

# **A Collaborative Framework for Representation and Harmonization of Clinical Study Data Elements Using Semantic MediaWiki**

**Guoqian Jiang, Harold R. Solbrig (Mayo Clinic)**

**Dave Ibersen-Hurst, Rebecca D. Kush (CDISC)**

**Christopher G. Chute (Mayo Clinic)**

**1 Division of Biomedical Statistics and Informatics, Mayo Clinic  
College of Medicine, Rochester, MN, 55905**

**2 Clinical Data Interchange Standards Consortium (CDISC), Austin,  
TX, 78746.**

# Introduction

- **The infrastructure and information generated in biomedicine are largely disconnected and disjoint.**
- **Semantic interoperability - fundamental and critical for sharing information.**
  - **Data elements**
  - **Terminologies**
  - **Information models**

# Cancer Data Standards Registry and Repository (caDSR)

- ISO/IEC 11179 standard for metadata registries
- To represent the common data elements (CDEs) in the database
- implemented a set of APIs and tools



# Challenges for Metadata Community

- **The community is facing the harmonization scaling problem.**
- **The need for tooling to navigate the model space is urgent.**
- **To form better community adoption and governance, a more open, scalable and collaborative platform is desired.**

# Vision for Clinical Data Interchange Standards Consortium (CDISC)

- To build a global, accessible electronic library, which enables precise and standardized data element definitions that can be used in applications and studies to improve biomedical research and its link with health care.

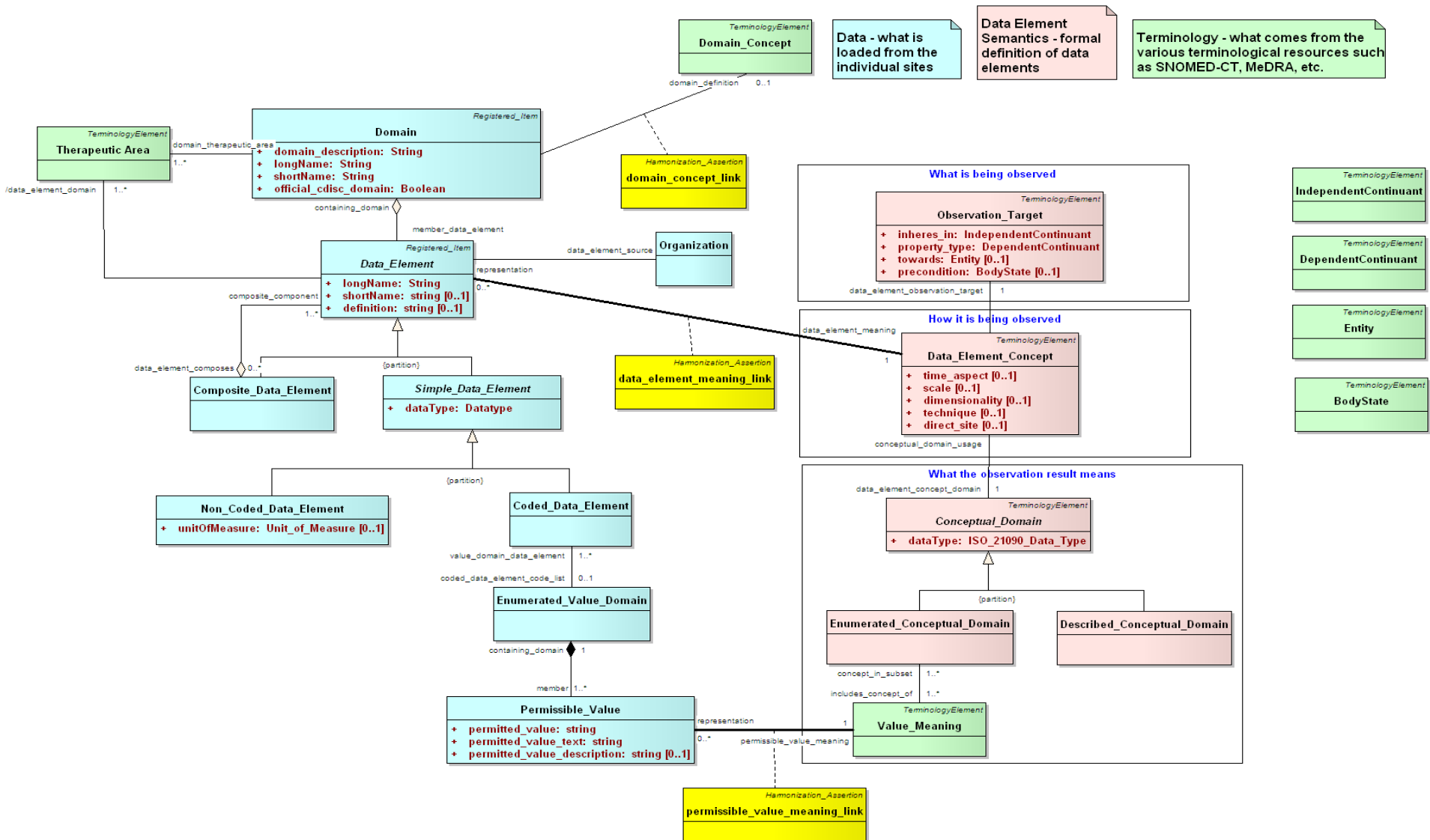
# Semantic Wiki/LexWiki

- Wiki as a collaborative system – community generated content.
- Semantic wiki as an platform – support different levels of the formality continuum (Free text -> OWL).
- LexWiki - a collaborative authoring platform for large-scale biomedical terminologies.
  - BiomedGT – Biomedical Grid Terminology
  - CTCAE - Common Terminology Criteria for Adverse Events
  - WHO ICD11 – the International Classification of Disease
  - NeuroLex - the Neuroscience Lexicon
  - XMDR - eXtended MetaData Registry (XMDR) Project
  - CSHARE - CDISC Shared Health and Research Electronic Library

# Objectives

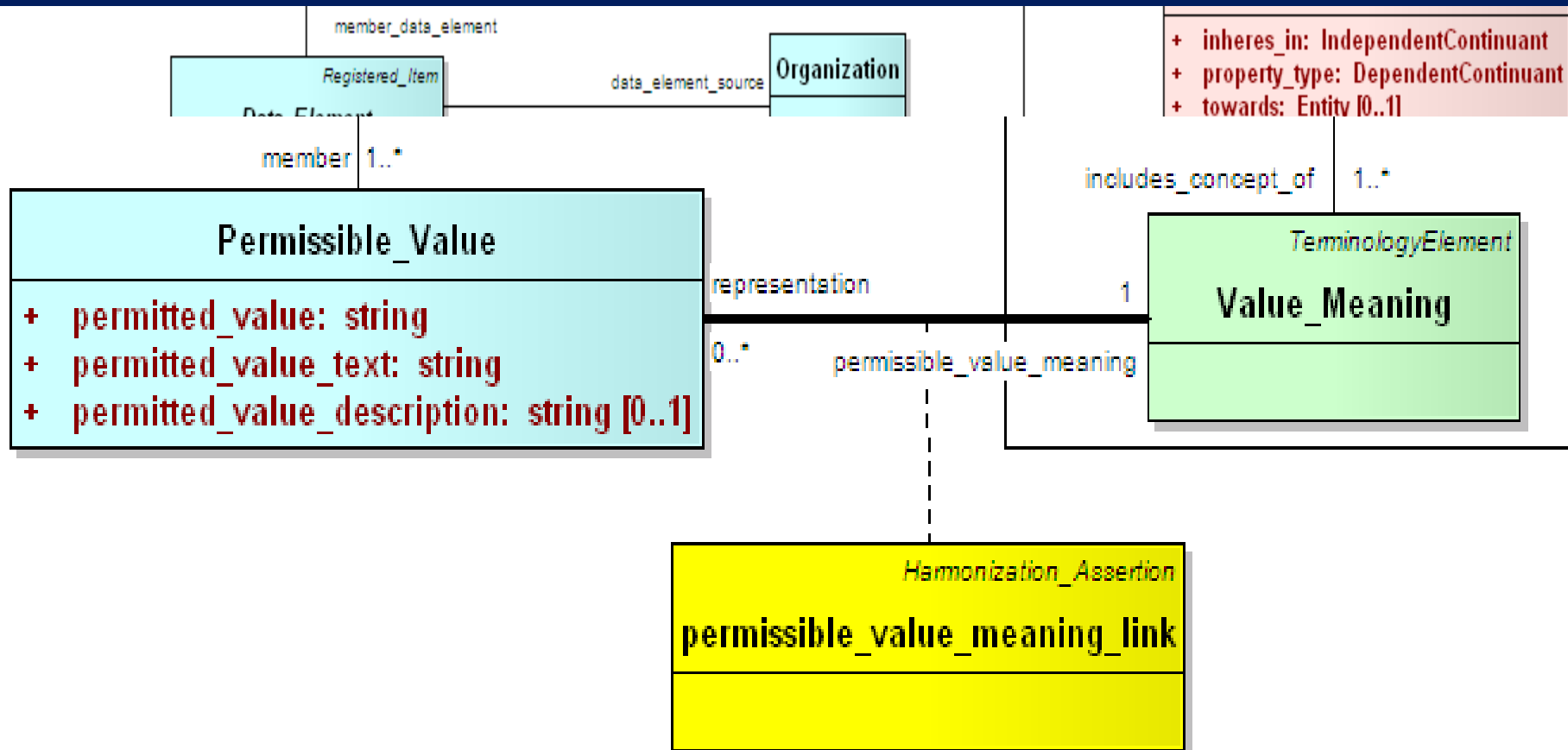
- **We propose a collaborative framework for representation and harmonization of clinical study data elements.**
- **We implement a prototype of CDISC Shared Health and Research Electronic Library (CSHARE) using Semantic MediaWiki.**
- **We report the preliminary observations and evaluations of how the components worked and the lessons learnt.**

# Representation and Harmonization Framework in UML Model (Compatible with ISO 11179)

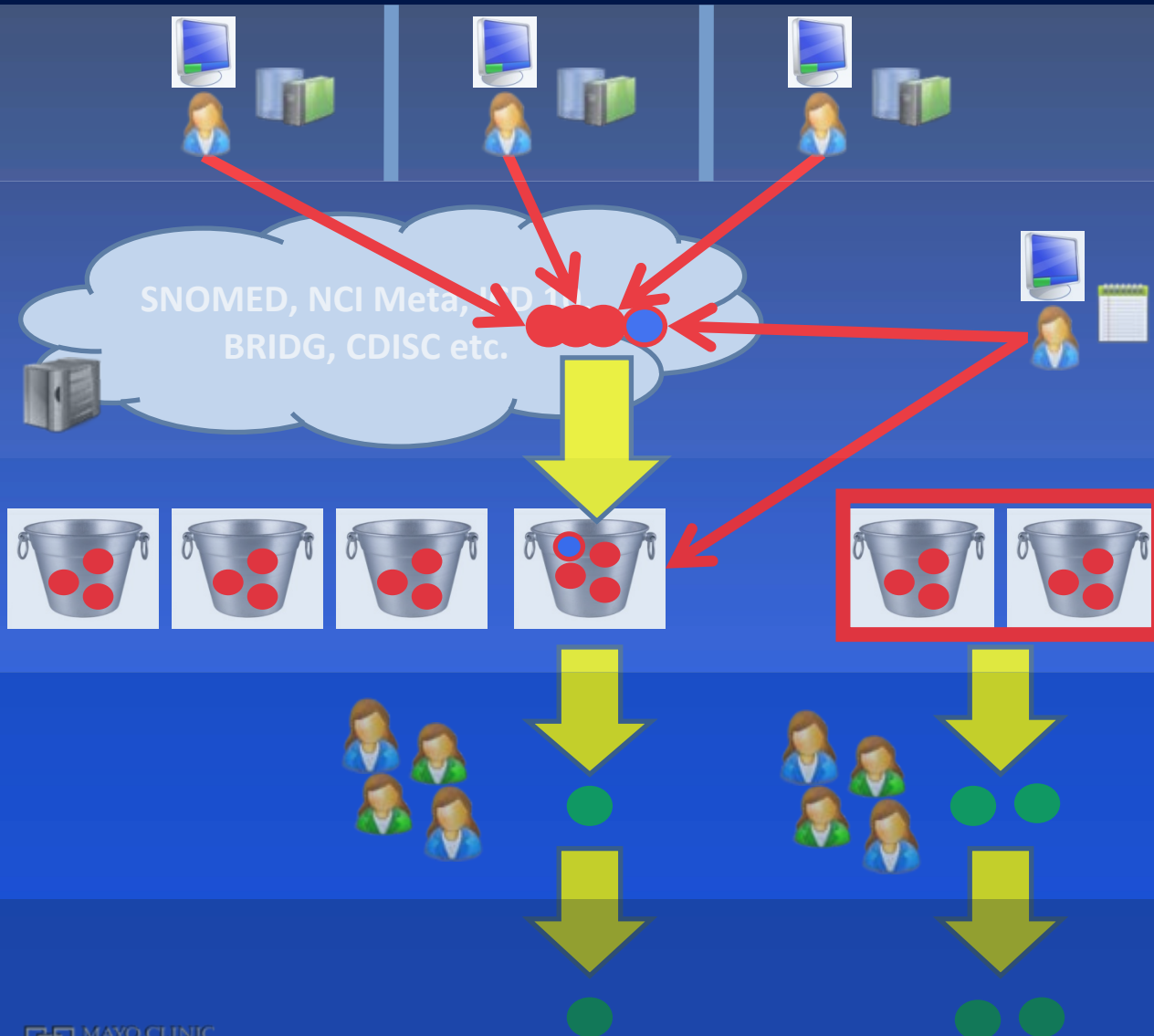




# Representation and Harmonization Framework in UML Model (Compatible with ISO 11179)



# Workflow process



Contribute

Link

Merge

Harmonize

New Definition

# CSHARE Prototype

- **A collaborative effort between Mayo Clinic, National Cancer Institute (NCI) and CDISC to evaluate the potential uses of collaborative technology in CSHARE project.**
  - **3 month time frame**
  - **Semantic MediaWiki Based**
  - **Five major participants:**
    - **MD Anderson**
    - **Eli Lilly**
    - **Glaxo Smith Klein (GSK)**
    - **Genzyme**
    - **Mayo Clinic**

# Baseline Terminology

- >300K entries
- LexGrid Model
- LexWiki Templates
- Wiki Categories
- Service Tools

CDISC Wiki

Main Page > SNOMED CT >

page discussion view source history

## SNOMED CT

### SNOMED-CT

- [−] Systematized Nomenclature of Medicine Clinical Terms(C0000000)
- [−] Body structure(C1268086)
  - [+] Acquired body structure(C0459112)
  - [+] Anatomical organizational pattern(C0687074)
  - [+] Anatomical site notations for tumor staging(C0475435)
  - [+] Morphologically altered structure(C1260954)
  - [\*] Nonspecific site(C0457192)
  - [\*] Normal anatomy(C1269644)
  - [+] Physical anatomical entity(C0506706)
  - [+] Topography not applicable(C0229959)
  - [\*] Topography not assigned(C0205627)
  - [\*] Topography unknown(C0229958)
- [+] Environment or geographical location(C0586399)
- [+] Event(C0441471)
- [+] Finding(C0037088)
- [+] Intervention or Procedure(C0184661)
- [+] Linkage concept(C0332280)
- [+] Medication(C0013227)
- [+] Object(C0347997)
- [+] Observable entity(C1285578)
- [+] Organism(C0029235)
- [+] Physical force(C0336996)
- [+] Qualifier value(C1285164)
- [+] Record artifact(C1689941)
- [+] Situation with explicit context(C0420206)
- [+] Social context(C0037414)
- [+] Special concept(C1298232)
- [+] Specimen(C0370003)
- [+] Staging and scales(C0432488)
- [+] Substance(C0439861)

CDISC Terminologies

- SNOMED CT
- NCI Thesaurus
- HL7 Version 3
- MeDRA
- ICD-9-CM
- ICD-10
- BRIDG
- ISO 21090

Toolbox

Browser

Powered By LexWiki

This page was last modified on 19 August 2009

# Individual Data Element - Description

[page](#)[discussion](#)[view form](#)[view source](#)[history](#)

## GSK Lesion Measurement Previously irradiated

(Curation Status: [InProgress](#))

[Description](#)[Values](#)[ConceptReference](#)

### Data Element Details

**Long Name:** Previously irradiated

**Short Name:** PREVIR

**Definition:** Whether a tumor was irradiated before the start of the study.

**Source Name:** [GSK](#)

**Data Type:** [CD](#)

**Units:** NO\_UNITS\_SPECIFIED

**Domain:** [Oncology Lesion Measurement](#)

**Therapeutic Area:** [Oncology](#)

# Individual Data Element – Value Set





[page](#)[discussion](#)[view form](#)[view source](#)[history](#)

## GSK Lesion Measurement Previously irradiated

(Curation Status: [InProgress](#))

[Description](#)[Values](#)[ConceptReference](#)

### Value Set: [GSK ZYESNO Value Set](#)

	 value code	 value name	 concept reference
<a href="#">GSK ZYESNO Value Set/N</a>	N	No	
<a href="#">GSK ZYESNO Value Set/U</a>	U	Unknown	
<a href="#">GSK ZYESNO Value Set/X</a>	X	Not applicable	
<a href="#">GSK ZYESNO Value Set/Y</a>	Y	Yes	
<a href="#">GSK ZYESNO Value Set/Z</a>	Z	Not done	

# Individual Data Element – Concept Reference

page

discussion

view form

view source

history

## GSK Lesion Measurement Previously irradiated

(Curation Status: [InProgress](#))

Description

Values

ConceptReference

### Data Element Concept

Concept Reference: [NCIM Irradiation\(C1282930\)](#) (**SNOMED**)

Concept Reference: [NCIM VPersonal history of irradiation\(C0481620\)](#) (**SNOMED**)

# Collaboration Feature – Discussion Tool

page

discussion

view source

history

refresh



## Talk:GSK Lesion Measurement Previously irradiated

---

I believe that radiation is a substance administration in BRIDG. I'm not sure how one would represent the fact that the "substance" is radiation. Is radiation a kind of Material? Maybe whether a tumor was previously irradiated involves `Material.name=Radiation`, `PerformedSubstanceAdministration.actualDateRange` before start of study, `PerformedSubstanceAdministration.targetSite=` this lesion.



# Underlying Semantic Annotations - RDF Triples

special page



## Browse wiki

### GSK Lesion Measurement Previously irradiated

<b>11179 coded data element code list</b>	GSK ZYESNO Value Set + ⓘ
<b>11179 containing data set</b>	Oncology Lesion Measurement + ⓘ
<b>11179 dataType</b>	DT CD + ⓘ
<b>11179 data element domain</b>	Oncology + ⓘ
<b>11179 data element meaning</b>	NCIM Irradiation(C1282930) + ⓘ, NCIM VPersonal history of irradiation(C0481620) + ⓘ
<b>11179 data element source</b>	GSK + ⓘ
<b>11179 definition</b>	Whether a tumor was irradiated before the start of the study.
<b>11179 longName</b>	Previously irradiated + 🔍
<b>11179 shortName</b>	PREVIR + 🔍
<b>11179 unitOfMeasure</b>	NO_UNITS_SPECIFIED + 🔍
<b>Modification date</b>	27 August 2009 16:04:16 + 🔍
<b>Categories</b>	CSHARE WorkflowCurationStatusInProgress, GSK Data Element

hide properties that link here

# RDF/OWL Rendering

```
- <rdf:RDF>
  <!-- Ontology header -->
  - <owl:Ontology rdf:about="">
    <swikt:creationDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-03-07T19:04:18-06:00</swikt:creationDate>
    <owl:imports rdf:resource="http://semantic-mediawiki.org/swikt/1.0"/>
  </owl:Ontology>
  <!-- exported page data -->
- <swikt:Subject rdf:about="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/GSK_Lesion_Measurement_Previously_irradiated">
  <rdfs:label>GSK Lesion Measurement Previously irradiated</rdfs:label>
  <swikt:page rdf:resource="http://informatics.mayo.edu/cshare/index.php/GSK_Lesion_Measurement_Previously_irradiated"/>
  <rdfs:isDefinedBy rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:ExportRDF/GSK_Lesion_Measurement_Previously_irradiated"/>
  <rdf:type rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Category-3ACSHARE_WorkflowCurationStatusInProgress"/>
  <rdf:type rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Category-3AGSK_Data_Element"/>
  <wiki:Property-3A11179_coded_data_element_code_list rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/GSK_ZYESNO_Value_Set"/>
  <wiki:Property-3A11179_containing_data_set rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Oncology_Lesion_Measurement"/>
  <wiki:Property-3A11179_dataType rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Category-3ADT_CD"/>
  <wiki:Property-3A11179_data_element_domain rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Oncology"/>
  <wiki:Property-3A11179_data_element_meaning rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Category-3ANCIM_Irradiation(C1282930)"/>
  <wiki:Property-3A11179_data_element_meaning rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/Category-3ANCIM_VPersonal_history_of_irradiation(C0481620)"/>
  <wiki:Property-3A11179_data_element_source rdf:resource="http://informatics.mayo.edu/cshare/index.php/Special:URIResolver/GSK"/>
- <wiki:Property-3A11179_definition rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
  Whether a tumor was irradiated before the start of the study.
  </wiki:Property-3A11179_definition>
  <wiki:Property-3A11179_longName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Previously irradiated</wiki:Property-3A11179_longName>
  <wiki:Property-3A11179_shortName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PREVIR</wiki:Property-3A11179_shortName>
  <wiki:Property-3A11179_unitOfMeasure rdf:datatype="http://www.w3.org/2001/XMLSchema#string">NO_UNITS_SPECIFIED</wiki:Property-3A11179_unitOfMeasure>
</swikt:Subject>
```

# Harmonization Assertion

## — Lexical definition and ISO 21090 datatypes

page discussion edit with form edit history delete move protect watch refresh

### Edit LexWiki CSHARE Data Element: GSK Lesion Measurement Previously irradiated

Description Values Concept Reference

Curation status:  (Please assign a curation status first.)

Long Name

Name:

Short Name

Name:

Definition

Definition:

Source Name

Source Name:

Data Type

Data Type:

Unit

Unit:

Dataset Name

Domain Name:

Therapeutic Area:

#### DEFINITION

A CD is a reference to a concept defined in an external code system, terminology, or ontology.

#### DT CD

DT LVP CD

DT BAG CD

DT LIST CD

DT NPPD CD

DT COLL CD

DT DSET CD

DT HIST CD

# Harmonization Assertion

## — Concept Reference and BRIDG linkage

### Concept Reference

Concept Reference:


BRIDG


### DEFINITION

The anatomical site or system that is the focus of the observation. For example, gastrointestinal, cardiovascular,

Equivalent BRIDG

 BRIDG PerformedObservation targetSiteCode

 BRIDG PerformedObservationResult targetSiteCode

 BRIDG PerformedObservation targetSiteLateralityCode

# Slicing and Dicing

page

discussion

edit

history

delete

move

protect

watch

refresh



## CSHARE Lesion Measurement

**TILES** • TABLE

sorted by: [page name](#); then by...  grouped as sorted

- Genzyme Lesion Measurement METHOD**

domain	Oncology Lesion Measurement
source	Genzyme
datatype	DT CD
concept	NCIM Measurement method(C1299991)
bridg mapping	BRIDG PerformedObservation methodCode
is mapped	true
maps from	CDE Lesion Measurement Method of Measurement, coded
value set	Genzyme METHOD6 Value Set
long name	METHOD
short name	METHOD
definition	Method used to measure the specified lesion

- GSK Lesion Measurement Method**

domain	Oncology Lesion Measurement
source	GSK
datatype	DT CD
concept	NCIM Measurement method(C1299991)
bridg mapping	BRIDG PerformedObservation methodCode
is mapped	true
maps from	CDE Lesion Measurement Method of Measurement, coded

**source**



1 Genzyme  
1 GSK  
1 Mayo

**is mapped**

1

1 false   
3 true

**datatype**

1

3 DT CD   
1 DT INT   
1 DT ST

**concept**

1

size(C0449453)   
1 NCIM Lesion(C0221198)   
3 **NCIM Measurement method(C1299991)**   
1 NCIM Palnable(C0522499)

**bridg mapping**

3 BRIDG PerformedObservation methodCode

# New Definition

[page](#)[discussion](#)[edit with form](#)[edit](#)[history](#)[delete](#)[move](#)

## CDE Lesion Measurement Method of Measurement, coded

[Harmonization](#)[Description](#)[Clump](#)[ConceptReference](#)

### Equivalent Data Elements

Property Name	Property Value	Matchiness	Comments
Data Element From Genzyme:	<a href="#">Genzyme Lesion Measurement METHOD</a>	(none)	(none)
Data Element From GSK:	<a href="#">GSK Lesion Measurement Method</a>	(none)	(none)
Data Element From Eli Lilly:	<a href="#">EliLilly Lesion Measurement Lesion Method of Measurement</a>	(none)	(none)
Data Element From Mayo:	<a href="#">Mayo Lesion Measurement Method of Evaluation</a>	(none)	(none)
Data Element From BRIDG:	<a href="#">BRIDG PerformedObservation methodCode</a>	(none)	(none)

(Curation Status: [InProgress](#))

# Discussion

- It was difficult to find the set of terminological components that were needed for classification.
- Definitions require a model.
- The ability to map value meanings to common terminology increases the ability to discover overlap.
- None of the terminologies carried good value set definitions.
- Data types played a key role in classification.
- The BRIDG model was too coarse for harmonization.
- Units played an insignificant role in the harmonization process.

# Summary

- **Prototype was quite successful**
- **Semantic MediaWiki was demonstrated to be a viable platform for:**
  - **Prototyping**
  - **Modeling and data loading**
  - **Collaborative discussion**
  - **Analyzing and harmonizing metadata**
  - **Semantic annotation**
  - **Import and export**



# Next Step

- **Prototype Phase 2**
  - **More content**
  - **Workflow**
  - **Ajax integration w/ BioPortal**
  - **New paradigms for harmonization**
  - **Spreadsheet based tools**
  - **Usability analysis**
- **Production**
  - ...

# Links/Contact

- CSHARE Demo Wiki: (cshare/share)  
<http://informatics.mayo.edu/cshareDemo>
- LexWiki: <https://cabiq-kc.nci.nih.gov/Vocab/KC/index.php/LexWiki>
- CDISC: <http://www.cdisc.org/>
- Guoqian Jiang:  
[jiang.guoqian@mayo.edu](mailto:jiang.guoqian@mayo.edu)

# Questions ?

