation Tool were:

- verizon.net, nmdp.org, saic.net, comcastbusiness.net, direcpc.net, scenpro.com

Form Builder (as of Feb 2011)

- 101 unique visitors to Form Builder this month.
- Form Builder was linked to by 31 distinct web sites.
- Unique Visitors the past 3 months: Nov 83, Dec 75, Jan 101
- Trend Year/Year - 12% Decrease, Average of 86 unique visitors for past 3 months (97 same period in 2009-10).
For the month, the top 5 non-NIH domains using Form Builder were:

- nmdp.org, Comcast.net, saic.com, verizon.net, direcpc.com

**UML Model Browser (as of Feb 2011)**

- 104 unique users visited the site this month.
- Unique visitors past 3 months: Nov 99, Dec 96, Jan 104
- Trend Year/Years - 22% Decrease, Average 99 unique visitors for past 3 months (138 same period 2009-10)

For the month, the top 5 non-NIH domains using the site were:

- cox.net, duke.edu, nmdp.org, scenpro.com, uams.edu

**caDSR API services (as of June 2011)**

- 161 unique users visited the caDSR API site this month.
- caDSR API was linked to by 17 distinct web sites.
- The top NIH referring URL was the Freestyle API (see statistics below).
- Unique visitors past 3 months: Nov 531, Dec 163, and Jan 63

The top 10 non-NIH domains accessing caDSR API were:


**Freestyle services (as of Feb 2011)**

- 52 unique users visited the site.
- Freestyle was linked to by 12 distinct web sites.
- Unique visitors past 3 months: Nov 33, Dec 33, Jan 43
- Trend – Freestyle was incorporated in SIW for searching for existing caDSR content when annotating models. It is currently used by caIntegrator.
- Average 43 unique visitors per month.

For the month, the top 5 non-NIH domains using the site were:

- UNKNOWN (4,849 accesses), saic.com, cox.net, gatech.edu, as12448.com

**Appendix B - Support for Clinical Trials**

Several cancer centers are supported by caDSR CDE development for their clinical trials. Those institutions are listed here.

**Cancer Centers Supported by caDSR CDE Development for Clinical Trials**

- Abramson Cancer Center of the University of Pennsylvania
- Albert Einstein Medical Center
- ACRIN - American College of Radiology Imaging Network
- Arizona Cancer Center
- UCI/Chao Family Comprehensive Cancer Center
- Duke University
- Massachusetts General Hospital/Dana Farber
- Georgetown University / Lombardi Medical Center
- Johns Hopkins Medical Institute
- MD Anderson Cancer Network
- National Cancer Institute Center for Cancer Research
- National Cancer Institute Cancer Diagnosis Program
- Northwestern University
- Oregon Health & Science University
- University of Colorado Health Sciences Center
- St. Joseph Hospital of Orange
- St. Joseph Hospital / Candler
- UCSF Carol Franc Buck Breast Care Center
- University of Arkansas Medical Sciences
- University of Minnesota
- University of Nebraska Medical Center
- University of Pennsylvania Health Sciences
- University of Washington Medical Center
- Virginia Commonwealth University
Appendix C - Models Registered in caDSR

The appendix lists the models that use and have registered CDEs in caDSR, listed alphabetically and by version, within the NCI program/context owner. Be advised that some of these models are retired. Details about each of the models, as registered in caDSR, can be found by browsing the models via the UML Model Browser.

This appendix includes the following sections.

- caBIG® (NCI cancer Biomedical Informatics Grid) Data Services
- PS&CC (NCI Population Sciences & Cancer Control)
- caCORE (NCI CORE Infrastructure)

caBIG® (NCI cancer Biomedical Informatics Grid) Data Services

AIM 1.0, 1.5, 2.0, 3.0
Bioconductor 1.0
BiospecimenCoreResource 1.0
BRIDG 1.0, 2.1
C3D Connector 1.5
C3PR 1.0, 2.0, 2.8
caAERS 1.0, v2.0
caArray 1.1, 2.0, 2.1, 2.4 (Internal)
caBIO 4.0, 4.1, 4.2, 4.3
cacORRECT 1.0
CAD Markup 1.0
CAD Order 1.0
caDSR 4.0
cElimir 1.0, 2.0
caFE Server 2.0
caGrid_Metadata_Models 1.0
caIntegrator 2.0, 2.1
Caisis 3.5
caLIMS2 1.1
caMOD_2.5
caNano 1.0
caNanoLab 1.0, 1.4, 1.5
Cancer Molecular Pages 1.0
CAP Cancer Checklists 1.0
Cardiovascular Model 1.0
caTIES 1.0, 2.0
caTISSUE CAE 1.2
caTISSUE Core 1.0, 1.1, 1.2
ccaTissue_Core_caArray 1.0
ccaTissue_Suite 1.0, 1.1, 1.2
caTRIP Annotation Engine 1.0
caTRIP Tumor Registry 1.0
caXchange 1.0
CDC NCPHI Proof of Concept 1.0
Center for Epidemiologic Studies Depression Scale (CES-D) 1.0
CGWB 2.0
ChemBank 1.1
Chromosomal Segment Overlap Finder Across Samples 1.0
Clinical Trials Lab Model 1.0
Clinical Trials Object Data System (CTODS) .53
CoCaNUT (Colon Cancer Knowledge Utility Toolbox) 1.0
Copy Number Analysis Tool 1.0
CTEP Enterprise Services 1.1
CTMS Metadata 1.0
DemoService 1.0
DigitalModelRepository (DMR) 1.0
DNAcopy Analytical Service 1.0
DSD (Dynamic Service Deployment) 1.0
EVS Core Grid Analytical Service 1.0
GeneConnect 1.0
GeneNeighbors 1.0
GenePattern 1.0
GenePattern Based Copy Number Analytical Service 1.0
Generic Image 1.0
Generic Parameters 1.0
Genomic Identifiers 1.0
geworkbench 1.1
GoMiner 1.0
Grid-enablement of Protein Information Resource (PIR) 1.1, 1.2
ImageMiner 1.0
ISO21090v1_0 1.0
Additions to this List

Publications Citing caDSR and CDEs

Papers, Abstracts, Posters, and Talks

Appendix D - Publications

This appendix includes the following sections:

- Additions to this List
- Publications Citing caDSR and CDEs
- Papers, Abstracts, Posters, and Talks

Additions to this List

To have a citation added please contact the NCI Help Desk.

Publications Citing caDSR and CDEs

3. Elizabeth J. Conwin, PhD, RN, Shirley M. Moore, PhD, RN, Andrea Plotsky, MPH, Margaret M. Heikember, PhD, RN, Susan G. Dorsey, PhD, RN, Drenna Waldrop-Valverde, PhD, Donald E. Bailey Jr., PhD, RN, Sharron L. Docherty, PhD, RN, Joanne D. Whitney, PhD, RN, Carol M. Musil, PhD, RN, Cynthia M. Dougherty, PhD, APRN, Donna J. McCloskey, PhD, RN, FAAN, Joan K. Austin, PhD, RN, & Patricia A. Grady, PhD, RN,
Papers, Abstracts, Posters, and Talks


43. Yu Rang Park, MS and Ju Han Kim, M.D., Ph.D."Metadata registry and management system based on ISO 11179 for cancer clinical trials information system", AMIA Annual Symposium, 2006


