caDSR Wiki

Contents of this Page

caDSR Links and Information

- Current caDSR Documentation
- Link to OneData
- caDSR II API Home Page
- Privacy Policy Statement

Related Links

- Enterprise Vocabulary Services (EVS) wiki

CBIIT Links

- CBIIT website
- NCI Cancer Data Science on X (formerly Twitter) @ NCIDataSci
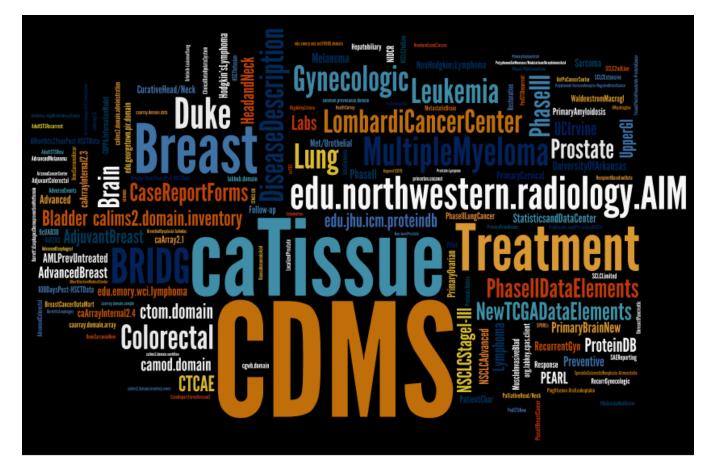
Metadata Needs and Support

The caDSR (Cancer Data Standards Registry and Repository) is a metadata repository based on the ISO/IEC 11179 Metadata Registry standard. It supports a broad community of users both inside and outside of NCI that have requirements to ensure the longevity and consistency of biomedical research data by registering metadata standards in caDSR.

What is a metadata standard? A metadata standard is a high level document which establishes a common way of structuring and understanding data, and includes principles and implementation issues for utilizing the standard. There are many metadata standards purposed for specific disciplines.

Content owners and end users include NCI and its partners in clinical trials, academic institutions (including NCI Designated Cancer Centers, SPOREs and NCTN/ETCTN), other NIH institutes (including NICHD, NHLBI, NCATS and NIDCR), other federal agencies (in particular the FDA), pharmaceutical companies, standards development organizations (e.g. CDISC) and a range of international biomedical organizations.

Requirements from researchers and/or their supporting informatics groups drive the creation of metadata in the caDSR. Metadata content development usually starts with a request for assistance by a researcher planning clinical or research data collection. Metadata curators work with the user and EVS to identify appropriate vocabulary while identifying a mix of new and existing CDE content to support the scientific requirement. Curators *always* attempt to reuse existing metadata (where that content supports the scientific requirement) as a way to help scientists ensure the compatibility of their data with other data collected across the enterprise, and sometimes researchers request that their content be harmonized with specific existing projects.

The word cloud that follows illustrates the broad variety of collections of data elements that are reposited in the caDSR for various communities and types of studies.

About caDSR

CBIIT's management of metadata began as part of an effort to support CTEP's reporting for breast cancer trials, and from a need to develop and disseminate standards that would ensure consistency and accuracy in reporting across the NCI Clinical Trial Network (NCTN/ETCTN) and Lead Protocol Organizations (LPOs). This led to the establishment of a centralized resource and associated web-based tools for creating, clearly documenting, and sharing human- and machine-readable data descriptions. The need to maintain and share data about data, or metadata, became the basis for the NCI's repository of CDEs, metadata, and data standards, what is now known as the caDSR. A CDE Steering Committee was formed to define what kind of metadata was needed for the repository. Driven by the needs from community to create, share, and manage CDEs over time, a set of metadata attributes was established, which included attributes such as human friendly name(s), text definition(s), valid values, unique identifiers, and workflow status. Consultation with appropriate experts identified ISO 11179, an international standard for metadata registries, as meeting the needs identified by the CDE Steering committee. As time went on, more groups wanted to record their data elements and share them via the caDSR, so additional features were added, including extensions of ISO 11179 to enable storage of metadata describing Case Report Forms (CRFs) that use CDE metadata as the basis for questions on the CRFs.

As more groups recorded their data elements in caDSR, the difficulty in creating high-quality names and definitions for data elements became recognized as a best practice for clarifying the meaning of the data, but also a challenge for data-element curators. Consequently, ISO 11179-5 Naming Principles were used to establish naming conventions that could be applied across groups. The same naming conventions are used in the National Information Exchange Model (NIEM). Since NCI had the EVS terminology services available, this was seen as a reasonable means by which to aid this task by giving curators access to well-formed and NCI-preferred names for concepts that form the name of the CDEs. At present, curators find concepts in EVS based on synonym or concept id searches, and the EVS preferred-term name and definition streamline their task. The challenge of ensuring that duplicate CDEs were not created led to leveraging the parts of the ISO 11179 metamodel along with a preference for the use of NCIt concepts so the system could semi-automatically recognize and promote reuse of existing content. NCIt is a specialized cancer terminology that includes additional knowledge from the literature about these concepts and is modeled as an OWL ontology were concepts have various types of relationships and mappings defined to other concepts. It also includes mappings to the UMLS where they exist. Therefore, a link to NCIt concepts from caDSR content can help test similarity between CDEs. Links to NCIt concepts can also provide access to researchers for exploring the meaning of a data that conforms to a given CDE in greater depth.

Although this activity began as a means to support CTEPs trials networks, the caDSR now supports a much wider audience. This includes clinical trials run by the NCI intramural program, the Center for Cancer Research (CCR), and Division of Cancer Prevention (DCP); Specialized programs of Research Excellence (SPOREs), Cancer Centers and other academic medical centers, other NIH institutes and centers as well as standards groups such as CDISC and a variety of international partners.

A note on the term "Common Data Element" (CDE). While originally intended to mean a data element that was reused across groups, the term has come to mean any description of a variable and its valid values. In this document, we will use the standard NCI version of this definition, which is to say a variable description (including valid values) described in the caDSR using its implementation of the ISO 11179 variable, regardless of whether the element in question has been used more than once.

This Wiki Space

This is the home page for the caDSR wiki space. You may edit pages if you are working on them with the authors. You are welcome to leave comments. This wiki space includes the following pages:

Documentation

For a complete list of current caDSR Tool user documentation, application guides, release notes, and FAQs, see the caDSR Documentation wiki page.

Email Lists and Forums

List/Forum	Email address or URL	Description
caDSR Users List	CADSR_USERS@LIST.NIH.GOV	Archive for content users such as Curators
caDSR Developers List	CADSR_SOFTWARE_DEVELOPERS@LIST.NIH.GOV	Archive for developers using caDSR Metadata, such as UML Model owners (subscription required)
caDSR Tools Download List	CADSR_TOOLS_DOWNLOAD@LIST.NIH.GOV	For adopters
All NIH List Servers	No email address	Index of all NIH mail lists
No list/forum	NCIAppSupport@mail.nih.gov	NCI Application Support

How to Cite caDSR

To cite the NCI Semantic Infrastructure, use the following reference.

Komatsoulis, G.A., Warzel, D.B., Hartel, F.W., Shanbhag, K, Chilukuri, R, Fragoso, G., de Coronado, S, Reeves, D.M., Hadfield, J.B., Ludet, C., and P.A. Covitz (2007) "caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability." *Journal of Biomed Informatics.* 2008 February; 41(1): 106--123. Published online 2007 April 2. doi: 10.1016/j.jbi.2007.03.009.

To cite the caDSR OneData general software, use the following reference.

NCI caDSR OneData. <https://cadsr.cancer.gov/onedata/Home.jsp> National Cancer Institute, Center for Biomedical Informatics and Information Technology, 01 Oct. 2010. Web. 17 Jan. 2013.

To cite a specific collection of forms in caDSR, or to cite caDSR CDEs in a particular collection or group, use a reference like the following.

"[Protocol or Classification Scheme Name]." [Context name], [Node name], NCI caDSR OneData. <https://cadsr.cancer.gov/onedata/Home.jsp> National Cancer Institute, Center for Biomedical Informatics and Information Technology, 01 Oct. 2010. Web. 17 Jan. 2013.

Example: " CALGB: 10603 Treatment Form ." CTEP, Protocol Forms, NCI caDSR OneData. <https://cadsr.cancer.gov/onedata/Home.jsp> National Cancer Institute, Center for Biomedical Informatics and Information Technology, 01 Oct. 2010. Web. 17 Jan. 2013.

Example: "NAACCR 11.1." PS & CC (NCI Population Sciences & Cancer Control), Classifications, Division of Population Cancer Control and Population Sciences, NCI caDSR OneData. <https://cadsr.cancer.gov/onedata/Home.jsp> National Cancer Institute, Center for Biomedical Informatics and Information Technology, 01 Oct. 2010. Web. 17 Jan. 2013.