

Grounding a Hyper-Ontology on mCODE Conceptual Model and Foundational Ontologies for Semantic Interoperability in Oncology

Mirna El Ghosh¹, Christel Daniel¹, Catherine Duclos², Varvara Kalokyri³, Jean Charlet¹,
Melanie Sambres¹, Gianna Tsakou⁴, Manolis Tsiknakis³ and Ferdinand Dhombres¹

¹Sorbonne Université, Inserm, Université Sorbonne Paris-Nord, LIMICS, Paris, France

²Université Sorbonne Paris-Nord, Inserm, Sorbonne Université, LIMICS, Paris, France

³Institute of Computer Science, Foundation of Research and Technology Hellas, Heraklion, Greece

⁴MAGGIOLI S.P.A., Research and Development Lab, Marousi, Greece

FOAM@FOIS 2024

July 15, 2024

Enschede, Netherlands



Co-funded by
the European Union



EUCAIM
CANCER IMAGE EUROPE

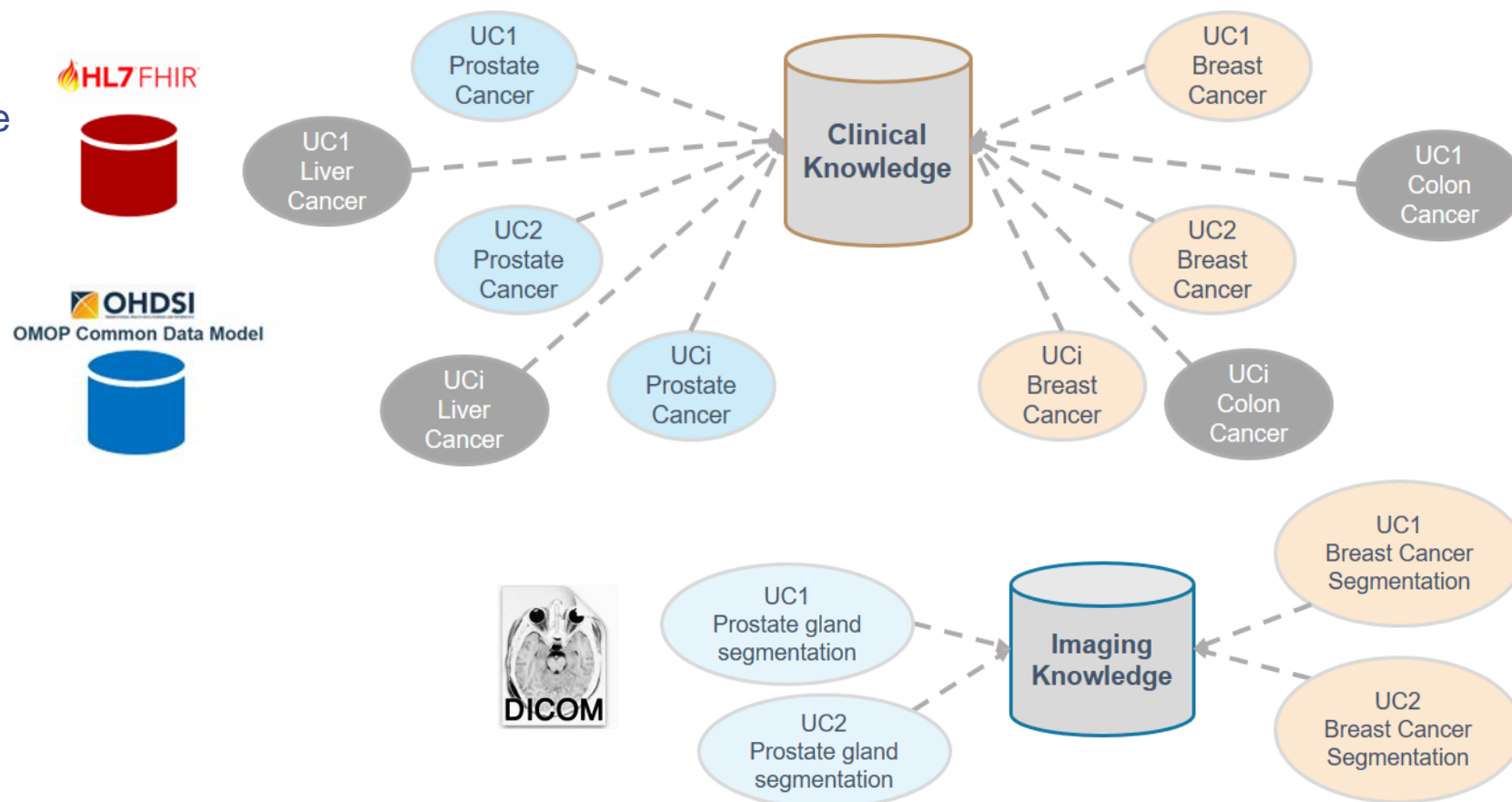
- EC-Co-Funded project
- Consists of 76 partners from 14 countries
- Data repositories: “AI for Health Imaging” Network (AI4HI), a cluster of 5 large EU-funded projects on big data and AI in cancer imaging
- European Research Infrastructures (e.g., Euro-BioImaging, BBMRI, EATRIS and ELIXIR)

An **ontology integration approach** toward **semantic interoperability** among **heterogeneous cancer image data models** and **distributed big cancer data repositories**



Use Cases (Input)

- 16 use cases (prostate, breast, colon, liver, lung, colorectal)
- Standard concepts (OMOP/FHIR)
- Imaging: DICOM SEG
- Heterogeneity/disparity of knowledge



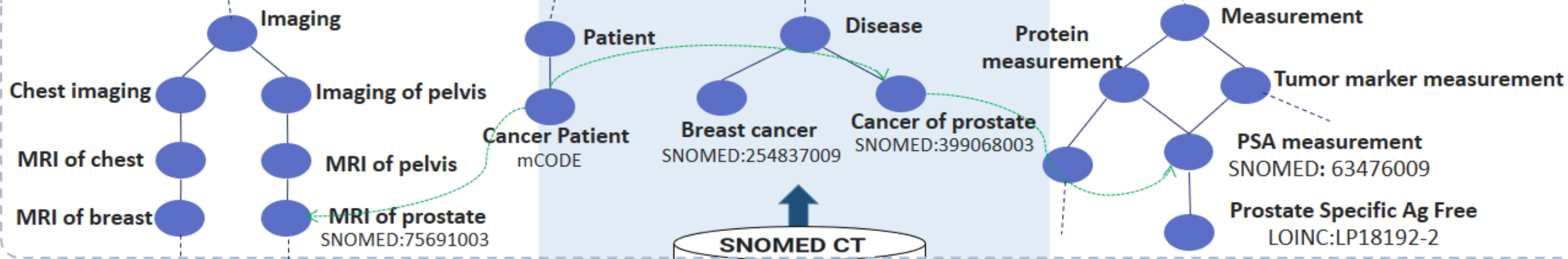
Ontology Integration Approach

Semantic Search

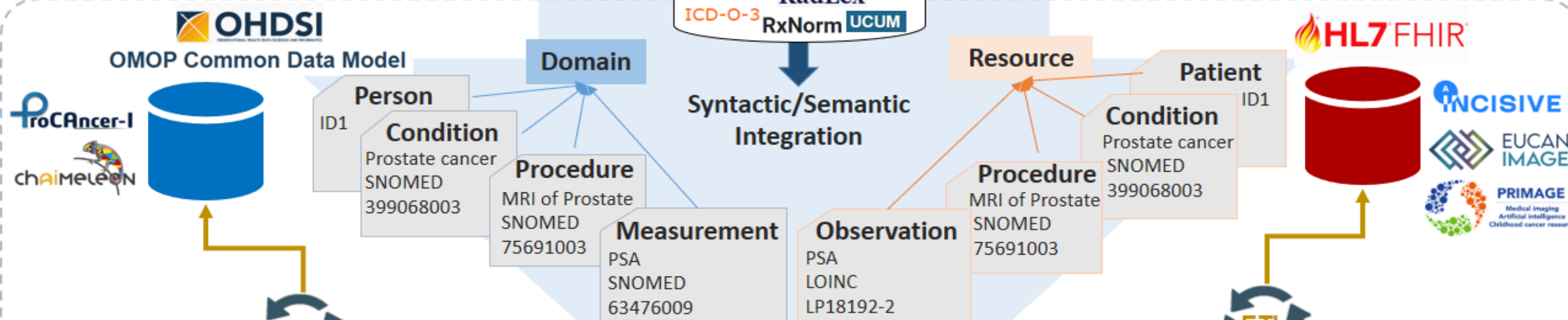
Federated Queries



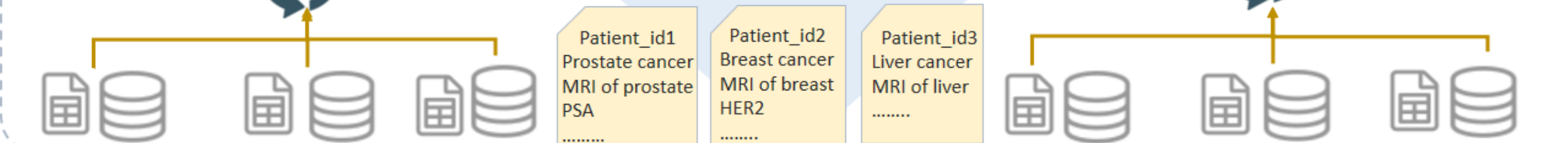
Semantic Common Meta-Model



Heterogeneous Data Standards

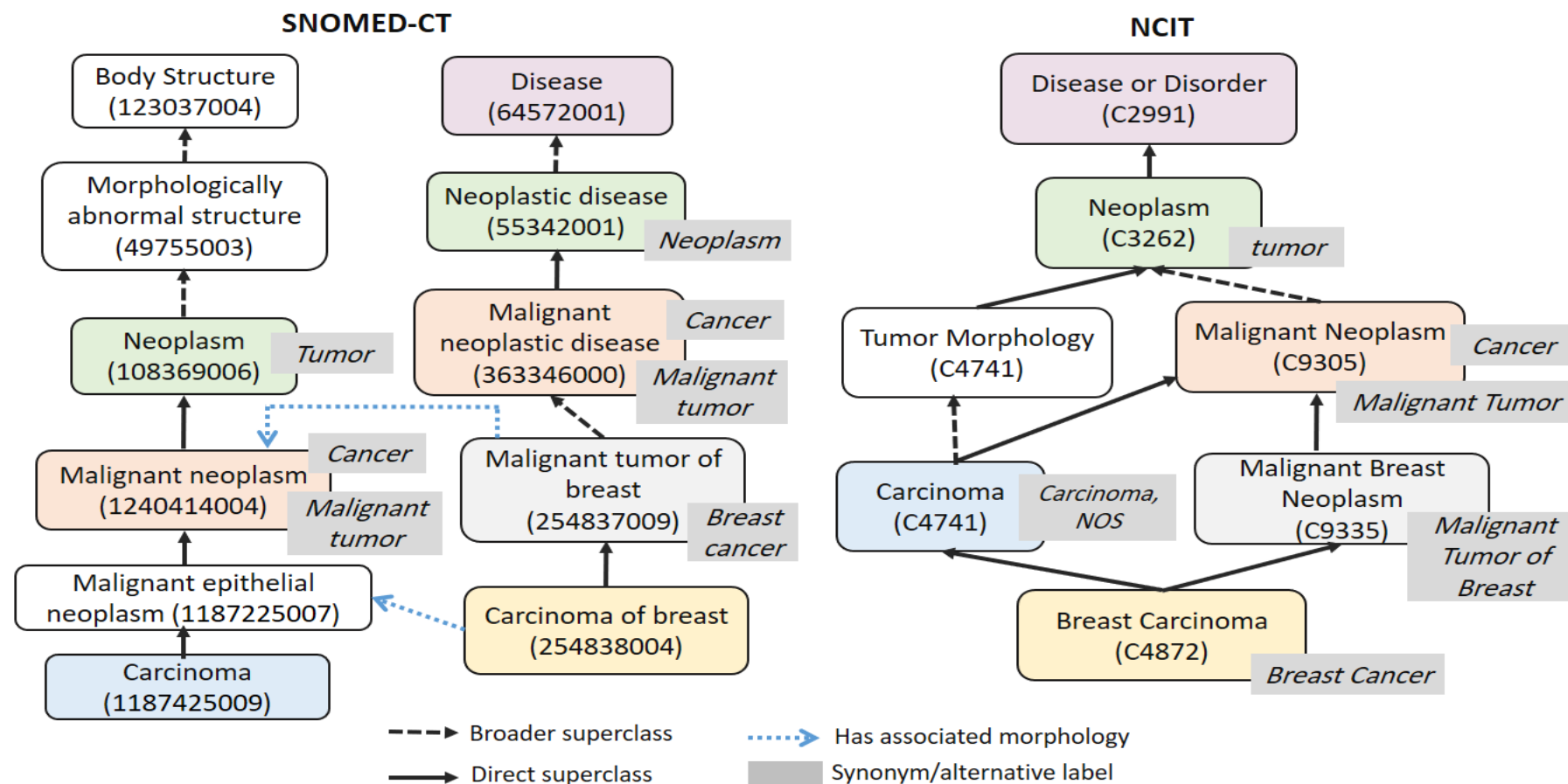


Local Sources



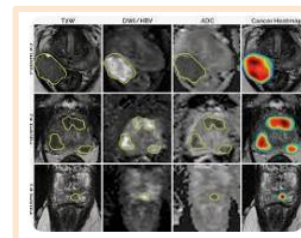
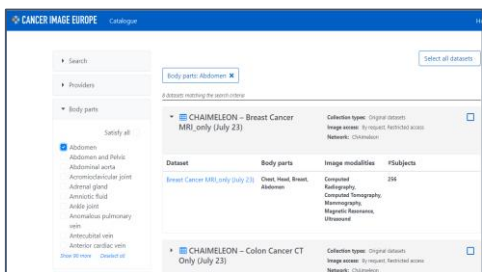
Semantic Heterogeneity

- Conflicting or heterogeneous semantic representation of basic oncology concepts (Disease/Morphology)
- Common well-founded semantic representation



EUCAIM Hyper-Ontology

- **Definition:** a FAIR-compliant semantic meta-model that abstracts and integrates heterogeneous and complex cancer imaging and clinical/biological knowledge represented using various standards (OMOP, FHIR, DICOM)
- **Purpose:** semantic search/federated querying/exploring data collections/image annotation
- **Requirements:** 1) *quality* requirements stated as Competency Questions (CQs) in the ORSD (Ontology Requirements and Specification Document); 2) *specific* requirements defined with the help of experts considering the challenges of the syntactic/semantic heterogeneity/disparity of the provided knowledge.



FR1: To define the basic cancer types.

CQ1: What are the leading cancer types?

Prostate cancer, Colon cancer, Breast cancer, Rectal cancer, Lung cancer, Neuroblastoma, Diffuse intrinsic pontine glioma, Colorectal Cancer, Primary malignant neoplasm of liver, Malignant neoplasm of colon and/or rectum, Primary malignant neoplasm of breast.

FR2: To define the specific cancer types.

CQ1: Are there any specific types of breast cancer?

Primary malignant neoplasm of female breast (SNOMEDCT, 363346000), Primary malignant neoplasm of breast with axillary lymph node invasion (disorder) (SNOMEDCT, 1082901000112103)

CQ2: Are there any specific types of prostate cancer?

Benign prostatic hyperplasia (SNOMEDCT, 266569009), Hormone refractory prostate cancer (SNOMEDCT, 427492003), Hormone sensitive prostate cancer (SNOMEDCT, 722103009)

CQ3: Are there any specific types of liver cancer?

Liver cell carcinoma (disorder) (SNOMEDCT, 109841003), Secondary malignant neoplasm of liver (SNOMEDCT, 94381002)

FR3: To define the main tumor staging methods and values.

CQ1: What tumor staging methods are specified for breast cancer?

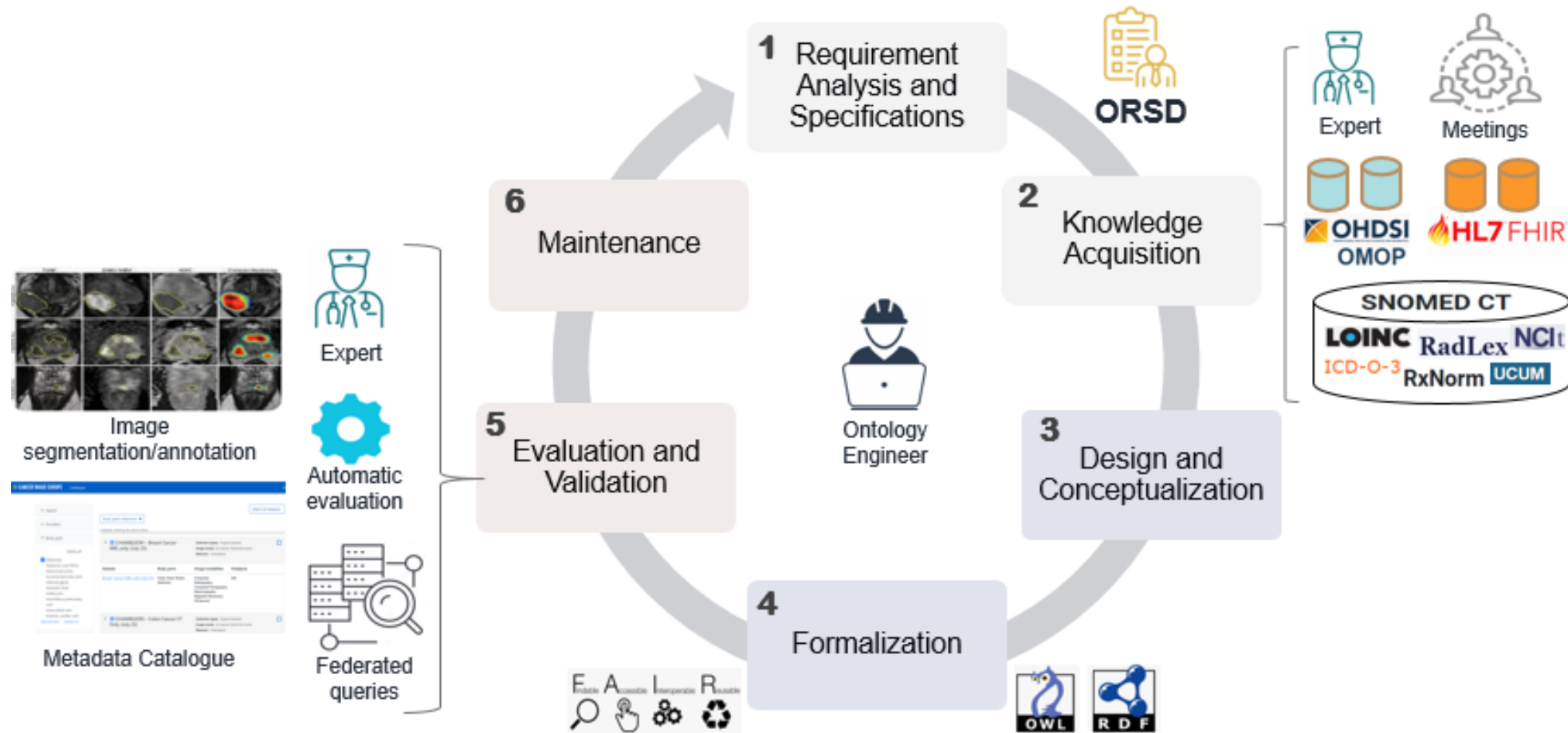
Edition of American Joint Commission on Cancer, Cancer Staging Manual used for TNM staging (observable entity) (SNOMEDCT, 443941007)

CQ2: What tumor staging (categorical) values are specified for breast cancer?

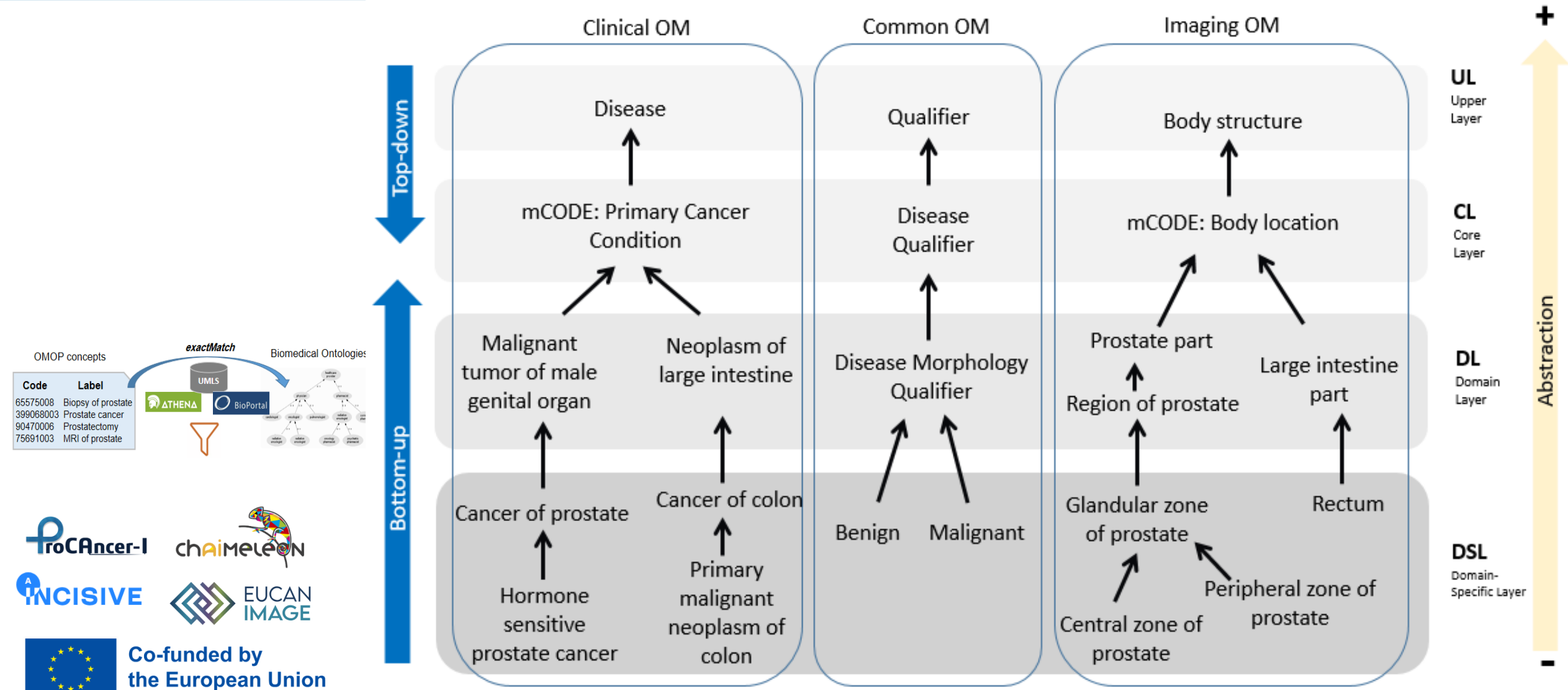
American Joint Committee on Cancer clinical T category allowable value (qualifier value) (SNOMEDCT, 1222585009), American Joint Committee on Cancer clinical N category allowable value (qualifier value) (SNOMEDCT, 1222588006), Tumor histopathological grade status values (tumor staging) (SNOMEDCT, 258244004)

FR4: To define the histology types of cancers.

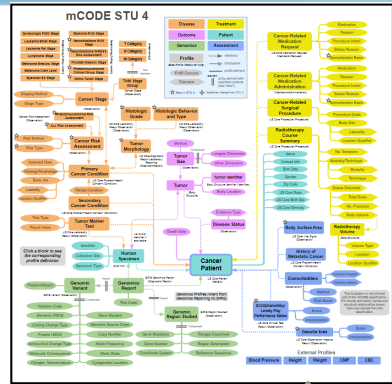
Hybrid Iterative Well-Founded Approach



Design and Conceptualization: Modular and Layered Architecture



Design and Conceptualization: Modular and Layered Architecture



Clinical Expert

Top-down

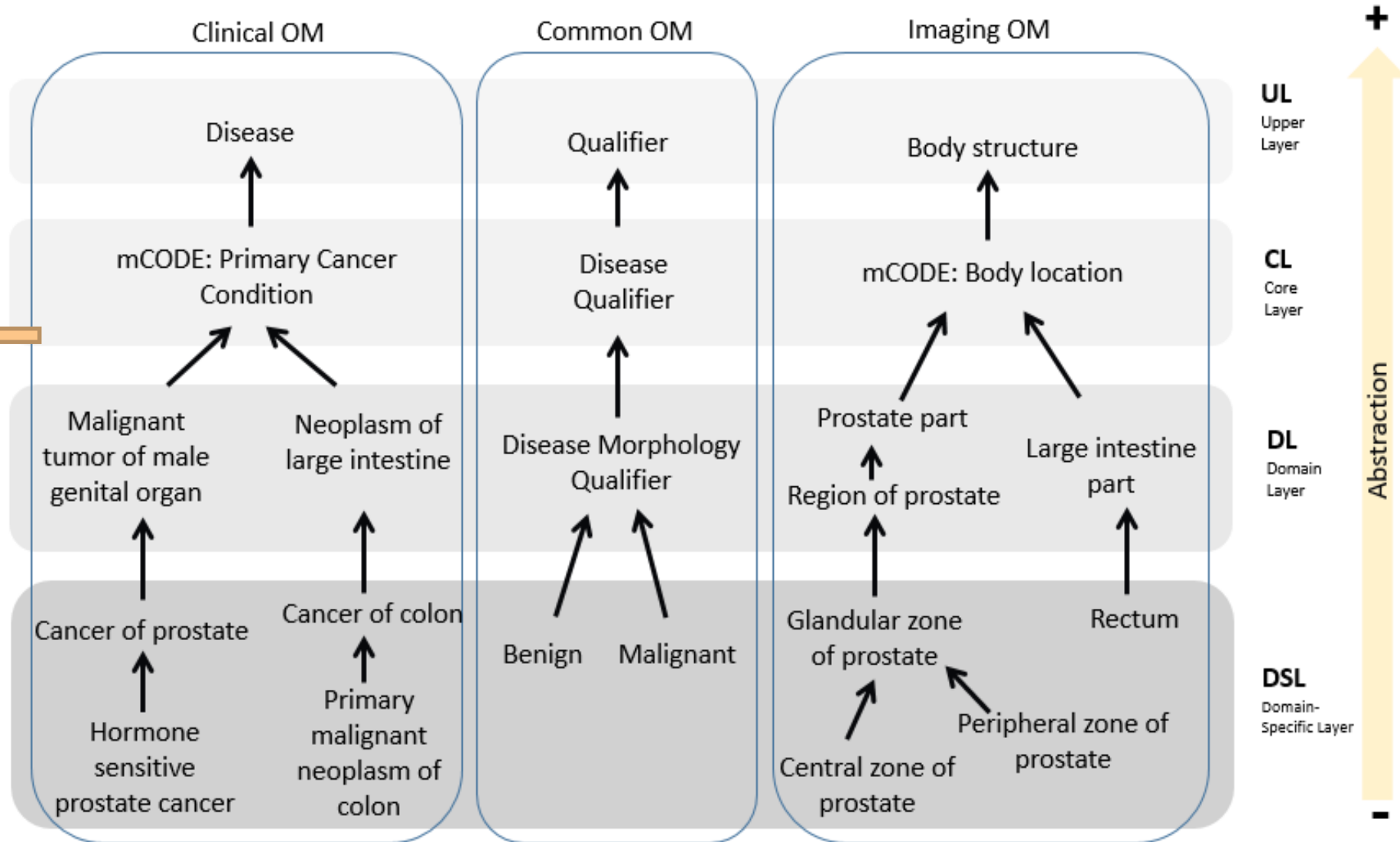
Bottom-up

OMOP concepts

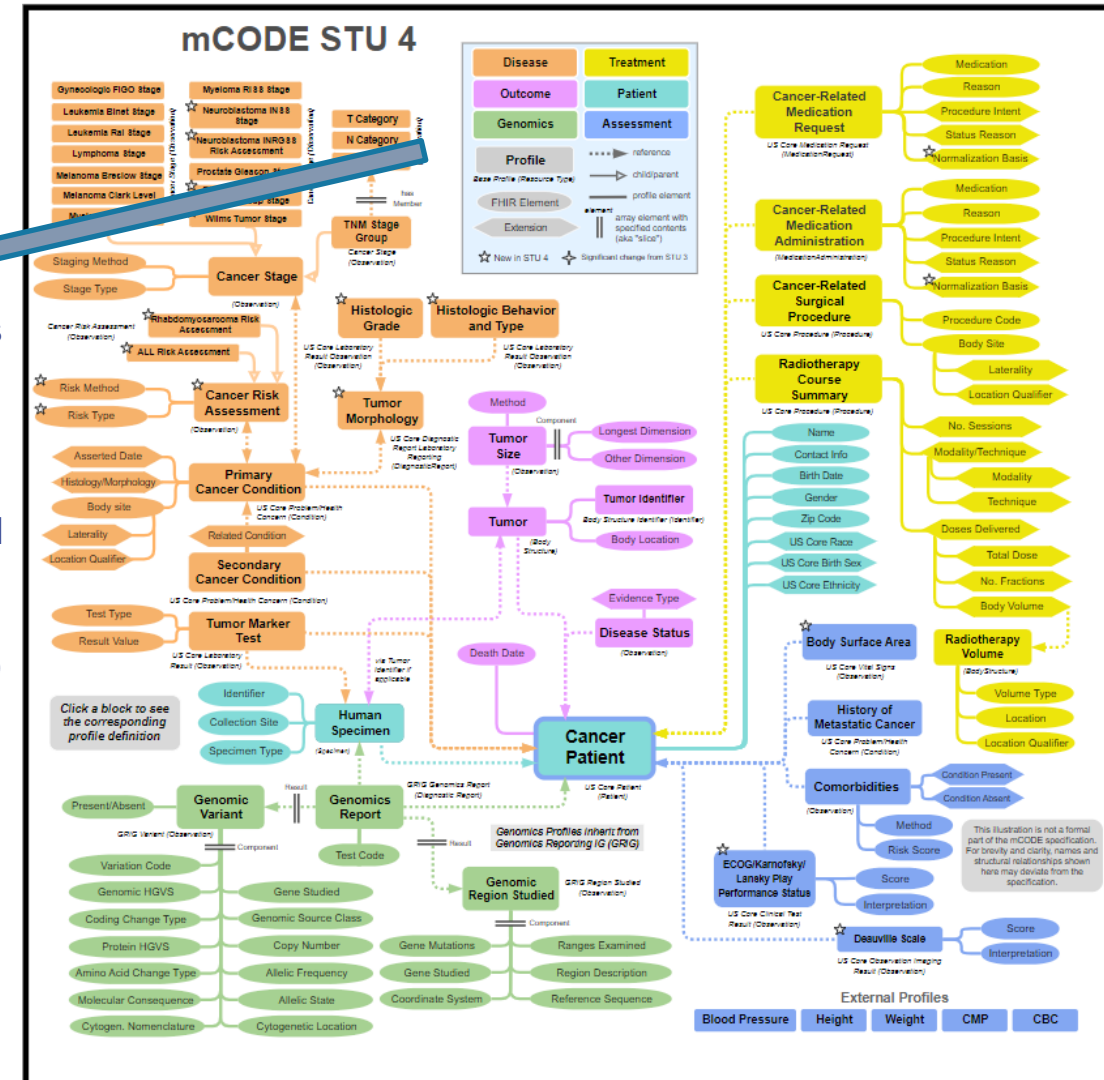
exactMatch

Biomedical Ontologies

Code	Label
65575008	Biopsy of prostate
399068003	Prostate cancer
90470006	Prostatectomy
75691003	MRI of prostate



- **Domain of Knowledge implementation guide** intended to show how to represent clinical concepts in oncology
- **Thematic groups:**
 1. **Disease Characterization:** cancer staging, and tumor characteristics (histological classification, grade, morphology, and behavior of tumor cell);
 2. **Health Assessment:** patient's general health before and during treatment, such as comorbidities, performance assessment, laboratory tests, and history of metastatic cancer;
 3. **Cancer Treatments:** reporting of procedures (surgery and radiotherapy) and medications used to treat a cancer patient or relevant to that treatment;
 4. **Outcomes:** disease status, tumor, and tumor size.



3 Disease Characterization

The mCODE Disease Characterization group includes data elements specific to the diagnosis and staging of cancer. This includes:

- **Cancer Diagnosis** - the date and location (body site/position and laterality) of the cancer diagnosis.
- **Tumor Characteristics** - histological classification, grade, morphology, and behavior of the tumor cell, compared to that of a normal cell.
- **CancerStage** - describes the severity of an individual's cancer based on the magnitude of the original (primary) tumor as well as on the extent cancer has spread in the body. Understanding the stage of the cancer helps doctors to develop a prognosis and design a treatment plan for individual patients. Staging calculations leverage results from the previous two categories, along with prognostic factors relevant to the cancer type, in order to assess an overall cancer stage group (source: [AJCC](#)).

3.1 Representing the Cancer Diagnosis

The cancer diagnosis combines the type, site, and certain characteristics of the cancer. Depending on the EHR and provider organization, different code systems may be used, such as:

- Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)
- International Classification of Diseases, 10th version, Clinical Modifications (ICD-10-CM)
- International Classification of Diseases for Oncology, 3rd version (ICD-O-3)

Because the use of these code systems vary in different institutions, mCODE supports all three. Two elements and one extension of the [FHIR Condition Resource](#) are involved with coding the cancer diagnosis: `Condition.code`, `Condition.bodySite`, and the `HistologyMorphologyBehavior` extension. How these attributes are used, depending on the code system, is captured in the table below:

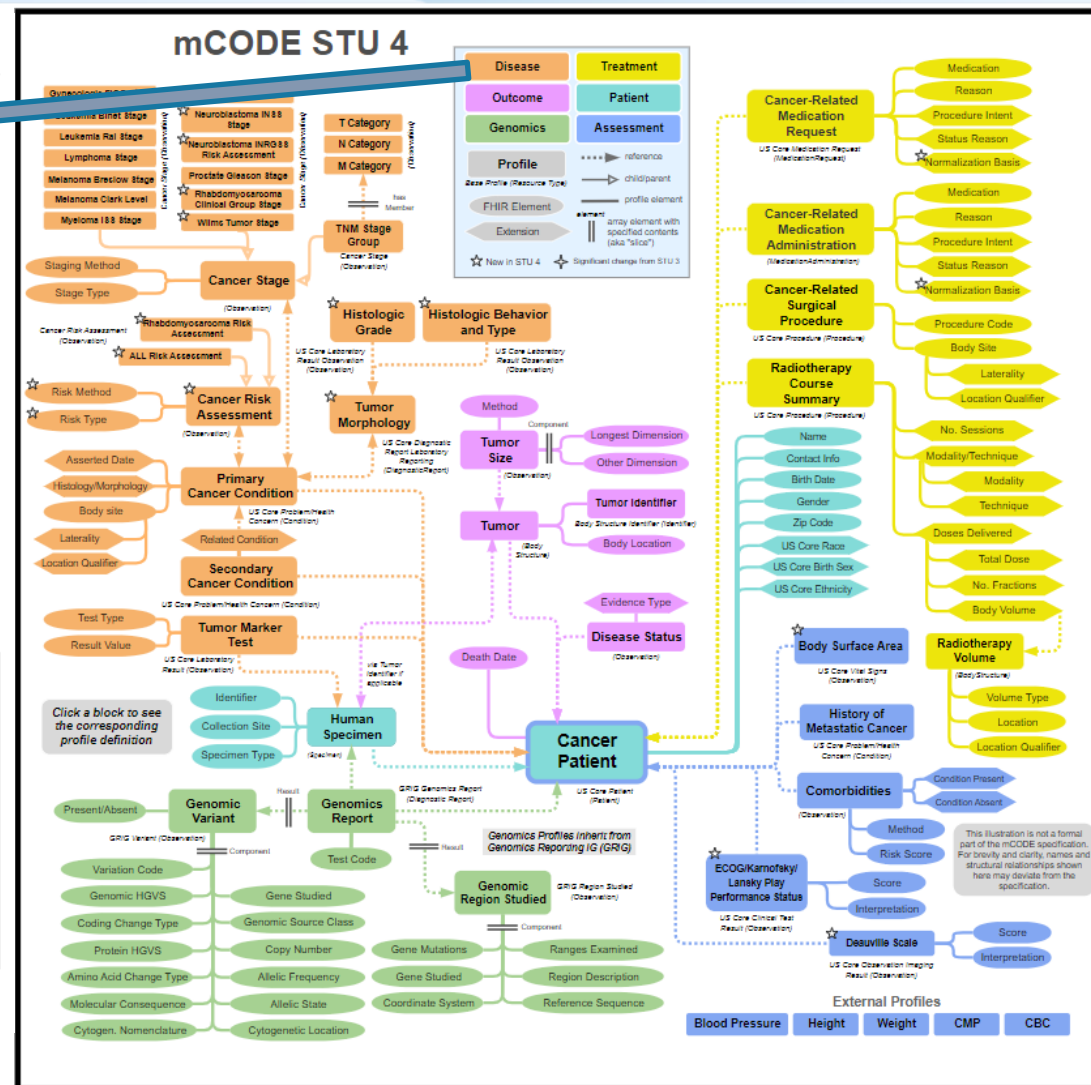
Encoding	Code	Histology Morphology Behavior Extension	Body Site
SNOMED Encoded	Any SNOMED CT code in the Primary Cancer Disorder Value Set	Any SNOMED CT code in the Histology Morphology Behavior Value Set	Any descendant of 123037004 "Body structure"
ICD-10-CM Encoded	Any ICD-10-CM primary code (precoordinated)	omit	optional; must be consistent with primary code if provided but may contain more detail
ICD-O-3 Encoded	SNOMED CT code 55342001 "Neoplastic disease (disorder)", 363346000 "Malignant neoplastic disease (disorder)", or 20376005 "Benign neoplastic disease (disorder)"	Full ICD-O-3 morphology code*	Any ICD-O-3 Topology Code

*The ICD-O-3 morphology code should have the form HHHH/B G where HHHH is the 4-digit histology code, B is the 1-digit behavior code (i.e., /1, /2, or /3 suffix for primary cancers, and /6 suffix for secondary cancers), and G is the grade or equivalent in leukemias and lymphomas (1 digit). For more information, see [ICD-O Third](#)



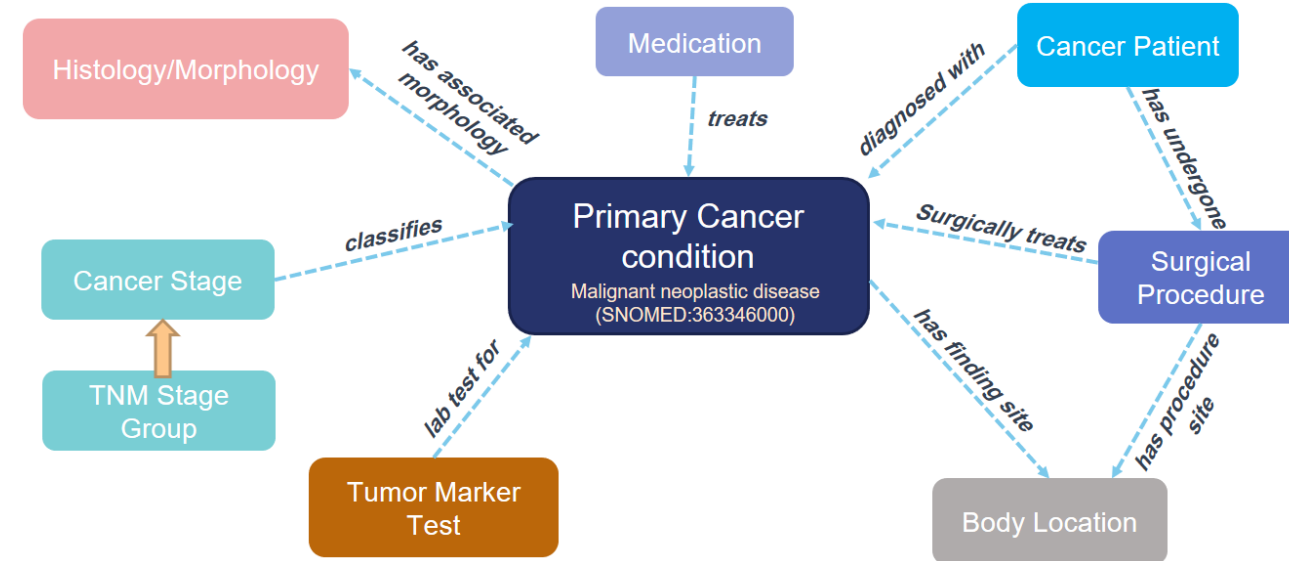
Co-funded by
the European Union

[2] <https://build.fhir.org/ig/HL7/fhir-mCODE-ig/>



mCODE-based Ontological Model

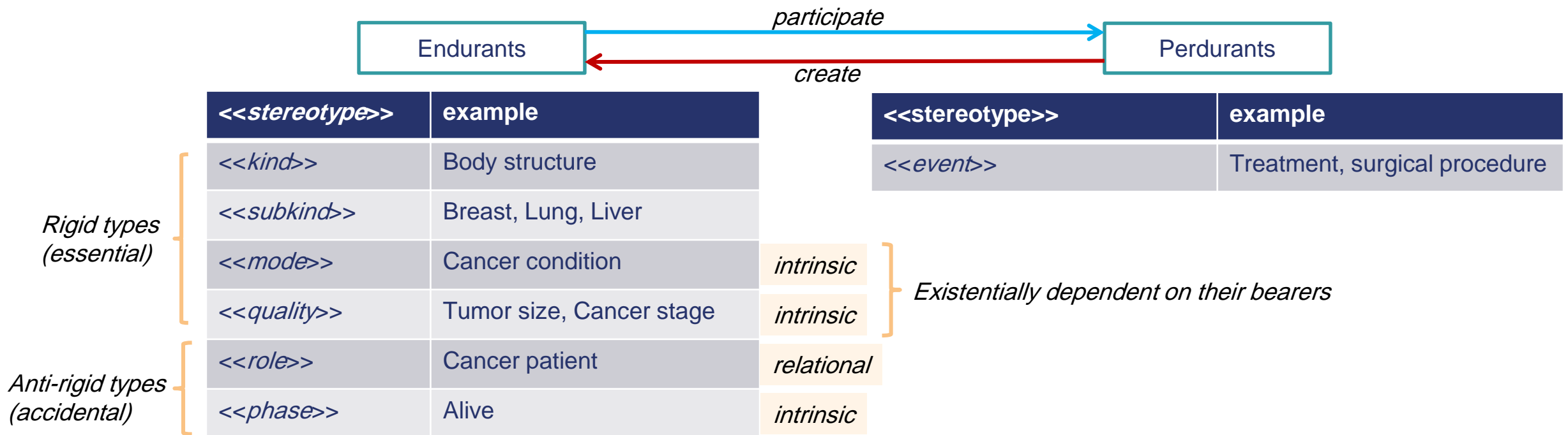
- **Ontology unpacking process** [3]: Ontological analysis to reveal the ontological conceptual model of mCODE, aiming to achieve semantic clarity on oncology's main concepts.
- **Ontology-oriented grounding approach:** foundational ontologies (UFO [4])
- **Ontology-driven conceptual modeling language:** OntoUML [4]



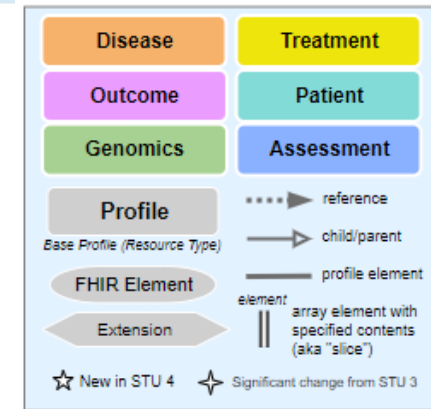
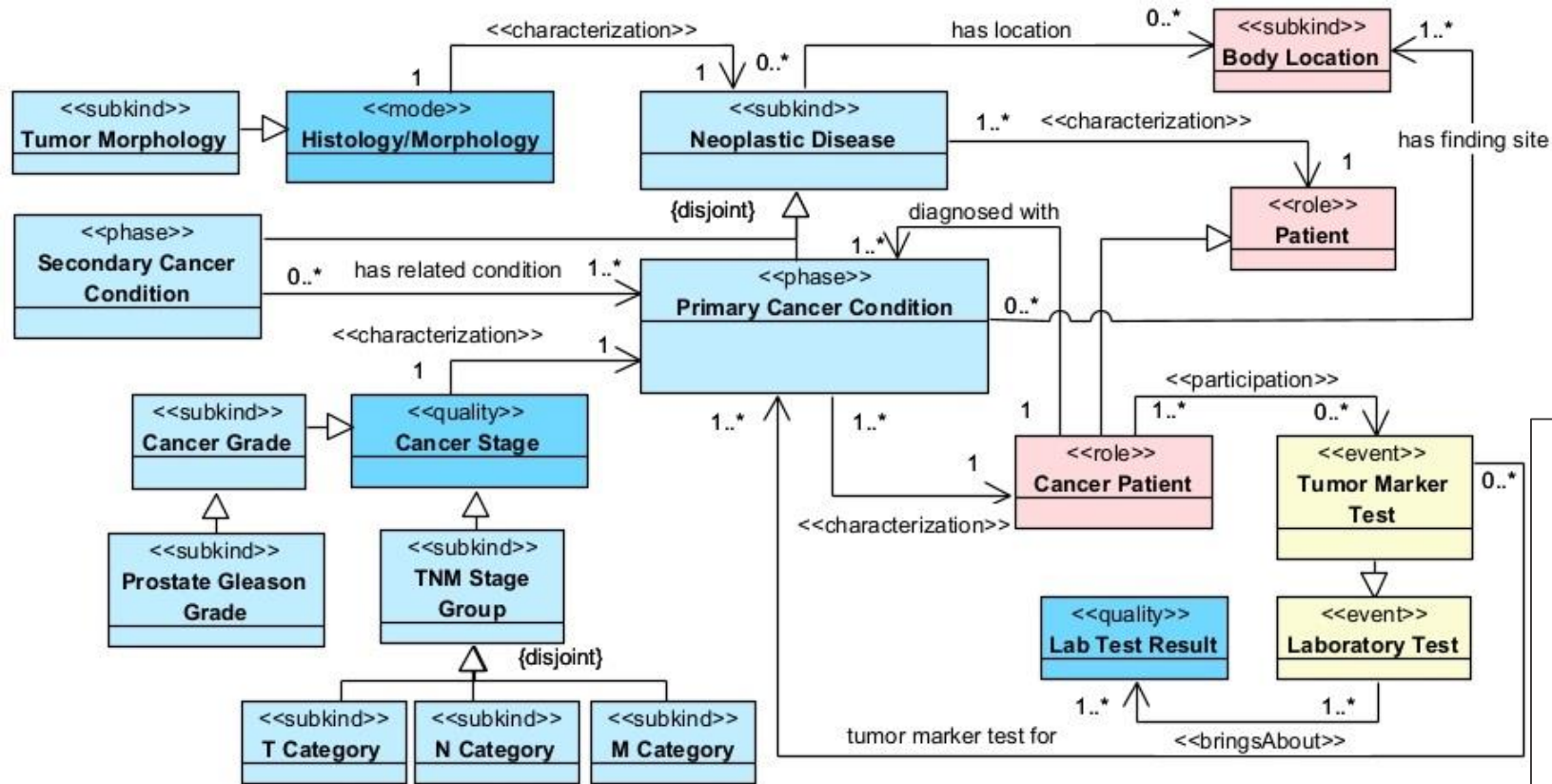
- *“Language truly ontological by nature”* [5]: explicitly represent the ontological nature of the domain
- *Ontology-Driven Conceptual Modeling (ODCM) language*: meta-model complies with UFO (modeling primitives reflect the ontological distinctions and axiomatization of UFO)



- “*Language truly ontological by nature*”[5]: explicitly represent the ontological nature of the domain
- *Ontology-Driven Conceptual Modeling (ODCM) language*: meta-model complies with UFO (modeling primitives reflect the ontological distinctions and axiomatization of UFO)



Ontology Unpacking Process



3 Disease Characterization

The mCODE **Disease Characterization** group includes data elements specific to the diagnosis and staging of cancer. This includes:

- **Cancer Diagnosis** - the date and location (body site/position and laterality) of the cancer diagnosis.
- **Tumor Characteristics** - histological classification, grade, morphology, and behavior of the tumor cell, compared to that of a normal cell.
- **CancerStage** - describes the severity of an individual's cancer based on the magnitude of the original (primary) tumor as well as on the extent cancer has spread in the body. Understanding the stage of the cancer helps doctors to develop a prognosis and design a treatment plan for individual patients. Staging calculations leverage results from the previous two categories, along with prognostic factors relevant to the cancer type, in order to assess an overall cancer stage group (source: AJCC ¹).

3.1 Representing the Cancer Diagnosis

The cancer diagnosis combines the type, site, and certain characteristics of the cancer. Depending on the EHR and provider organization, different code systems may be used, such as:

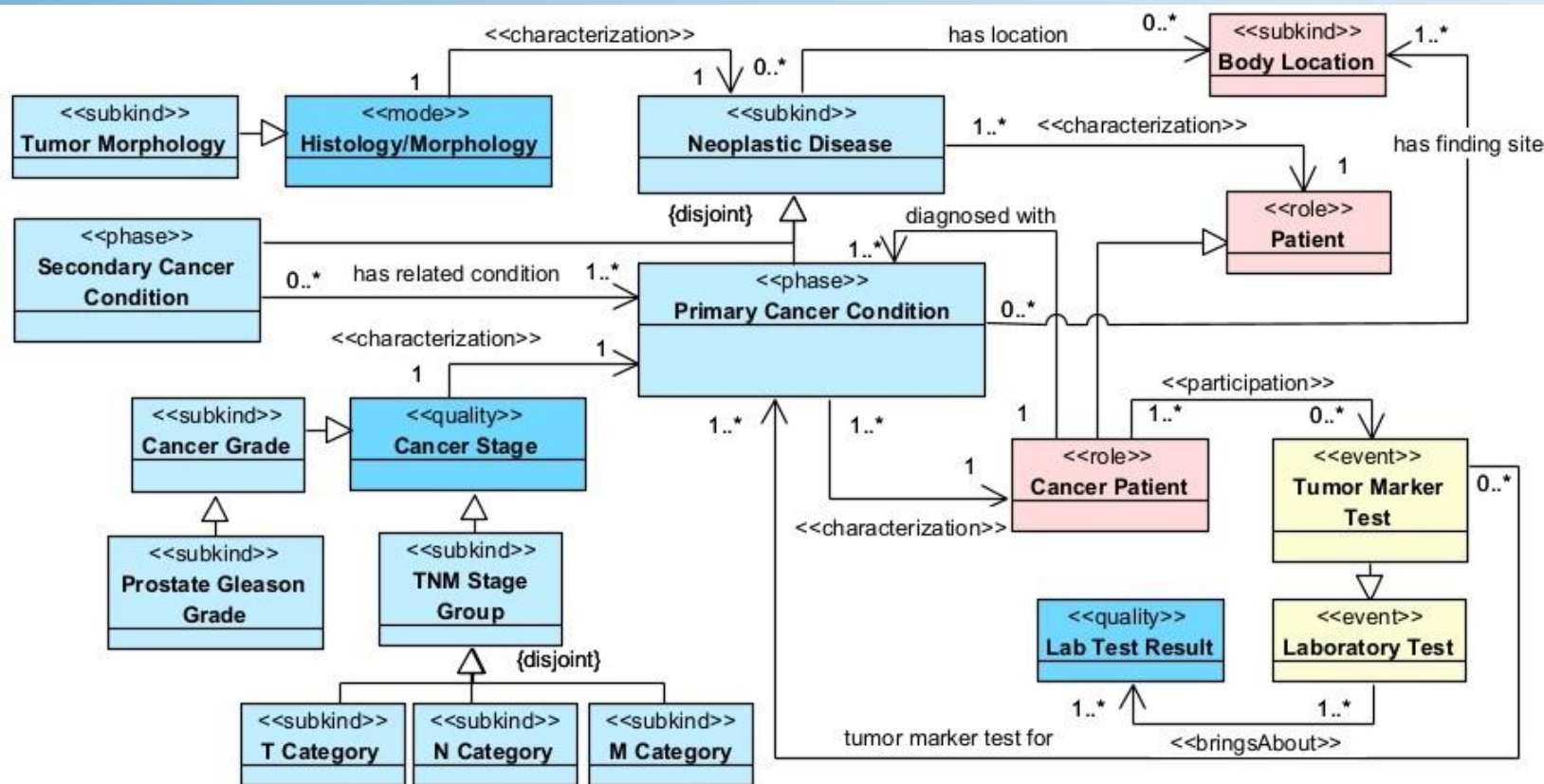
- Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) ²
- International Classification of Diseases, 10th version, Clinical Modifications (ICD-10-CM) ³
- International Classification of Diseases for Oncology, 3rd version (ICD-O-3) ⁴

Because the use of these code systems vary in different institutions, mCODE supports all three. Two elements and one extension ⁵ of the FHIR *Condition* Resource ⁶ are involved with coding the cancer diagnosis: *Condition.code*, *Condition.bodySite*, and the *HistologyMorphologyBehavior* extension. How these attributes are used, depending on the code system, is captured in the table below:

Encoding	Code	Histology Morphology Behavior Extension	Body Site
SNOMED Encoded	Any SNOMED CT code in the <i>Primary Cancer Disorder Value Set</i>	Any SNOMED CT code in the <i>Histology Morphology Behavior Value Set</i>	Any descendant of 123037004 "Body structure"
ICD-10-CM Encoded	Any ICD-10-CM primary code (precoordinated)	omit	optional; must be consistent with primary code if provided but may contain more detail
ICD-O-3 Encoded	SNOMED CT code 55342001 "Neoplastic disease (disorder)", 363346000 "Malignant neoplastic disease (disorder)", or 20376005 "Benign neoplastic disease (disorder)"	Full ICD-O-3 morphology code*	Any ICD-O-3 Topology Code

*The ICD-O-3 morphology code should have the form HHHH/B/G where HHHH is the 4-digit histology code, B is the 1-digit behavior code (i.e., /1, /2, or /3 suffix for primary cancers, and /6 suffix for secondary cancers), and G is the grade or equivalent in leukemias and lymphomas (1 digit). For more information, see ICD-O Third

Core Layer: Disease Characterization



CQ1) How are cancer conditions classified?

CQ2) Who has been diagnosed with primary cancer?

CQ3) In what body location is the cancer condition identified?

CQ4) Is there any histology/morphology associated with the cancer condition?

CQ5) How is the cancer stage/grade specified?

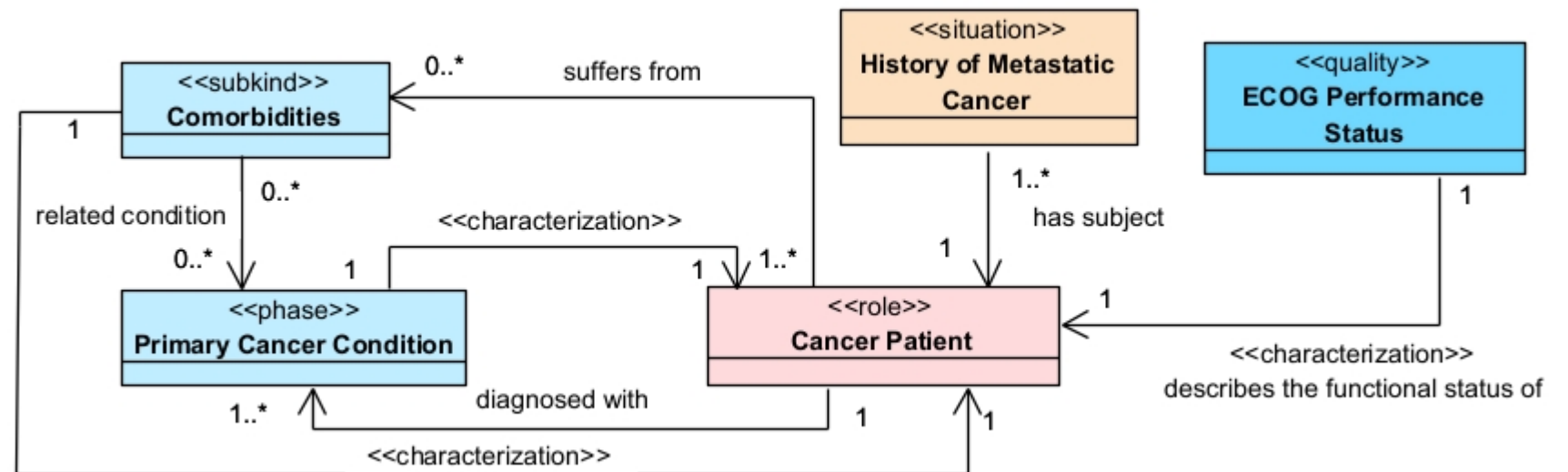
CQ6) Is the cancer condition classified by a certain stage/grade?

CQ7) Are there any tumor marker tests for the cancer condition?

CQ8) Are there any results generated by lab tests?

Core Layer: Health Assessment

- CQ1) Are there any comorbidities the cancer patient suffers from?
- CQ2) Are there any related cancer conditions for these comorbidities?
- CQ3) How is the cancer patient's functional status assessed?
- CQ4) Is the history of metastatic cancer recorded?



Core Layer: Cancer Treatment

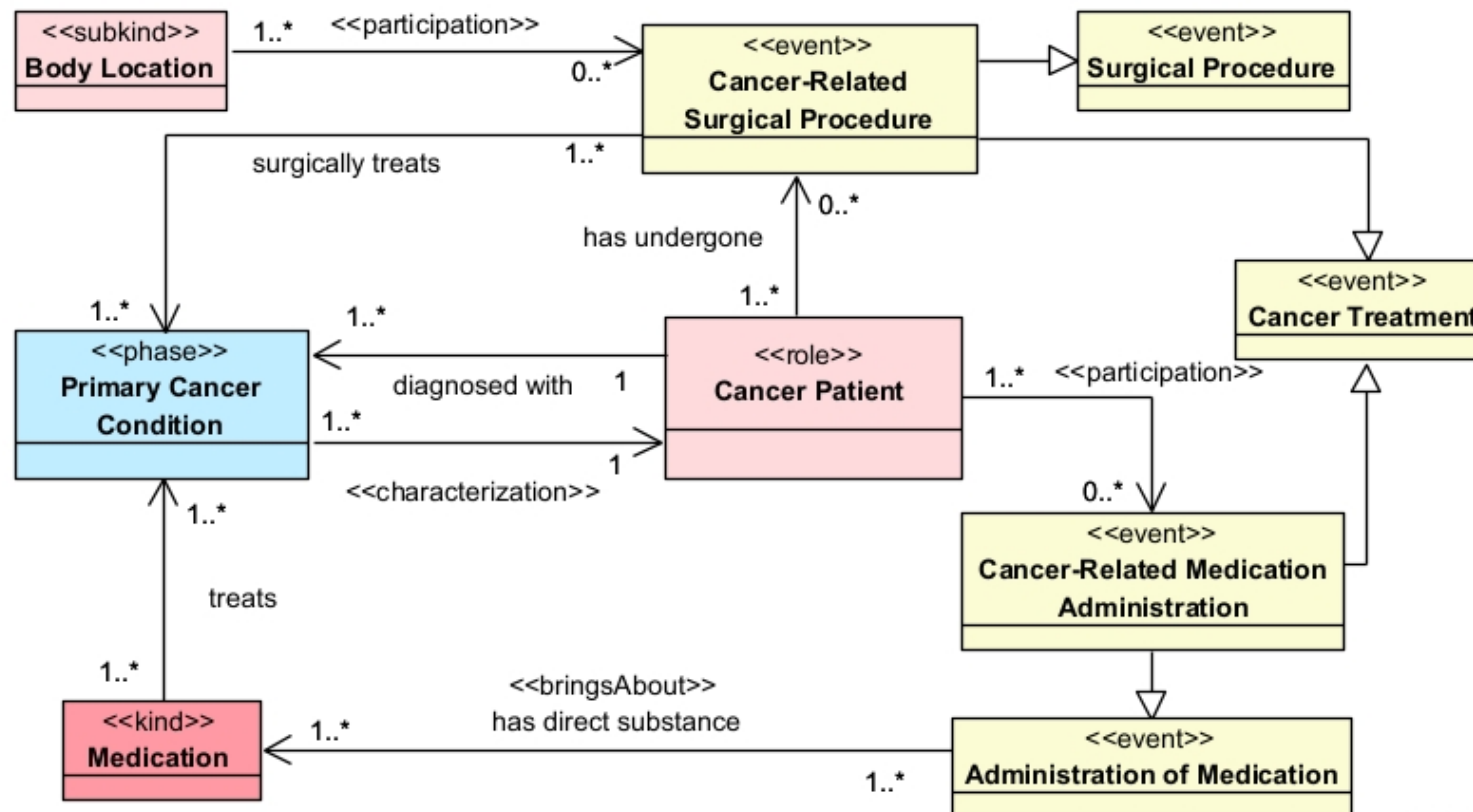
CQ1) Has the cancer patient undergone any surgical procedure?

CQ2) What cancer condition does the surgical procedure treat?

CQ3) What body location is affected by the surgical procedure?

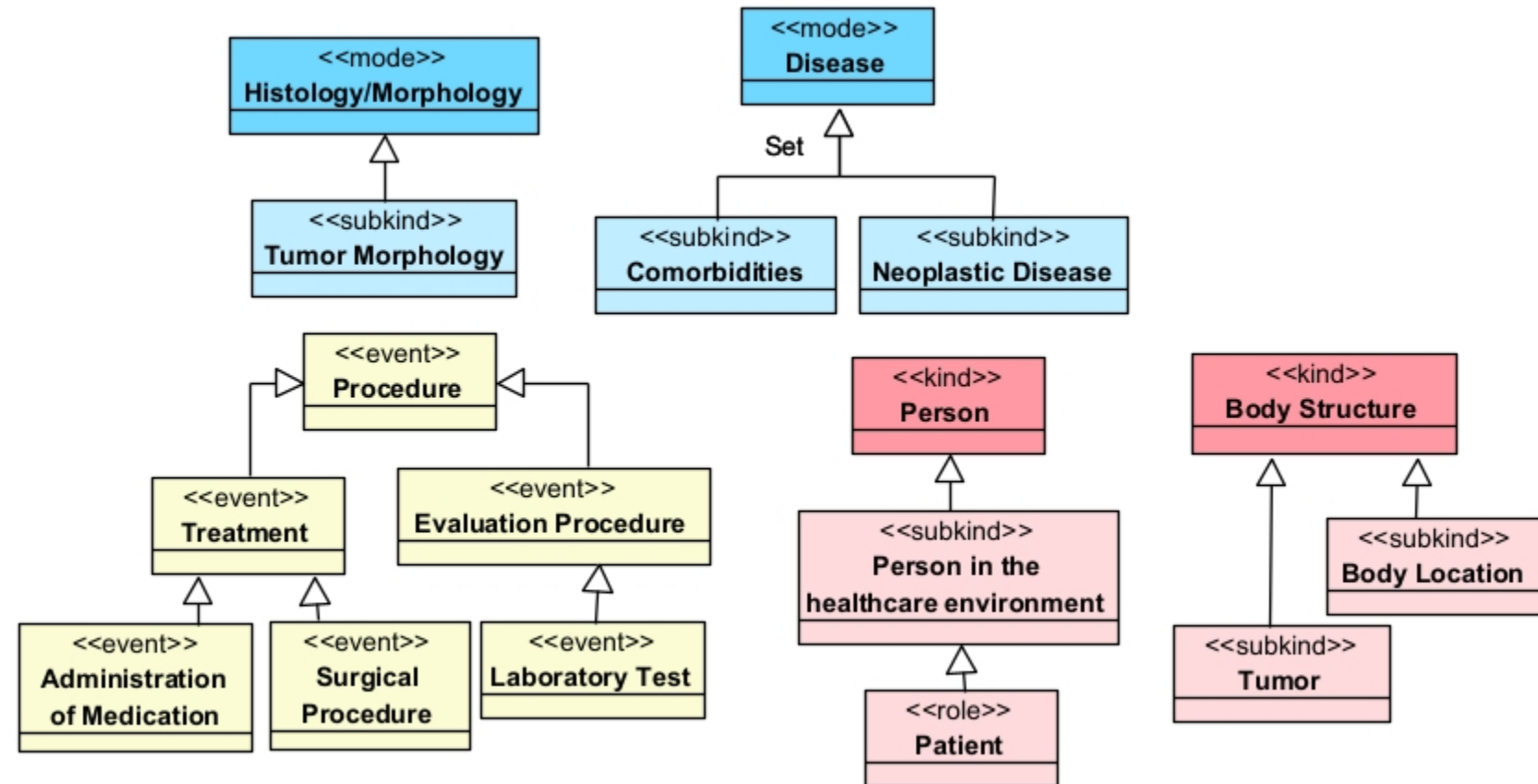
CQ4) Is any medication prescribed to treat the cancer condition?

CQ5) How is the medication prescribed?



Upper Layer

- Generic categories: Disease, Morphology, Treatment, Procedure, Staging/Grading, Body Structure, etc.

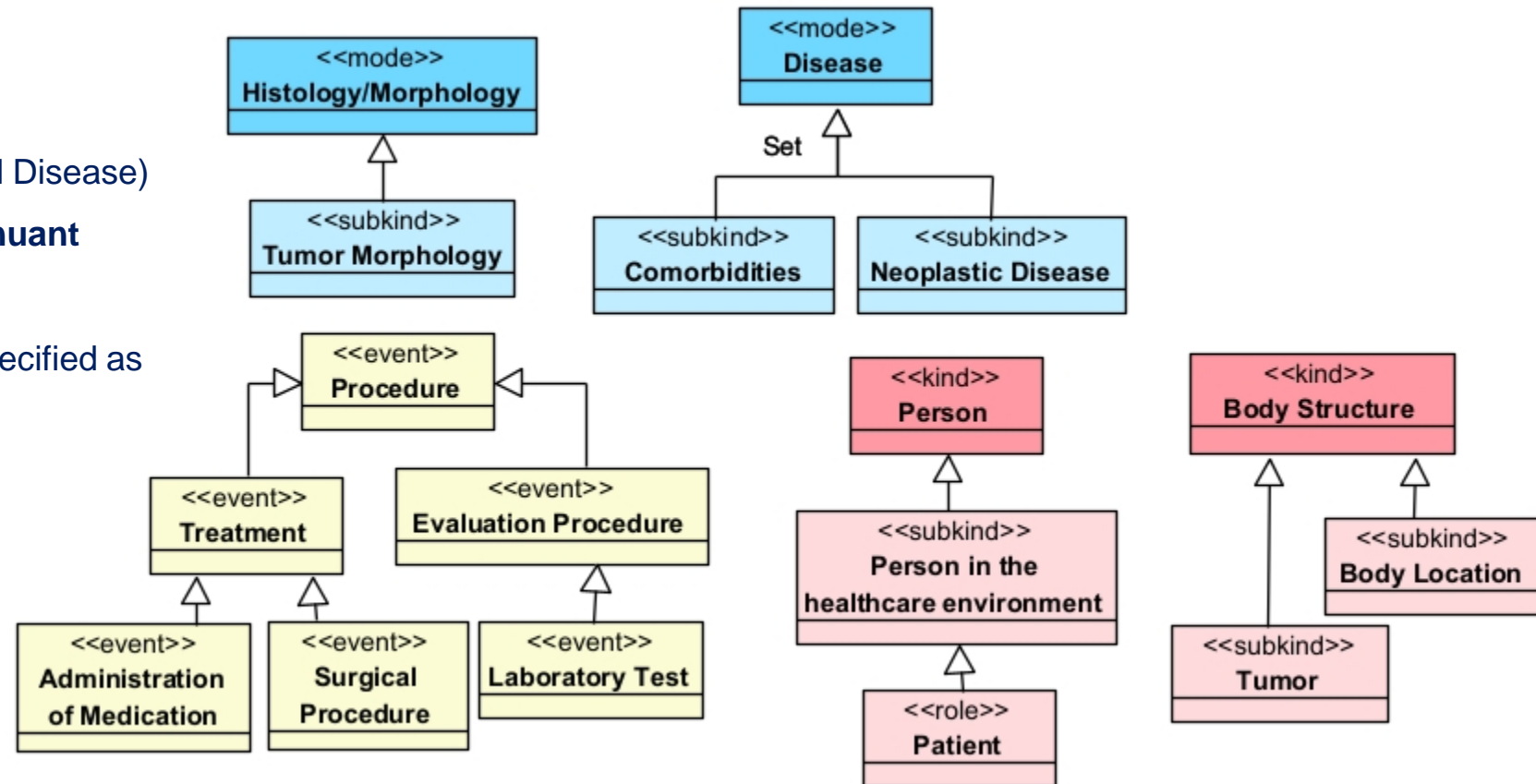


Upper Layer

- Generic categories: Disease, Morphology, Treatment, Procedure, Staging/Grading, Body Structure, etc.

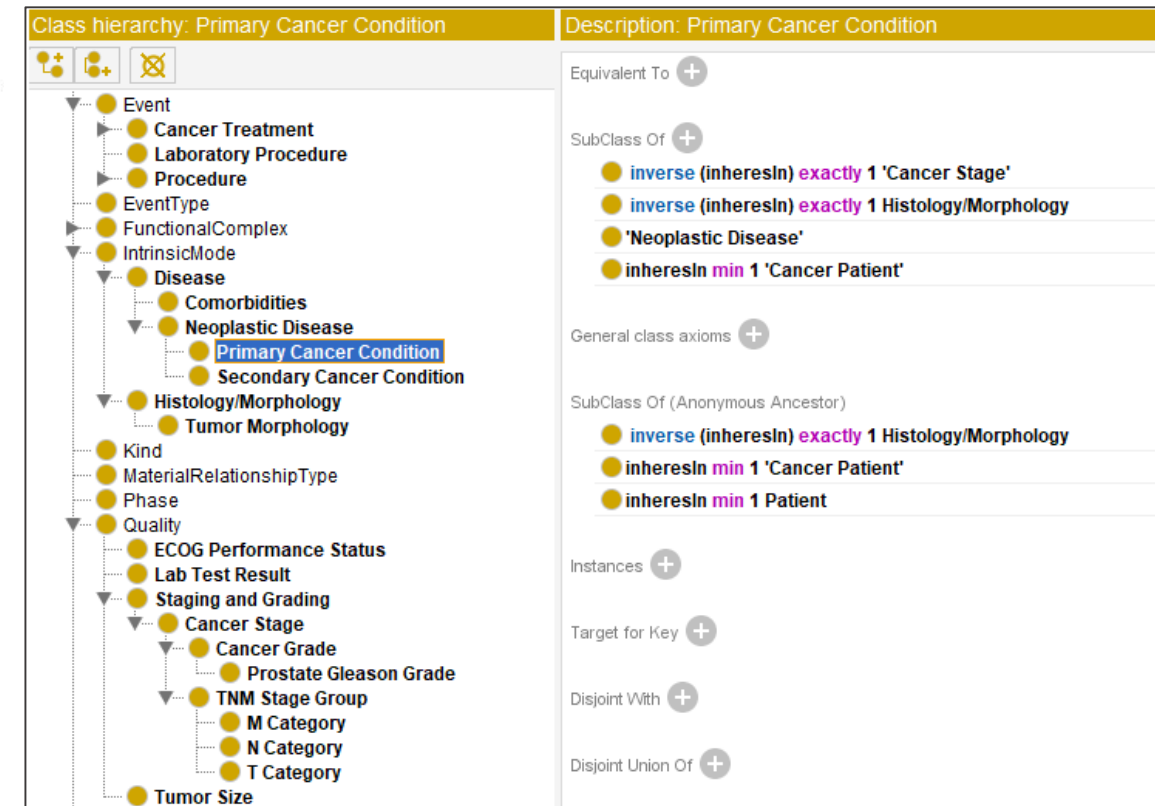
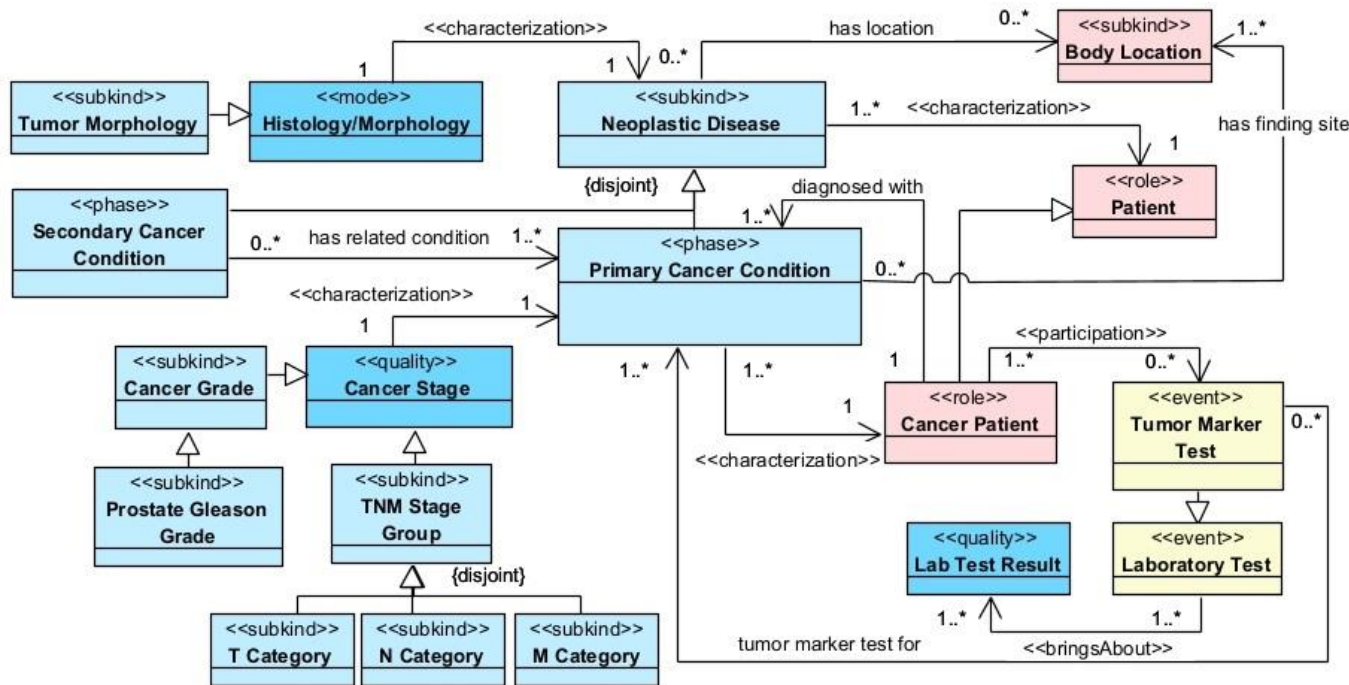
Using BFO [5]:

- *intrinsic modes* (Histology/Morphology and Disease) specified as **Generically dependent continuant** (BFO:0000031)
- *endurants* (Patient and Body Structure) specified as **Independent continuant** (BFO:0000004).

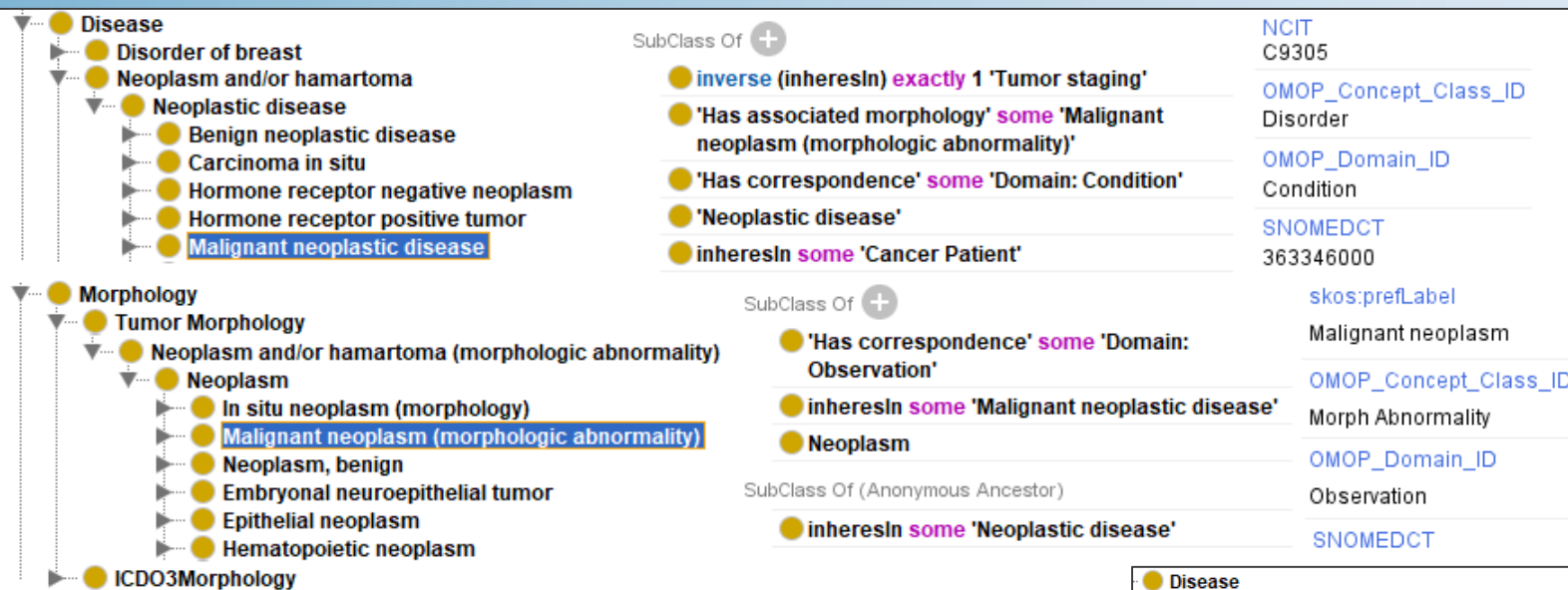


Integration with Bottom-Up

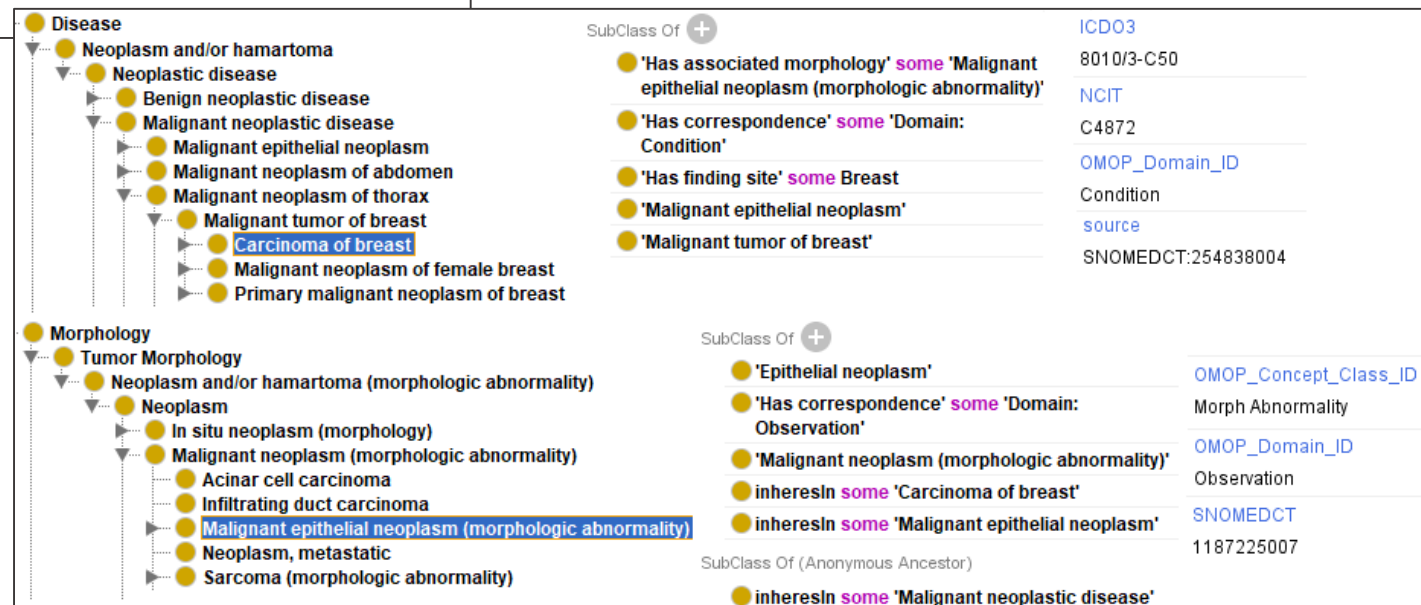
- Reference conceptual is transformed into **operational ontology** formalized using OWL.



Results



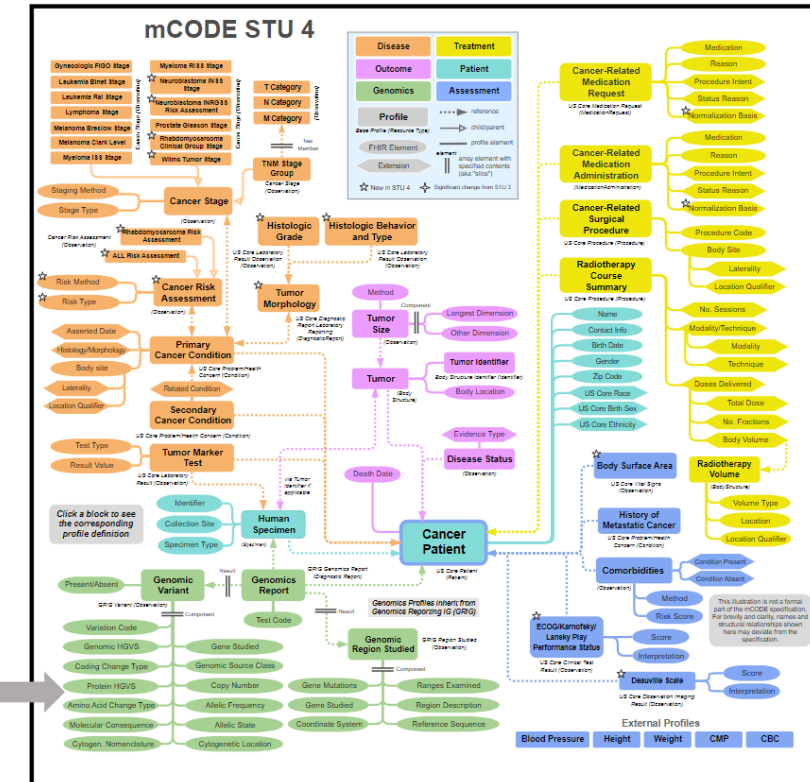
<https://doi.org/10.5281/zenodo.12583826>



Co-funded by
the European Union

Perspectives

- With the help of domain experts:
 - evaluate the mappings and results
 - maintain the upper layer's main categories
- Validate the hyper-ontology using real-world cancer diagnosis scenarios and application-based tasks (federated queries, annotation of images, exploring of data collections)
- Extension of the biological context (genomics)
- Related works: ontological unpacking of a Viral Conceptual Model (VCM) [3, 6]



Thank you

mirna.el-ghosh@inserm.fr

LIMICS, Sorbonne University



Co-funded by
the European Union



EUCAIM
CANCER IMAGE EUROPE