

Metadata as the Key to Semantic Interoperability in Clinical Data Systems

Brian Campbell
The EMMES Corporation

Society for Clinical Trials May 16 – 18 Vancouver, Canada
May 16, 2011

NCI's Call for Standardization in Clinical Data Collection

- "Armitage Report"
 - August 1997
 - Call for uniform data collection
 - to promote efficient protocol implementation
 - to enhance the ability to share and compare data

Birth of the CDE Initiative

- Sponsored by NCI's Cancer Therapy Evaluation Program (CTEP) with support from the NCI Center for Bioinformatics
- Initiated for Phase III NCI-sponsored Cooperative Group oncology trials
- Mandated the use of CDEs as semantically annotated metadata

What are Metadata?

Metadata serves as the semantic backbone of the CDE. Metadata are descriptive terms used to unambiguously define the data, but are not the data themselves.

A search for 'Agent' retrieves the following:

Agent Name	Taxol
NSC Number	007

Do you really know what you have?

Without controlled, conceptually based metadata you can't be certain.

FDA Metadata

Agent - A chemical compound administered to a human being to treat an existing disease or condition, or prevent the onset of a disease or condition.

NSC Number - Identifier given to a chemical compound by the Nomenclature Standards Committee of the US Food and Drug Administration (FDA).

Agent Name - Taxol [The trade name of an antineoplastic medication that interferes with the growth of cancer cells and slows their growth and spread in the body. Taxol is used in the treatment breast, ovary and lung cancers, and AIDS-related Kaposi's sarcoma.]

NSC Number - 007 [The numeric identifier given to the chemical compound Taxol [paclitaxel] by the Nomenclature Standards Committee of the US Food and Drug Administration (FDA).]

CIA Metadata

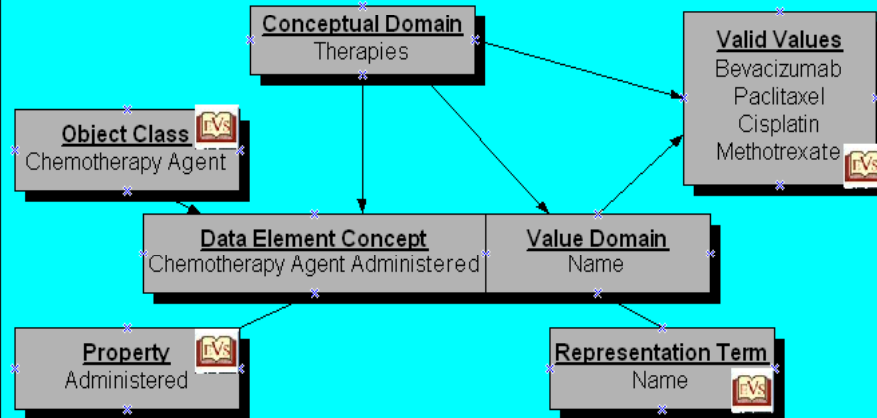
Agent - A sworn intelligence agent; a spy.

NSC Number - Coded identifier given to an intelligence agent by the National Security Council.

Agent Name - Taxol [Code name given to Valerie Plame, an undercover agent of the Central Intelligence Agency (CIA)]

NSC Number - 007 [The code number used to identify Valerie Plame, an undercover agent of the CIA.]

caDSR ISO11179 Implementation



EVS Tie-In

The image displays two screenshots from the NCI Terminology Browser. The left screenshot shows the 'Graft Versus Host Disease' entry. It includes a table of identifiers with columns for name and code, and a section for relationships to other concepts. The right screenshot shows the 'NCI Thesaurus Taxonomy' tree, with 'Graft Versus Host Disease' highlighted under the 'Diseases and Disorders' category.

NCI Terminology Browser - Microsoft Internet Explorer

Address: http://ncicb.nci.nih.gov/NCITBrowser/GetStateAndProperty.do?toolmarktag=1&conceptname=Graft_Versus_Host_Disease

Links: [of_coder_banner\[1\].gif](#)

ISSUES

Graft Versus Host Disease [Printable Page](#) [History](#) [Graph](#)

Identifiers:

name	Graft_Versus_Host_Disease
code	C3063

Relationships to other concepts:

Disease_Has_Finding	Immunodeficiency
<input type="checkbox"/>	<input type="checkbox"/>

Information about this concept:

DEFINITION
NCI-GLOSSIA reaction of donated bone marrow or peripheral stem cells against a person's tissue.


DEFINITION
NCI|An incompatibility reaction (which may be fatal) in a subject (host) of low immunological competence (deficient lymphoid tissue) who has been the recipient of immunologically competent lymphoid tissue from a donor who lacks at least one antigen possessed by the recipient host; the reaction, or disease, is the result of action of the transplanted cells against those host tissues that possess the antigen not possessed by the donor. Seen most commonly following bone marrow transplantation, acute disease is seen after 5-49 days and chronic disease weeks to months after

NCI Thesaurus Taxonomy

- Abnormal Cell
- Anatomic Structure, System, or Substance
- Biochemical Pathway
- Biological Process
- Chemotherapy Regimen
- Clinical or Research Activity
- Conceptual Entity
- Diagnostic, Therapeutic, and Research Equipment
- Diagnostic or Prognostic Factor
- Disease, Disorder or Finding
 - Diseases and Disorders
 - Behavior-Related Disorder
 - Cancer-Related Condition
 - Amyloidosis
 - Graft Versus Host Disease**
 - Acute Graft Versus Host Disease
 - Chronic Graft Versus Host Disease
 - Intestinal Graft Versus Host Disease
 - Meigs' Syndrome
 - Paraneoplastic Syndrome
 - Precancerous Condition
 - Spinal Cord Compression
 - Superior Vena Cava Syndrome
 - Tumor Lysis Syndrome
 - Veno-Occlusive Disease
 - Congenital or Acquired Anatomic Abnormality
 - Degenerative Disorder

Two-Tiered Process

- Step 1: Develop a set of disease-specific template case report forms (CRFs) that represent 80% of the questions routinely incorporated into an average set of CRFs
- Step 2: Implement a review process to ensure compliance and allow for incorporation of new questions as specific protocols dictate



Step 1 Template CRF Development A: Review Existing CRFs

- Identify areas of clinical trial specialization
- Gather active CRFs from participating sites
- Identify reporting standards
- Create "strawman" CRFs based on analysis



Step 1 B: Consensus Building

- Collaborative committee process
- Diversify the committee members
(Statisticians, Data Monitors, CRAs, Research Nurses, MDs, etc.)
- Face-to-face meetings and conference calls

Step 1

C: Curation of CDEs

- Utilize a metadata standard (i.e., ISO 11179)
- Use a controlled, concept-based vocabulary or ontology (i.e., NCI Enterprise Vocabulary Service (EVS))
- Employ publicly-available web-based tools (i.e., NCICB Data Standards Repository (caDSR))

Step 2 Compliance Review Process

A: Define the Process

- Ensure all stakeholders understand respective responsibilities
- Develop lines of communication
- Make all stakeholders accountable


Step 2

B: Allow for Evolution

- Allow for scientific advances
- Process should be dynamic to allow for efficiencies
- Leverage emerging technology (e.g., Oncology Patient Enrollment Network (OPEN), electronic data capture systems, data exposure systems)

Current Template Development Status

- Developed template CRFs for
 - 11 individual cancers
 - Bone Marrow Transplant Clinical Trials Network
 - National Institute for Dental and Craniofacial Research
 - AIDS Malignancy Consortium
- Currently under development
 - Substance Abuse module for an electronic health record sponsored by the National Institute for Drug Abuse



Current Compliance Review Status

- CDE-compliance review performed on more than 200 protocols and associated CRFs
- Provided ongoing training to interested stakeholders to keep CDEs front and center
- Played integral role in leveraging OPEN and Medidata Rave for use by the Cooperative Groups



Questions?