

NCI Workshop

The Role of Ontology in Big Cancer Data

Session 3: Cancer big data and the Ontology of Disease

Bethesda, Maryland

May 13, 2015

Oncology in SNOMED CT



Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA



U.S. National Library of Medicine



Disclaimer

- ◆ “The views and opinions expressed do not necessarily state or reflect those of the U.S. Government, and they may not be used for advertising or product endorsement purposes”



SNOMED CT and oncology

- ◆ SNOP (~1965)
 - Systematized Nomenclature of Pathology
- ◆ SNOMED International (~1990)
 - Morphology and topology harmonized with the International Classification of Diseases for Oncology (ICD-O)
- ◆ SNOMED CT (~2000)
 - General clinical terminology



Breast cancer in SNOMED CT

The screenshot displays the IHTSDO SNOMED CT Browser interface. At the top, it shows the release information: "Release: United States Edition 20150301" and the perspective: "Perspective: Full". The main search area shows a search for "breast cancer" with 56 matches found in 0.152 seconds. The search results are filtered by language (English) and semantic tag (finding). The concept details for "Malignant tumor of breast (disorder)" (SCTID: 254837009) are shown on the right, including its parents (Malignant neoplasm of thorax (disorder) and Neoplasm of breast (disorder)) and children (16 concepts). A callout box highlights the finding site: "Breast structure" and associated morphology: "Malignant neoplasm of primary, secondary, or uncertain origin".

IHTSDO SNOMED CT Browser

© IHTSDO 2015

Search Mode: Partial matching search mode

Status: Active components only

Group by concept

Filter results by Language

english 56

Filter results by Semantic Tag

finding 32

situation 8

disorder 6

substance 4

Search: breast cancer

56 matches found in 0.152 seconds.

Concept	Description
Breast cancer	Malignant tumor of breast (disorder)
CA - Breast cancer	Malignant tumor of breast (disorder)
Suspected breast cancer	Suspected breast cancer (situation)
At risk of breast cancer	At risk of breast cancer (finding)
Familial cancer of breast	Familial cancer of breast (disorder)
Seen by breast cancer nurse	Seen by breast cancer nurse (finding)
Triple-negative breast cancer	Triple negative malignant neoplasm of breast (disorder)
At risk of breast cancer (finding)	At risk of breast cancer (finding)

Concept Details: Malignant tumor of breast (disorder)
SCTID: 254837009

Parents

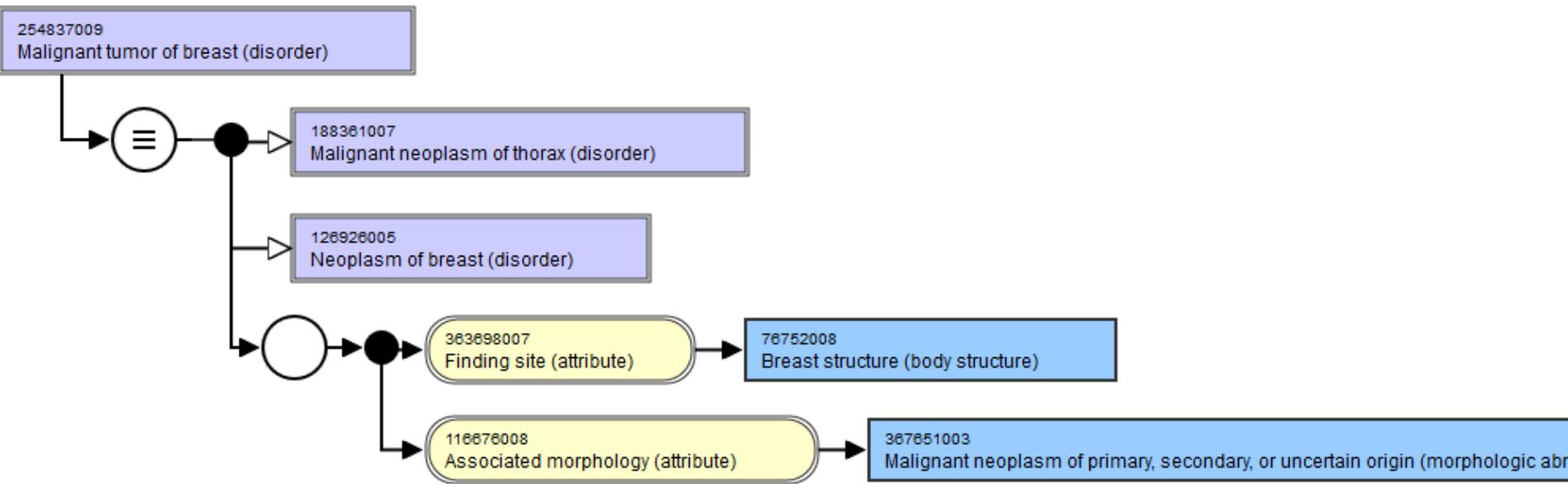
- Malignant neoplasm of thorax (disorder)
- Neoplasm of breast (disorder)

Children (16)

- Carcinoma of breast (disorder)
- Familial cancer of breast (disorder)
- Hormone receptor positive malignant neoplasm of breast (disorder)
- Local recurrence of malignant tumor of breast (disorder)
- Malignant lymphoma of breast (disorder)

Finding site → Breast structure
Associated morphology → Malignant neoplasm of primary, secondary, or uncertain origin

Breast cancer in SNOMED CT



Main types of cancer-related entities

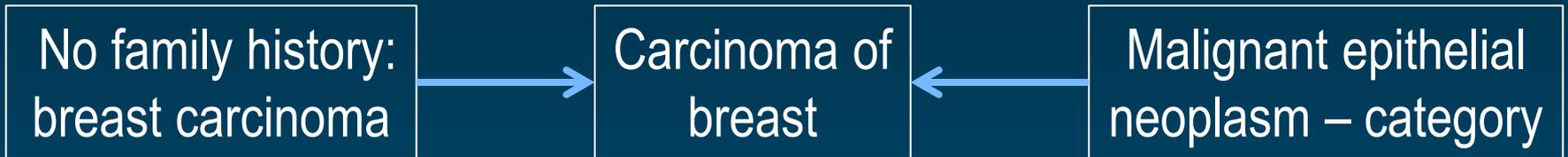
“situation”

disorders

morphologic
abnormality

*associated
finding*

*associated
morphology*



Filter results by Semantic Tag

Types of cancer-related entities

disorder

236

◆ Sarcoma lower inner quadrant of female breast

finding

212

◆ T3: Breast tumor >5 cm in greatest dimension

record artifact

150

◆ CAP Cancer Checklist; Breast: Excision Less Than...

situation

120

◆ Suspected breast cancer

procedure

66

◆ BRCA1 mutation carrier detection test

substance

30

◆ Breast cancer type 1 susceptibility protein

tumor staging

23

◆ Tumor-node-metastasis (TNM) breast tumor staging

observable entity

17

◆ Distance of tumor from closest surgical margin (breast)

assessment scale

14

◆ Mental adjustment to cancer scale

staging scale

10

◆ Nottingham histologic grading system

Coverage (pre-coordinated)

Category	Top-level concept	# descendants
Disorder	Malignant neoplastic disease	3449
Finding	Tumor finding	1562
Record artifact	Pathology report	71
Situation	- Family history of neoplasm	57
	- History of neoplasm	91

**1.75% of SNOMED CT concepts
are directly related to cancer**



Coverage (post-coordinated)

- ◆ Compositional grammar
 - Based on the SNOMED CT “concept model”
- ◆ “Refinement” of existing concepts
 - Specific morphology
 - Primary vs. secondary
 - Specific anatomical structure
 - Laterality
- ◆ E.g., Family history of breast carcinoma in the mother

Use cases

- ◆ Clinical documentation
- ◆ Clinical research
 - Secondary use of data / observational research
- ◆ Cohort selection
 - For clinical quality measures (“Meaningful Use”)
 - For clinical research
- ◆ Clinical decision support

eMeasure Title	Breast Cancer: Hormonal Therapy for Stage IC-IIIIC Estrogen Receptor/Progesterone Receptor (ER/PR) Positive Breast Cancer		
eMeasure Identifier (Measure Authoring Tool)	140	eMeasure Version number	4.0.000
NQF Number	0387	GUID	ac639794-cb3f-4f4f-8fa6-680300d5ed4e
Measurement Period	January 1, 20XX through December 31, 20XX		
Measure Steward	American Medical Association-convened Physician Consortium for Performance Improvement(R) (AMA-PCPI)		
Measure Developer	American Medical Association-convened Physician Consortium for Performance Improvement(R) (AMA-PCPI)		
Endorsed By	National Quality Forum		
Description	Percentage of female patients aged 18 years and older with Stage IC through IIIIC, ER or PR positive breast cancer who were prescribed tamoxifen or aromatase inhibitor (AI) during the 12-month reporting period		

<https://vsac.nlm.nih.gov/>

Metadata

Member Of

Name:

Breast Cancer ER or PR Positive

Type:

Extensional

Steward:

American Medical Association-convened Physician Consortium for Performance Improvement(R)

OID:

2.16.840.1.113883.3.526.2.1387

Definition ID:

20150430

Program:

CMS, MU2 Update 2015-05-01 using this value set

Value Set Members Expansion ID: 20121025

Expanded Code List

View Toggle Clear

Page 1 of 1 20

View 1 - 4 of 4

Code	Descriptor	Code System	Version	Code System OID
416053008	Estrogen receptor positive tumor (disorder)	SNOMEDCT	2014-09	2.16.840.1.113883.6.96
416561008	Progesterone receptor positive tumor (disorder)	SNOMEDCT	2014-09	2.16.840.1.113883.6.96
417181009	Hormone receptor positive malignant neoplasm of breast (disorder)	SNOMEDCT	2014-09	2.16.840.1.113883.6.96
417742002	Hormone receptor positive tumor (disorder)	SNOMEDCT	2014-09	2.16.840.1.113883.6.96

View

Page 1 of 1 20

View 1 - 4 of 4

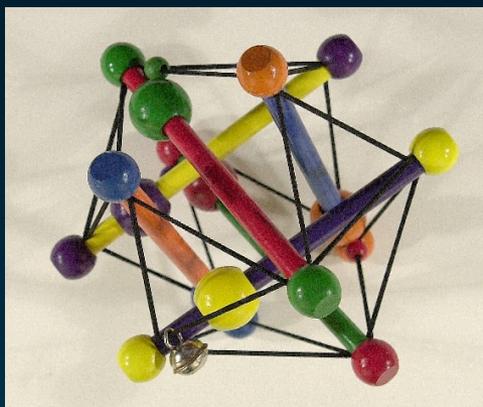
SNOMED CT in the U.S.

- ◆ International release developed by IHTSDO
 - 27 member countries, including U.S.
- ◆ NLM is the U.S. National Release Center
 - Distribute international release (UMLS license)
 - Released in January and July
 - Develop the U.S. extension / U.S. edition
 - Released in March and September
 - Take requests for changes or new concepts / terms
 - U.S. SNOMED CT Content Request System (USCRS)
 - <https://uscrs.nlm.nih.gov/>
 - E.g., cancer staging concepts added for clinical quality measures



Challenges

- ◆ Keep it simple (or hide complexity)
 - Reproducible representation (ontology developer)
 - Reproducible annotation (ontology user)
- ◆ Interoperability between biological and clinical data
 - Clinical data is Big Data
 - Necessary for translational research



Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov



Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA