INTERNATIONAL HEALTH TERMINOLOGY
STANDARDS DEVELOPMENT ORGANISATION

SNOMED CT

Content Improvement Project

Combined Inception and Elaboration phases

Project ID: Cancer synoptic reporting
Topic: Anatomic and molecular pathology observables for cancer protocol worksheets
Genetic findings and observables

Date 20180925
Version 0.18

Amendment History

Version	Date	Editor	Comments
0.01	20161108	James R Campbell W. Scott Campbell	Project proposal at request of Jane Millar, Ian Green
0.02		James R. Campbell	Exemplars added to document per review by Matt Cordell
0.10	20170608	James R. Campbell	Scope expanded to include international collaboration on cancer protocols; observables templates edited and updated for AP and MP
0.11	20170920	James R Campbell	Expand templates to include exemplar clinical findings; integrate ICCR harmonization into total workplan
0.12	20171007	James R. Campbell	Add publication of synoptic reports in HL7 CDA standard including map to FHIR API; update project plan
0.15	20171204	James R. Campbell	Develop project timeline and plan for construction phase to include international publication of Nebraska Lexicon© extension
0.16	20171213	W. Scott Campbell	Changed RCP to RCPath throughout document
0.17		James R. Campbell	Introduced modelling templates
0.18	20180925	James R. Campbell	Revised templates for modelling of AP, genomics and MP

Review Timetable

Review date	Responsible owner	Comments
20161218	Matthew Cordell	Request for revision; appendix A proposed
		(remove or add rows if necessary)

Table of Contents

1 Glossary

1.1 Domain Terms

Synoptic report	A document, historically developed as print copy, which summarizes and enumerates findings of importance identified by the pathologist in the course of examining and analyzing patient specimens submitted for pathologic analysis. A cancer synoptic report organizes observations of the pathologist and the clinical and molecular laboratories for a case which has been assigned the pathological diagnosis of malignancy. Synoptic reports including biomarker worksheets prepared and published by the College of American Pathologists© and the International Collaboration for Cancer Reporting serve as a standard for cancer case reporting internationally.
Histopathology	The examination and evaluation results of a biopsy or surgical specimen concluded by a pathologist, after the specimen has been processed and tissue sections have been placed onto glass slides for microscopic examination
Anatomic pathology	A medical specialty that is concerned with the diagnosis of disease based on the macroscopic, microscopic, biochemical and immunologic examination of organs and tissues
Molecular pathology	An emerging discipline within pathology which is focused on the study and diagnosis of disease through the examination of molecules and molecular structure within organs, tissues, tumors or bodily fluids. Molecules often studied include DNA, RNA and proteins.
Nucleotide sequence	An ordered series of base pairs in a DNA or RNA molecule that is the biological information structure which supports genetic transcription and encoding of proteins. Structurally, a gene is a sequence of nucleotides within a stand of DNA.
Chromosome	A double stranded (diploid) pair of DNA molecules which are reciprocal and bound by histones in humans. During mitosis chromosomes condense and become visible by light microscopy, consisting of two chromatids joined at a centromere.
Gene	A gene (from ancient Greek: γόνος, gonos, offspring, procreation) is a locus (or region) of DNA which is made up of nucleotides and is the molecular unit of heredity.[1][2][Wikipedia; noted October 2017]
Genome	A genome is the genetic material of an organism consisting of DNA and RNA including chromosomes, mitochondria, plasmids and chloroplasts in plants. [Wikipedia, noted October 2017]
Mutation	In biology, a mutation is the permanent alteration of the nucleotide sequence of the genome of an organism, virus, or extrachromosomal DNA or other genetic elements.[Wikipedia; noted October 2017]. Mutations may be sequence variants, which is an alteration in the nucleotide sequence of a particular gene, or chromosomal alterations wherein segments of chromosomes may be translocated, deleted, inserted or altered in number of copies.
Germ-line mutation	A mutation which is inherited from the parent. Germline mutations occur in the ovum or spermatozoa of the parents and are passed to the offspring at conception.
Somatic mutation	A mutation which is not inherited from the parent and hence occurs as an alteration in chromosome structure during life.

2 Introduction

2.1 Purpose

The purpose of this project is to review, elaborate, model, deploy and test expanded SNOMED CT content to serve the structured reporting of detailed anatomic pathology(AP) and molecular pathology(MP) data in the diagnosis, staging and treatment of cancer. The scope of the effort includes all IHTSDO members with interest in standardizing their cancer treatment protocols.

In the domain of molecular pathology, it subsumes and replaces Art227316 “Malignancy with gene mutation” and Art63284 “Genetic carrier of X”. It further collaborates with and develops material for Art6283 “Observable entity concept model” and collaborates with and develops material for the LOINC technology preview which is a work item in that project.

2.2 Audience and stakeholder domain

The audience for this document includes all standards terminology leaders, implementers, EHR vendors and clinical users of SNOMED CT and LOINC but is especially targeted at stakeholders from pathology and oncology professional societies including, but not limited to the College of American Pathologists(CAP)), Royal College of Pathology(RCPath), Royal College of Pathologists of Australasia (RCPA), Canadian Association of Pathologists, European Society of Pathologists, International Collaboration on Cancer Reporting(ICCR), American Society of Clinical Oncology(ASCO) and related organizations.

2.2.1 Input from stakeholders

Discussion within the iPaLM SIG including input from staff of CAP, RCPath, RCPA and other pathologist professional groups have widely affirmed the importance of structured data reporting of cancer synoptic reports.
As this project heavily invests in the deployment and use of content in the domain <<363787002 | Observable entity (observable entity) | (henceforth referenced as Observables in this document), the Observables project team and their consultants has been involved with this project and have participated from inception.

2.2.2 Degree of consensus on the statement of problem

Initiatives from a number of IHTSDO members including US, UK, Australia and Canada are underway in the domain of cancer treatment especially in the subject area of personalized or precision medicine which employs detailed genomic observations regarding patients and the cancer episodes they bear in order to guide diagnosis and treatment. It is a matter of wide agreement across these projects that the EHR must support increasingly detailed anatomic and molecular pathology data to guide and document treatment, as well as serve within research databases for investigation. Professional organizations for pathology, oncology and the clinical research community support these needs. There is also an ongoing discussion within the international community regarding trans-national collaboration and coordination of research and clinical use cases in cancer diagnosis. SNOMED CT is uniquely positioned to play a significant role in this work as an international terminology.

Pathology communities and users of pathology data are currently burdened by vendor systems that report only with earlier versions of SNOMED CT and several member countries including Australia, New Zealand, Sweden, and the United Kingdom, are looking into ways to transition to SNOMED CT. Many IHTSDO member countries have projects in progress on selected synoptic reports but are not equipped or have resources to tackle the undertaking of extending and deploying SNOMED CT for in pathology.

3 Statement of the problem or need

3.1 Summary of problem or need, *as reported*

Cancer synoptic worksheets, protocol checklists and tissue pathways are the prevailing form of text-based structured cancer reports based on anatomic pathology (histopathology) and molecular (genomic and proteomic) pathology assessments. Many national pathology societies produce and publish synoptic worksheets for use in their realms and these enumerate the clinically important data elements that should be reported by the pathologist for diagnostic and prognostic purposes by clinical care teams. The international community under the coordination of the International Collaboration on Cancer Reporting (ICCR) is in the process of harmonizing the data elements between national professional colleges to promote internationally consistent cancer reporting. A review of the academic literature and discussion with users of the cancer synoptic data demonstrates that structured reports will not realize their full potential and clinical data will not provide decision support and promote epidemiology until data elements are reported with computable, standardized clinical terminologies. The IHTSDO and LOINC committees have separately supported past projects to accomplish this for CAP synoptic protocol check sheets. Those sets of terminology for AP and MP have not been accepted or deployed primarily related to lack of collaboration and limitations of content. To date, no broadly accepted terminology code set has been successfully deployed for detailed reporting of cancer synoptic data elements. However it is a matter of general agreement that SNOMED CT (Observable entity) and LOINC represent the appropriate semantic domains and are specified by various governmental health agencies for use in recording structured laboratory and pathology data in the electronic health record.

3.2 Summary of requested solution

Develop, deploy and test an extension of SNOMED CT Clinical findings, Observable entities, Body structures, Substances and related domains which uniquely capture and encode the clinical content reported in Cancer synoptic reports. This will initially include 35 of the 82 types of malignancies reflected in synoptic worksheets developed and published by the College of American Pathologists(CAP)[3] of the United States. Concurrent LOINC code assignment for all Observables will serve those IHTSDO members whose national authorities have identified LOINC as the code resource for AP and MP. If publication of the dual content in collaboration with the US release center - National Library of Medicine (US) - leads to collaborative testing among IHTSDO members, proves successful and response from IHTSDO member community is favorable, the content will be proposed for promotion to SNOMED CT International edition for deployment by all IHTSDO members.

International harmonization will occur concurrently with development and will consist of review of content and translation of fully specified names and preferred terms by cooperating international professional societies coordinated by the ICCR. Subject matter from collaborating professional organizations including translations will be harmonized and an expanded and comprehensive core data set supporting precision medicine in cancer serving all IHTSDO members will be developed from this body of work. Deliverables will support full interoperation of cancer synoptic data across all IHTSDO members.

3.3 Statement of problem *as understood*

The SNOMED CT concept model does not currently support detailed or fully defined genomic or proteomic data within Observable entities, Clinical findings or related hierarchies. In addition to applying a harmonized SNOMED CT – LOINC concept model for Observable entities to the details of AP and MP observables, extensions to SNOMED CT content and concept model will need to be prepared and tested for MP observables which relate to genotypic and phenotypic evaluation results.

In a past project[4], Cancer synoptic worksheets from CAP were prepared and vetted for scientific content but were not validated by informaticians for representation of discrete, well-formulated data. Semantic analysis and systematic application of a harmonized concept model for Observable entities will be required for all 82 work sheets.

Some of the content within scope of the project exists in SNOMED CT, LOINC or both. Most of the content represents new material by virtue of the evolution of the science, or due to the higher degree of granularity offered by the harmonized concept mode for Observable entity. Conceptual content developed from semantic analysis of the CAP work sheets will need to be reviewed for content overlap with existing SNOMED CT or LOINC content. Existing SNOMED CT content will need to be fully defined when possible and tested. Existing LOINC concepts will be fully defined with the harmonized model and published within the expression data set agreed by the IHTSDO and Regenstrief. All Observables material will be reviewed and assigned LOINC codes by the LOINC committee.

Publication artifacts from the project will include an RF2 release of the extension content and any associated map and translation refsets for the IHTSDO community of use. In collaboration with Regenstrief, an ontology of LOINC concepts in OWL format accompanied by maps to SNOMED CT defining elements will be developed for metadata development in service of data warehouses for research and public health.

3.4 Detailed analysis of reported problem, including background

History of worksheet coding: CAP was the original owner and developer of SNOMED. For their member use within the discipline of anatomic pathology, they developed many Observables concepts[3] which were carried forward in the merger that developed SNOMED CT. However, systematic implementation of this code set in any electronic record software never occurred and all Observables concepts in SNOMED CT today are modelled as primitive as no concept model existed for Observables at the time of the original work. Approximately 160 concepts currently exist within the body of SNOMED CT from those efforts.
The laboratory LOINC committee developed LOINC codes for CAP concepts in anatomic pathology in projects during 2003 and 2006. These total 122 concepts in LOINC 2.56 and are labelled with Method of “CAP cancer protocols”.

Appearance of molecular biomarkers: CAP began with the addition of work sheets for tumor biomarkers in 2014. This was in response to the increasing clinical use of molecular and genomic data in the diagnosis, prognosis and treatment planning for cancer. They currently number 13 of the 82 types of cancer contained in the AP work sheets. The laboratory LOINC committee has been very active in recent years in molecular pathology and genomic observables, and 1379 concepts exist in LOINC 2.56 with the method “Molgen”. An additional 1252 molecular probes employing polymerase

chain reaction methods exist, targeting DNA and RNA of microorganisms for microbiology. The concepts are all defined as observations on tissue or other specimens and no elements in the LOINC model support the differentiation of somatic and germ-line mutations.

Consistency and alignment between ITHSDO members: IHTSDO members have professional societies with attention to pathology and laboratory medicine and there is no consistency at this time in the application and reporting of pathology data for cancer across realms. Pathologist professional organizations from many IHTSDO members formed the International Collaboration on Cancer Reporting(ICCR) in 2011. This organization has the stated objective of consolidating and standardizing cancer datasets for pathology reporting. At this time the collaboration has produced a handful of datasets.

The International Pathology and Laboratory Medicine (iPALM) SIG of the IHTSDO began a collaborative effort with the RCPath in London in the fall of 2015. At that meeting, pathology professionals from the US and UK met with consultant terminologists and leadership of the IHTSDO to discuss conceptual content and procedures for application of the Observables harmonized concept model in service of synoptic protocols. iPALM has convened regular meetings and teleconferences since that initial gathering and has expanded scope to involve professional participation from six IHTSDO member countries and the ICCR. Cross mapping of content between CAP, RCPath and RCPA protocols has already begun with an effort in colorectal and pancreatic cancer.

At an iPALM meeting in London April of 2017, RCPath, RCPA, ICCR and the Swedish Society of Pathologists sent delegations. Colorectal harmonization activities were reported by RCP and the Swedish delegate reported on pancreatic cancer coding. By unanimous agreement of the delegates this project was extended to include an ICCR harmonization report synoptic with the CAP cancer protocol reports developed by Nebraska as an informative source. The next project for harmonization was agreed as lung cancer.

At a Pathology project kickoff meeting in Bratislava in October of 2017, representatives of RCPath, RCPA, CAP, ICCR and the Swedish Society of pathologists met with IHTSDO leadership and agreed to a workplan in which additional pathology protocol worksheets would be vetted by the professional societies and terminology content would be published by Nebraska as Nebraska Lexicon© content in a module maintaining cancer protocol development from this project.

3.5 Subsidiary and interrelated problems

Challenges of genomic data: Neither SNOMED CT nor LOINC currently employ an anatomic subcellular anatomy, molecular process inventory or detailed molecular model for nucleic acids or cellular proteins. Both terminologies, especially LOINC, include terms in concept descriptors that relate to these subcellular features that have emerged from research on the human and cancer genomes as clinically relevant to human medical practice. At this time the clinical terminology has no defining features in the concept model underlying SNOMED CT.

Disconnects from scientific terminology resources: The scientific community has been active in organizing, structuring and publishing ontologies and databases which document the knowledge flowing from the explosion in research. For medicine, the Human Genome Organization supports the activities of the Human Gene Nomenclature Committee which maintains a database of human genes

that cross references the databases of DNA, RNA, proteins and molecular cellular processes at the basis of human life. These resources form a rich terminological knowledge-based network of reference material which has no computable links to the clinical terminologies of the EHR. Employing these resources as referential definitions of the language of clinical molecular pathology would be an important step in uniting the bench and the bedside.

Complexity of harmonized concept model: The semantic domain of Observable entities proposes to serve a broad spectrum of types of observation results originating in laboratory medicine, pathology, radiology, physical examination, patient history and more. In order to define such a varied spectrum of concepts, the harmonization concept model proposed is complex and hard to understand. This is a major threat to reproducibility of terminology development in this domain and early studies suggest that consensus-based templates applying the proposed model to individual use cases will be necessary to support uniform results in concept model deployment. Application of the proposed model to a new complex field such as molecular pathology and genomic/proteomic observables will pose a substantial challenge.

Issues from SNOMED CT - LOINC harmonization: The agreement between RI and the IHTSDO in 2008 specifies that no SNOMED CT international release concept identifiers may be developed or employed in communication or aggregation of Observables data. This places substantial restraints on the publication and promulgation of Observables harmonization for the IHTSDO community. The current agreement, negotiated with NLM, IHTSDO leadership and the Regenstrief Institute, specifies that publication of developed content will occur with only LOINC codes as concept identifiers but that SNOMED CT data systems may load the material employing extension concept codes unique to their site. SNOMED CT extension concepts and identifiers are not subject to this agreement and it was agreed in Bratislava in October 2017 that interim publication of material by the project will be confined to Nebraska Lexicon© extension concepts which could be shared with collaborating IHTSDO members and evaluated for future publication.

4 Risks and Benefits

4.1.1 Risks of not addressing the problem

The challenges to clinical terminologies imposed by Personalized (Precision) Medicine will not go away and are likely to rewrite standards for clinical documentation and conduct of health care. A central focus of this upheaval is the diagnosis and treatment of cancer. If SNOMED CT does not expand and refine the concept model to serve these use cases, it runs the risk of becoming irrelevant and facing new challenges from alternative terminologies yet to be developed or in use in the scientific community.

Increasingly, medicine is a collaborative effort and international in scope. Sharing of research and clinical data across national borders is becoming widespread. Interoperation of such data is becoming a requirement in order to assure effective aggregation of datasets. SNOMED has a unique historical position as a leader in terminology for pathology and to drop that baton SNOMED would suffer loss of international credibility and prestige.

At this time the world-wide community of use of the EHR has a schizophrenic attitude toward the deployment of Observables content. Some IHTSDO member realms employ LOINC codes while others shun their use but have no substitute for content. A viable, consistent approach to interoperation must be addressed or we will suffer more fragmentation of the use community and not follow the spirit of the 2008 harmonization agreement.

4.1.2 Risks of addressing the problem

Risk of merging with scientific term sets relate primarily to editorial migration of the principles and practices of the scientific references, with resultant semantic disconnect and/or ambiguity.

IHTSDO members may be confused with proposed procedures and rebel, asking for SNOMED CT compliant solution in publication of Observables.

Efforts to encode molecular pathology by extending the Observables concept model to include subcellular anatomy may diverge from other as yet unnamed terminology efforts, leading to an orphan terminology development not acceptable to the international community.

5 Requirements: criteria for success and completion

5.1 Criteria for success/completion

5.2 Strategic and/or specific operational use cases

GOAL 3: DEVELOP & EXECUTE A ROADMAP FOR THE COMPLETION OF RELEVANT CONTENT/MAPPING WORK FOR SNOMED CT THAT GIVES IHTSDO THE DIRECTION TO MARKET WITH ALL STAKEHOLDERS
GOAL 5: START DEVELOPMENT OF WORK TO POSITION IHTSDO AS A LEADER IN THE AREAS OF MOBILE HEALTH, CONSUMER HEALTH, GENOMICS, RESEARCH & BIG DATA ANALYTICS

5.2.1 Use case 1: Deployment of work sheet data in an EHR

Test case 1(Clinical care): Clinical acceptability of cancer worksheet structured/coded data in Epic (at Nebraska) as fit for purpose
Test case 2(Clinical research): Demonstration of the clinical utility of observations of tumor budding (a novel pathology feature in AP) in the prognosis and staging of colon cancer.

5.2.1.1 Fit with IHTSDO strategy

Deployment of structured AP and MP data for clinical cancer management relates to both goals 3&5 of the IHTSDO strategic plan. These data sets represent critical components of national plans for cancer diagnosis and treatment. The addition of molecular pathology data adds the important dimension of genomics to the project and assures the importance of this project for big data analytics. Involvement by multiple IHTSDO members with the iPALM project support the importance of this work to the IHTSDO.

5.2.2 Use case 2: Deployment of work sheet data in research data warehouse

Test case(Personalized Medicine): Assemble a cohort of ‘triple negative’(Estrogen receptor, progesterone receptor and Her2 receptor) breast cancer cases for staging research by oncology investigators integrated in a research dataset with EHR and tumor registry data

5.2.2.1 Fit with IHTSDO strategy

Extension of the SNOMED CT concept model for genomic and proteomics, binding of the concept model to ontologies from NCBI and the application of the extension data within a ‘big data’ warehouse uniquely position this project on the forefront of strategic goal #5.

6 Solution Development

6.1 Design

6.1.1 Outline of design

6.1.1.1 Prioritize the sequence of synoptic development with IHTSDO members and the ICCR. For those synoptic reports identified, iteratively analyze the semantics found in CAP and ICCR work sheets; define content and FSN within the SNOMED CT concept model for required observables; review with iPaLM, CAP and ICCR domain experts for subject matter and agree on content semantics and SNOMED CT definition

6.1.1.2 Partition Nebraska Lexicon© namespce into clinical, laboratory observables and synoptic pathology modules; organize terminology workspace for international publication and content sharing

6.1.1.3 Extend SNOMED CT content (primarily attributes and qualifier values) to support Observables definition in Anatomic Pathology

6.1.1.4 Extend the SNOMED CT concept model and add content in observables, body structures, substances and qualifiers to include genes and proteins as needed for Molecular Pathology use cases

6.1.1.5 Bind Genes, subcellular anatomy and proteins by reference to NCBI ontologies including HGNC, Uniprot and Gene Ontology and classify with SNOMED CT

6.1.1.6 Vet analyzed content with Observables project for definition and application of the concept model; define use cases and develop consensus templates for application of concept model; document all templates as part of an editorial observables guide

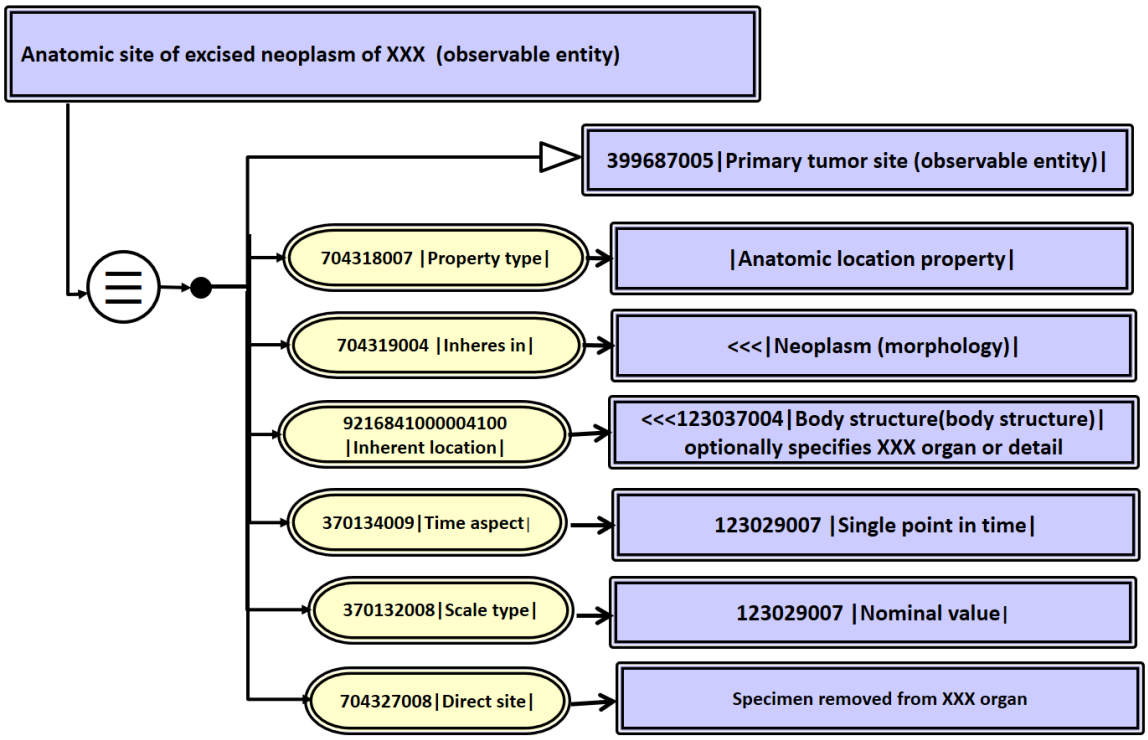
6.1.1.6.1 Anatomic pathology[6]

Morphologic and Histopathologic Tumor Observations

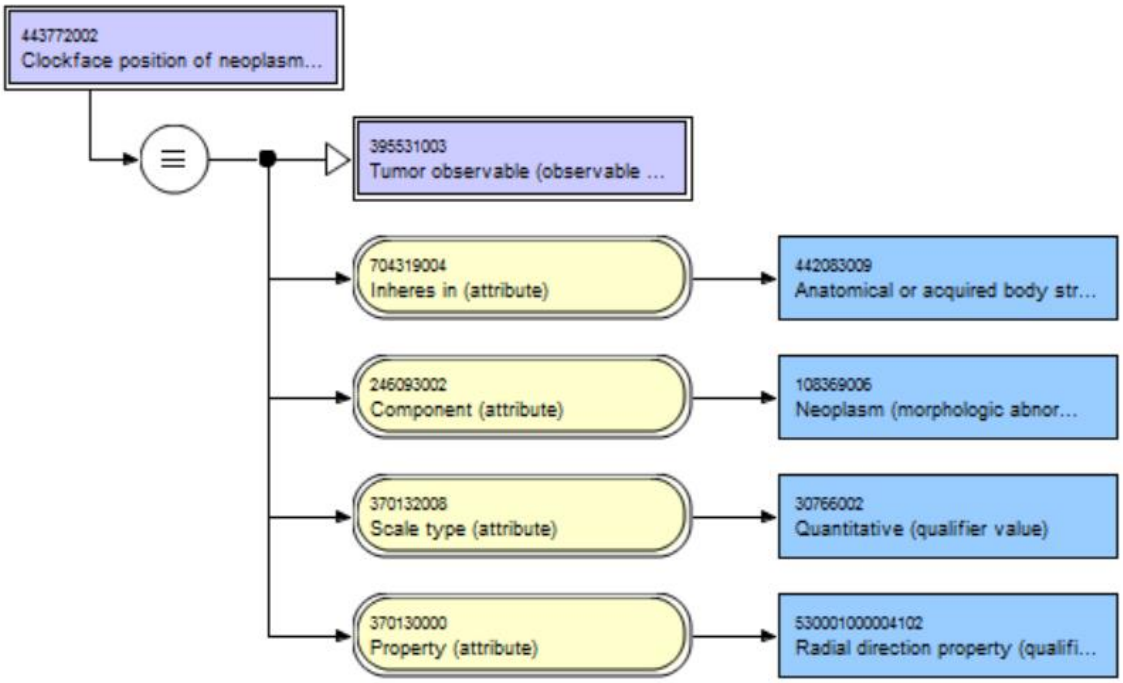
Routine histopathologic observables required addition of several <<|Measurement property(qualifier)| in order to fully define characteristics of the microscopic and morphologic observations made by pathologists.

Pre-coordinated SNOMED CT Observables had many references to ‘tumor’, ‘neoplasm’, ‘malignant neoplasm’ and various specifications of cancer histologic types. In order to properly classify these concepts within the SNOMED CT Abnormal and acquired body structures, ‘tumor’ in existing FSNs was interpreted to equate with 416939005|Proliferative mass(morphologic abnormality)| which is a supertype of both neoplasm and malignant neoplasm. Hence a surgical procedure which excises an enlarging mass (tumor) may be subsequently proven to be non-neoplastic, neoplastic or malignant on histologic examination.

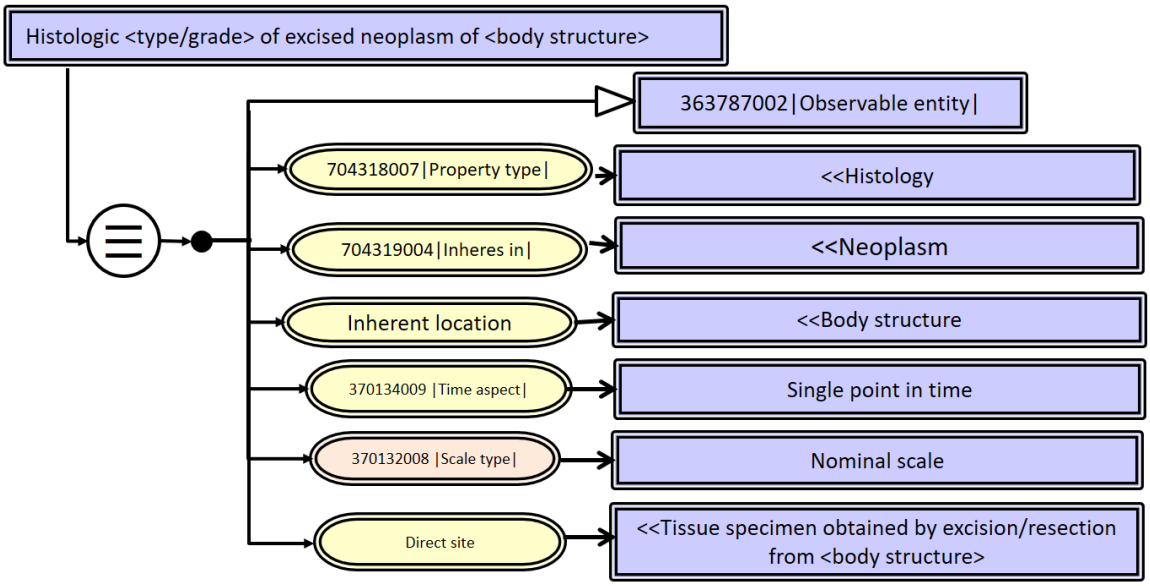
Anatomic site of excised neoplasm



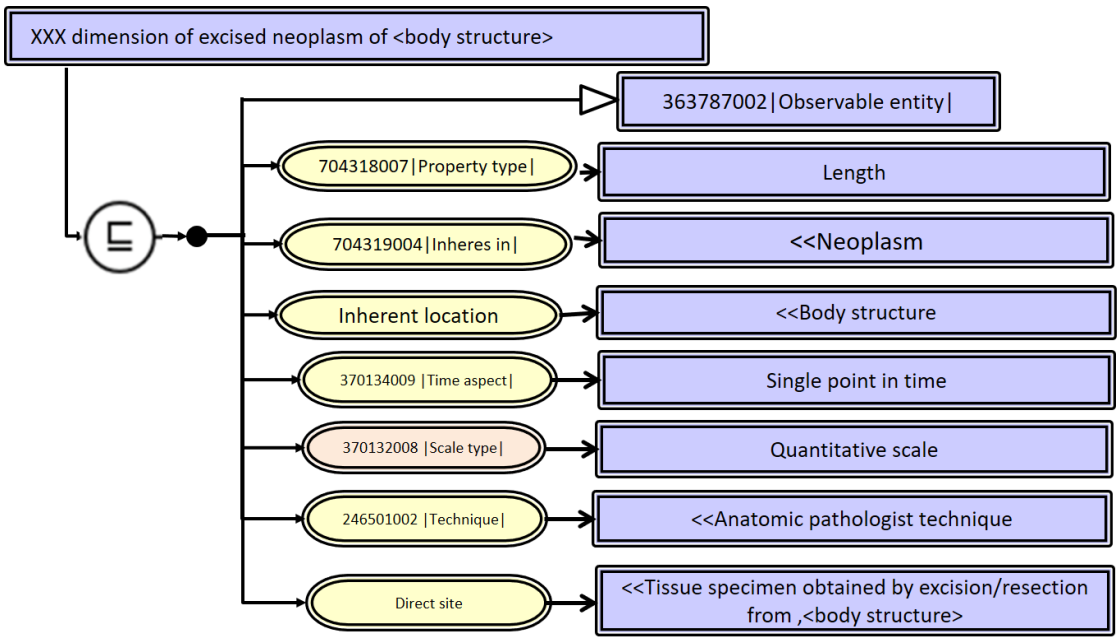
Clockface location of <neoplasm> within some <anatomic structure>



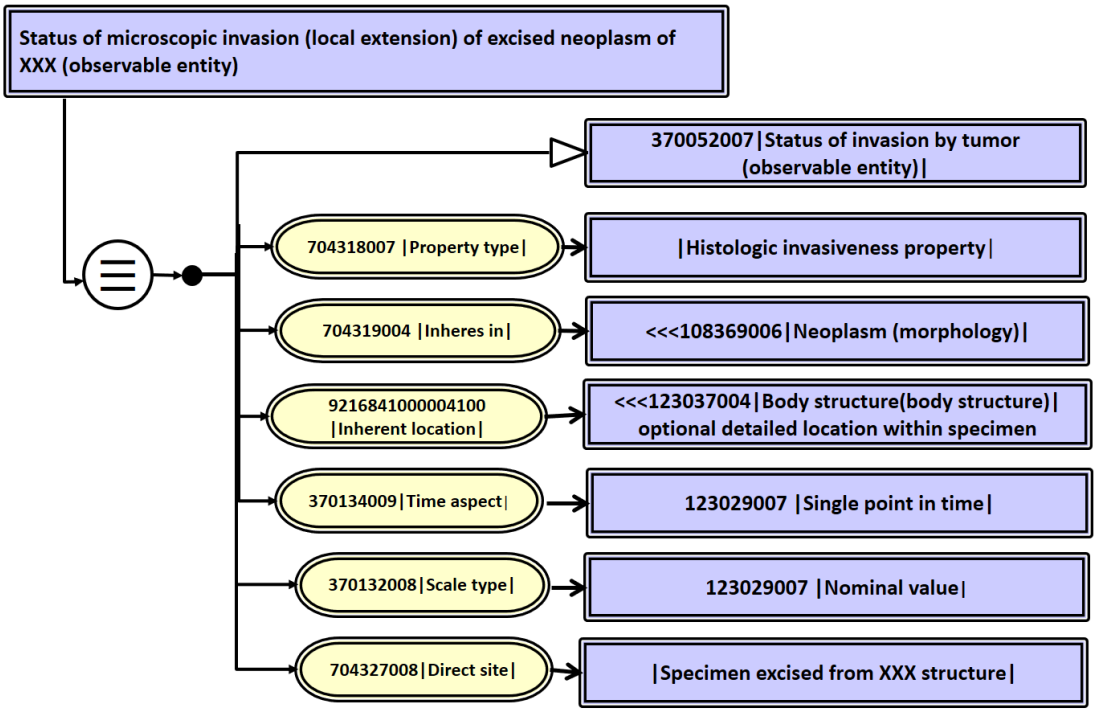
Histologic <type or grade> of excised neoplasm of <body structure>



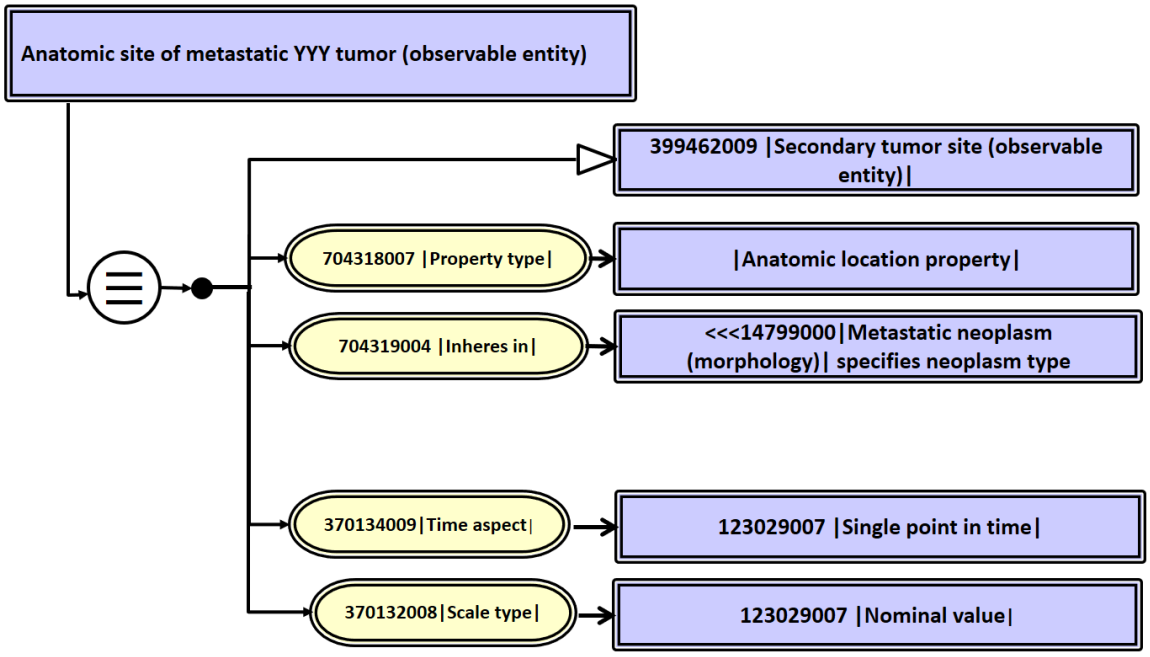
Tumor size of excised neoplasm



Status of microscopic invasion of excised tumor of XXX
Local invasiveness of excised neoplasm of XXX



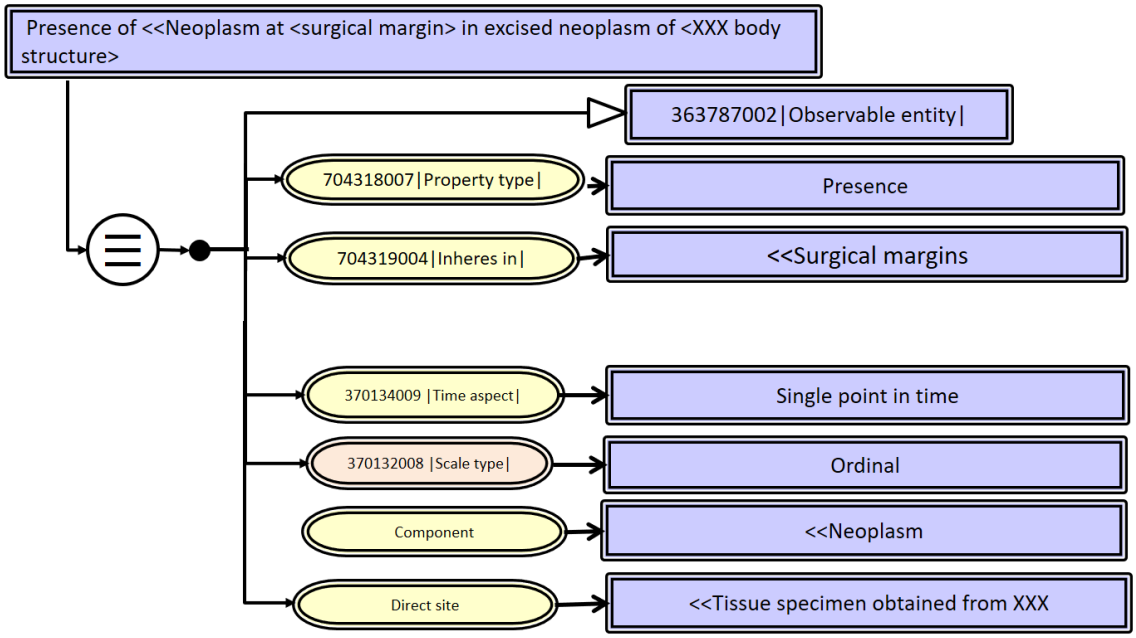
Anatomic site of distant metastasis of neoplasm type YYY



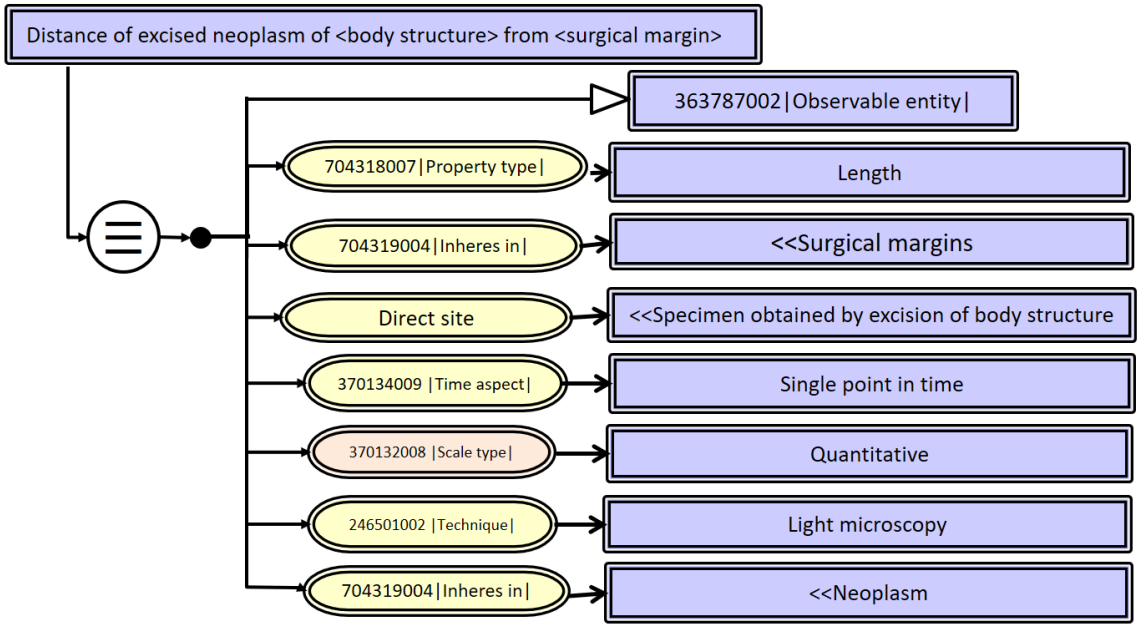
Observations of and about surgical margins

Presence(status) of XXX surgical margin involvement by excised neoplasm of <site body structure>

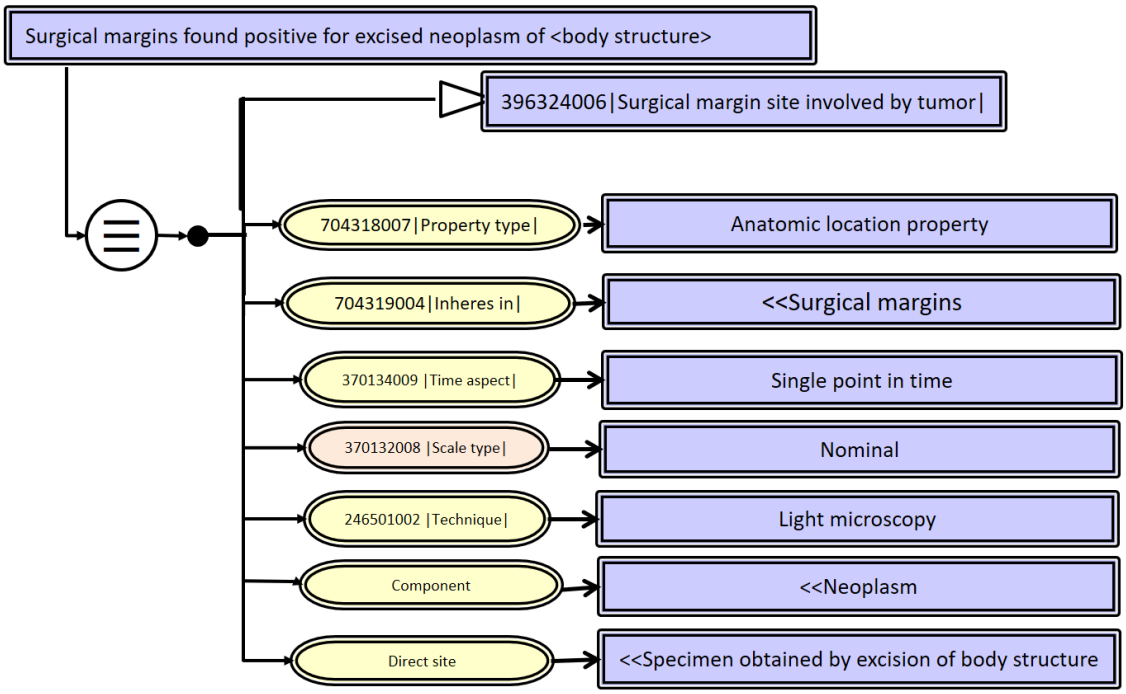
Is the neoplasm to be found at the surgical margin?



Distance of excised <neoplasm> of <body structure> from <surgical margin>



Surgical margins found positive for excised neoplasm of <body structure>

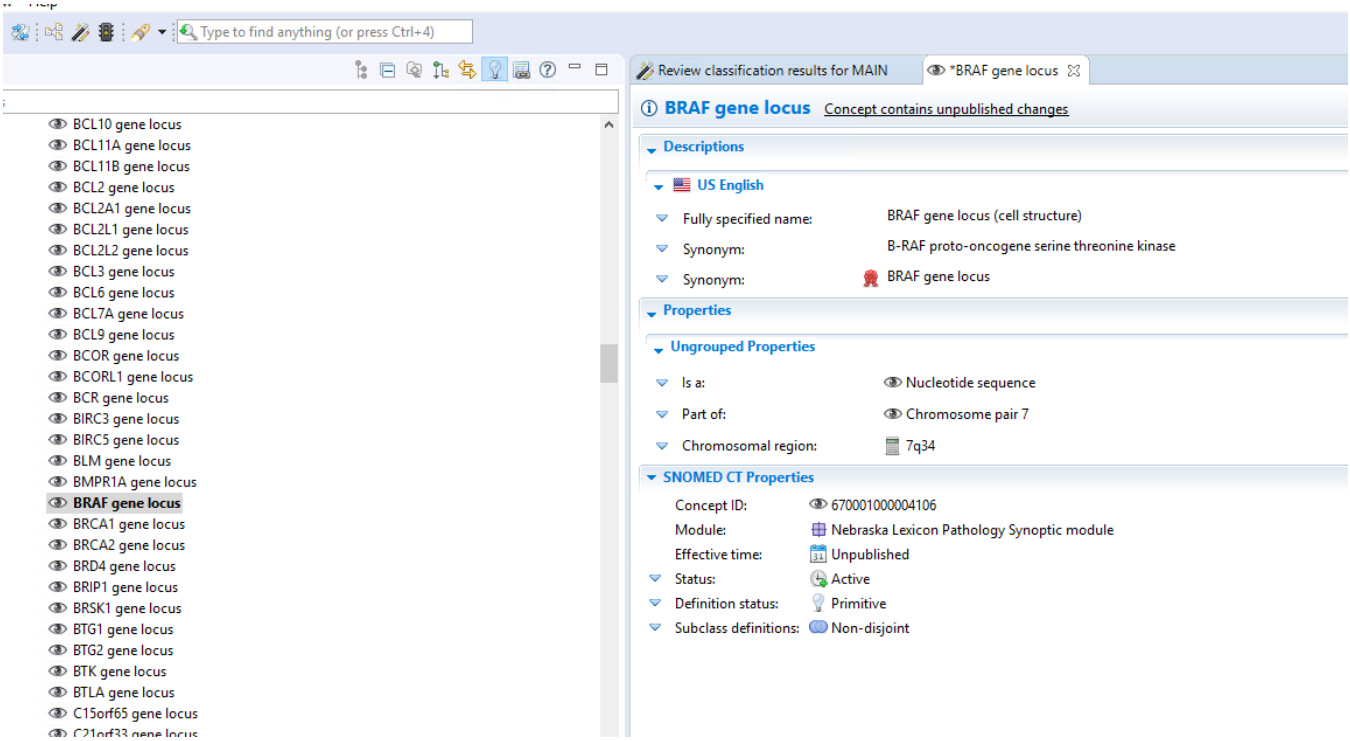


6.1.1.5.2 Molecular pathology and genomic observables[7],[8]

6.1.1.6..1 Genetic subcellular features

Nucleotide sequences and gene loci

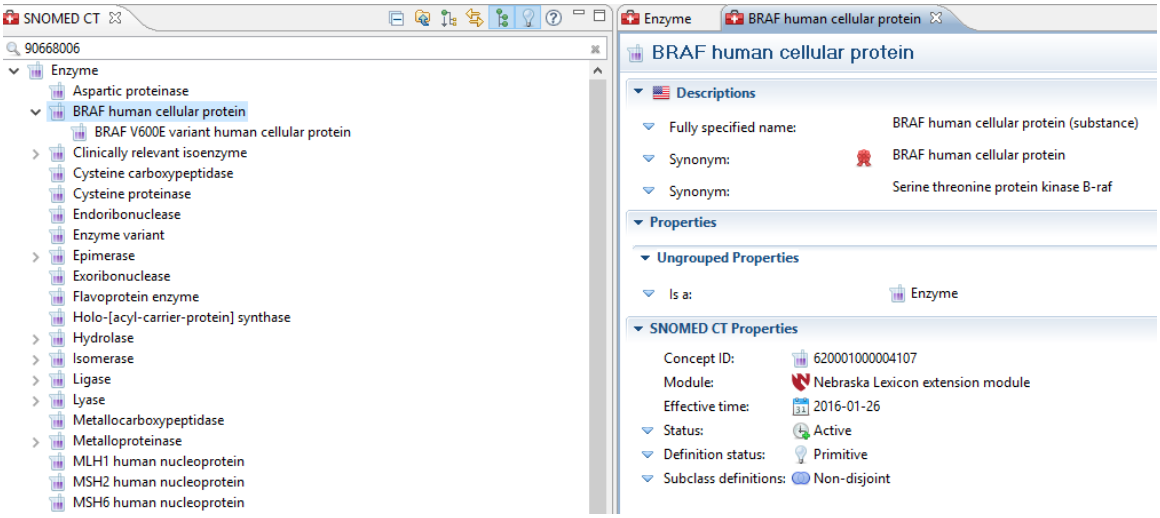
Gene loci and nucleotide sequences are structural features of the subcellular environment found in nuclear DNA. DNA and RNA may be found in the nucleus, mitochondria and the cytoplasm but nuclear and mitochondrial DNA specify the genetic makeup of the eukaryotic organism including humans. SNOMED International core contains no detailed representation of gene structure and those are necessary to fully define many MP observables. Nucleotide sequences are added as subcellular structures and subtypes of named gene loci are added and defined by reference to the HGNC structural data which has arisen from sequencing and characterization of the human genome. Sequence variants are further defined as subtypes of nucleotide sequence which represent specific structural gene alterations of clinical significance.



Proteins

Proteins are large biomolecules, or macromolecules, consisting of one or more long chains of amino acid residues. Proteins perform a vast array of functions within organisms, including catalysing metabolic reactions, DNA replication, responding to stimuli, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in protein folding into a specific three-dimensional structure that determines its activity.[Wikipedia, 2017]

In the SNOMED CT concept model, proteins are subtypes of 105590001|Substance| and are categorized functionally and by molecular properties. The concept model has not been elaborated for Substances, however as part of the molecular pathology developments for SNOMED CT, we anticipate providing definition by reference for proteins to Uniprot and other scientific databases.

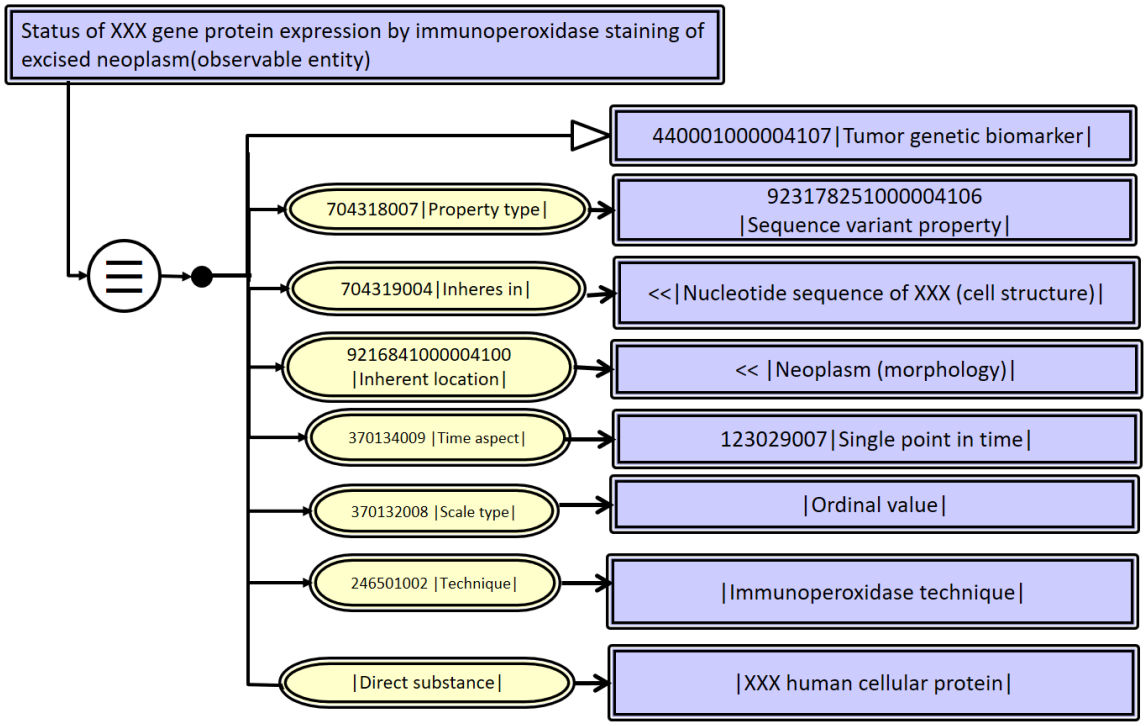


6.1.1.6..2 Tumor biomarkers; somatic mutations in cancer

Molecular observations regarding genetic makeup of tumors may be accomplished using: a)immunohistochemical analysis for (normal or) variant proteins as surrogates for nucleotide sequence, b)by directly observing nucleotide sequence data from neoplastic tissue, c)by amplifying or hybridizing neoplastic DNA or RNA and determining presence and location of the (normal or) abnormal nucleotide segments using fluorescent, radiological or other probes.
For indirect observation of sequence analysis, immunohistochemistry basically counts cells staining positive with protein-specific fluorescent stains. Both immunohistochemistry and sequencing observations are seeking to assess the detailed gene structural differences that in turn characterize the genome of the neoplasm, Hence both classes of these observables INHERE_IN the gene locus (nucleotide sequence) of interest but assess different phenotypic features - either nucleotide sequences or the proteins they produce.

Immunohistochemical data

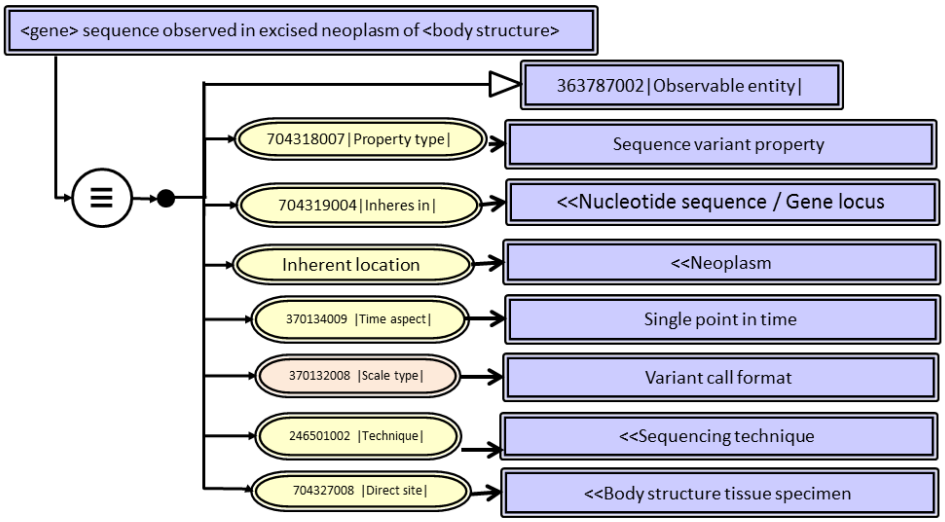
Expression status of <Protein> coded by <gene> within excised <neoplasm> of <body structure>



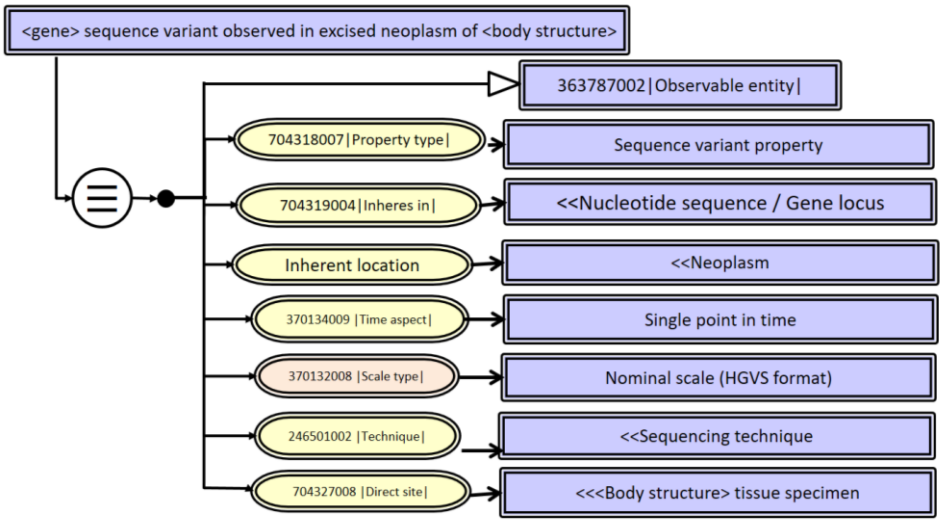
Neoplastic sequence data

Nucleotide sequence observations (single nucleotide polymorphisms) may be reported as: 1) full file (variant call format(vcf)) report of an entire chromosome or set of genes sequenced, 2) Human Gene Variant Society(HGVS) formatted data reporting only sequence variants found in a specific gene locus or RNA segment along with the reference genome or 3) pre-coordinated ordinal level data reporting the presence or absence of a specific previously characterized sequence variant. Sequence variants may be germ-line if they were inherited from the parents (congenital) or somatic if they occur developmentally later in life. The majority of cancers (malignant neoplasms) represent growths that originate as cell lines arising from somatic variants.

Nucleotide sequence (raw data in vcf format) detected in <gene locus> of excised neoplasm of <body structure>

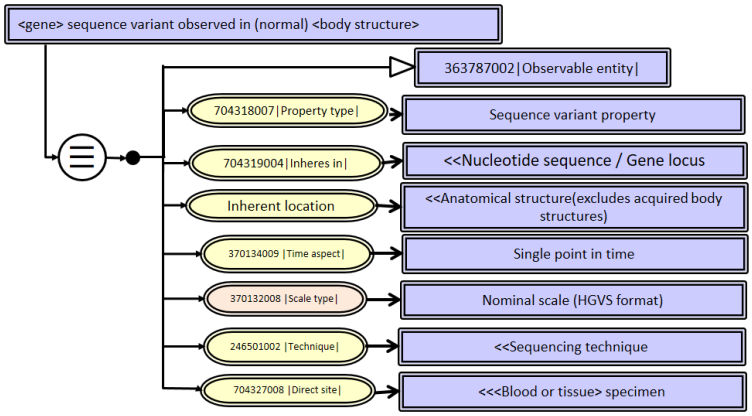


(Somatic) sequence variant detected in <gene locus> of excised neoplasm of <body structure>
What is the sequence variant identified in the gene of a morphologically abnormal specimen?

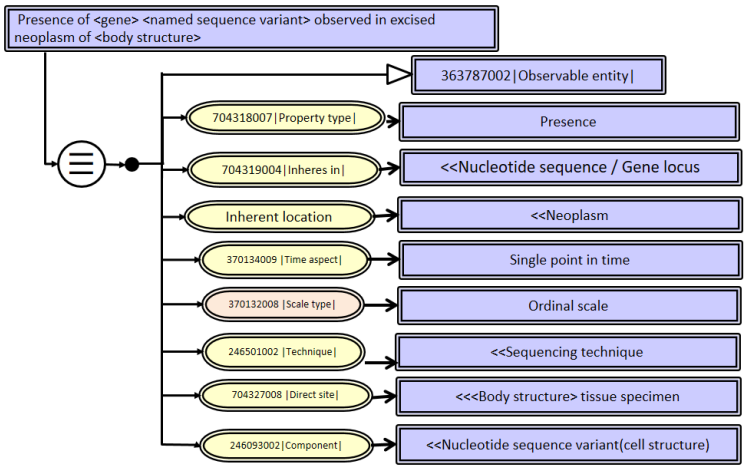


Germ-line sequence data are observations made about nucleotide sequences in tissue or cellular structures that are morphologically indistinguishable from the patient. Hence, while the somatic observable has *Inherent_location* <<108369006|Neoplasm(morphologic abnormality)|, germ-line observables have *Inherent_location* <<91723000|Anatomic structure(body structure)|.

(Germ-line) sequence variant detected in <gene locus> of <anatomical structure(excludes morphologically abnormal structures)>



Presence of (somatic) <sequence variant> in <gene locus> of excised neoplasm of <body structure>

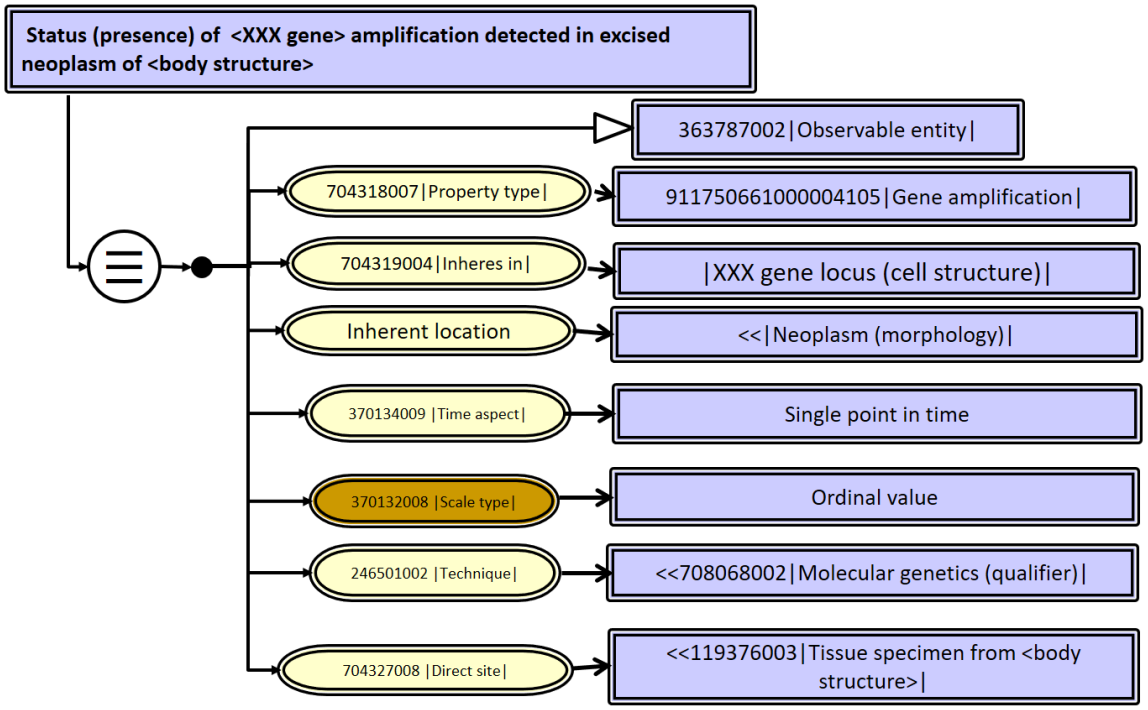


6.1.1.5.3 Chromosomal and gene structural variations

Nucleotide sequence variants represent only one type of genetic variation that may be instrumental in disease pathogenesis. The first genetic traits discovered in the history of medicine were chromosomal structural changes that could be observed microscopically with chromosomal banding techniques. Today, fluorescent in-situ hybridization supplemented by sequence data allows for more detailed analysis of chromosomal and gene structural variation. For this reason, new <<118598001|Measurement property(qualifier)| attributes were added including 911750461000004108|Chromosomal structural rearrangement (qualifier| and

911750661000004105|Gene amplification(qualifier)|. Gene amplification occurs when the number of expressed alleles of a gene increases beyond the two that are normal in the human genome. Gene amplification can occur when another copy of the chromosome is produced, when a gene is duplicated and translocated to another chromosomal position or in a variety of other structural changes.

Status of XXX gene amplification in excised <neoplasm> of <excised body structure

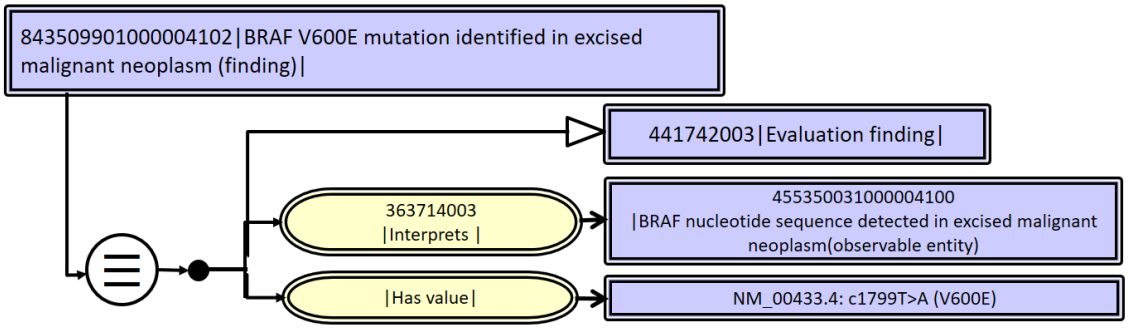


6.1.1.5.4 Molecular pathology and clinical findings

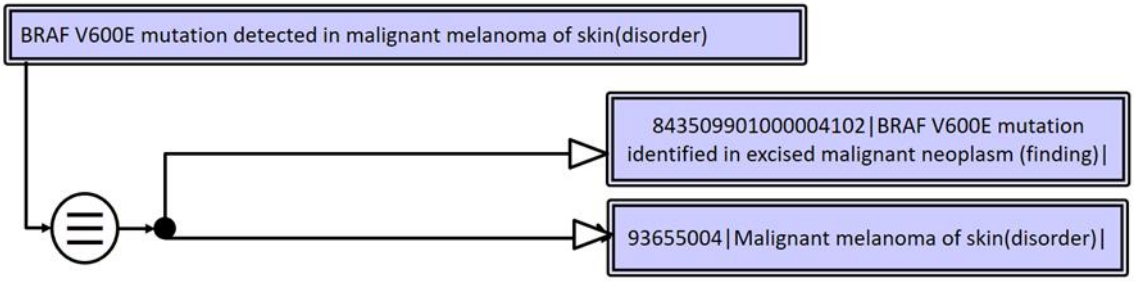
Employing genetic observable data in the definition of clinical findings

Somatic and germ-line observations of sequence variants and other chromosomal abnormalities can be employed to fully define congenital genetic diseases as well as specific subtypes of malignancies based upon genetic data.

BRAF V600E mutation identified in malignancy



BRAF V600E mutation detected in malignant melanoma of skin



6.1.1.7 Deploy and test new content in Nebraska Lexicon© extension namespace

6.1.1.8 Publish RF2 and OWL content in NLM UMLSTS for member comment; publish annotated CAP work sheets including coded observables and SNOMED CT valuesets; develop browser version for Nebraska Lexicon© to support external review and analysis

Discussions with NLM at the Wellington meeting initially agreed upon the following artifacts: 1)RF2 release structures for Nebraska Lexicon© including integrated technology preview employing LOINC code map for external communication and installation software for IHTSDO compliant implementation; 2)Observables ontology for harmonized LOINC-SNOMED CT content in OWL format; 3)CAP Cancer Worksheets© in pdf format annotated with Observables coding in LOINC and SNOMED CT valuesets.

Based upon revised project status agreed at Bratislava, an RF2 module representing all synoptic pathology project content will be published twice yearly from the Nebraska Lexicon© website[5] in cooperation with NLM. Content will be made directly available to all interested IHTSDO national release centers for evaluation and deployment.

6.1.1.9 Deploy content in UNMC Anatomic Pathology system (COPATH®) and interface coded and structured Observable-value pairs to the clinical enterprise employing HL7 messaging

6.1.1.10 Collaborate with vendors of sequencing hardware and industry knowledge developers to automate reporting of sequence data by HL7 V2.X messaging to clinical and research data systems

6.1.1.11 Outcome report: Publish scientific paper documenting changes in pathologist workflow and structured data integrated into electronic health record[3]

6.1.1.12 Outcome evaluation: Deploy interfaced content to research data warehouse and assess research test case

6.1.1.13 Outcome evaluation: Deploy interfaced content to the EHR (Epic®) and assess fitness of purpose for clinical use case of cancer diagnosis and treatment planning

6.1.1.14 Outcome evaluation: Assemble evaluation data from test deployments and submit with content for promotion to the International release of SNOMED CT

6.1.1.15 ICCR domain experts collaborating from non-english speaking countries will translate FSN and preferred terms for their NRC extension namespace as cancer protocol content is developed; language refsets will be published by Nebraska as part of their twice annual release

6.1.1.16 Develop implementation guide for HL7 CDA publication of synoptic reports employing vetted pre-coordinated coded content in SNOMED CT and LOINC

6.1.1.17 Outcome evaluation: Retrospectively mine structured data via NLP from historical CAP worksheets from Nebraska for breast, pancreas, lung and colorectal cancer content to FHIR API for interoperability; demonstrate interoperation of structured (HL7 V2) and summary (HL7 FHIR) data

6.1.1.18 Outcome evaluation: Paper reporting merged epidemiologic data for breast cancer from cooperating centers

6.1.2 Significant design or implementation decisions / compromises

Multiple face-to-face meetings of iPaLM and Observables project team have been necessary to develop the procedures for semantic analysis and structuring of the observables content of the worksheets. Frequent rechecks with the domain experts have been necessary to keep this activity on target.

Concept model extensions for Body structures and Substances have gone through several iterations with suggested enhancements including: 1) do not replicate scientific data in SNOMED CT which will then require curation and synchronization and 2) bind SNOMED CT extensions for molecular to authoritative scientific reference ontologies.

The concept model for Observables is complex and subject to non-reproducible application. As we have encountered new types of Observables in semantic analysis of the worksheets, we have

required consultation with the Observables project team and from those discussions we have developed templates for inclusion in the editorial guide for Observables.

6.1.3 Evaluation of Design

6.1.3.1 Exceptions and Problems

- The IHTSDO-Regenstrief agreement of 2008 severely limits publication and implementation of new and existing Observables content for SNOMED CT compliant electronic databases. However the semantic nature of data in the CAP cancer protocol worksheets clearly calls for substantial content <<363787002 | Observable entity (observable entity) |. Initial publication will occur in Nebraska Lexicon© extension formatting to maintain adherence to the IHTSDO-Regenstrief agreement.

6.1.3.2 Design Strengths

- Binding of SNOMED CT genes, proteins and cellular processes to recognized and authoritative scientific reference standards
- Detailed and fully defined structured datasets for anatomic and molecular pathology in cancer
- Employment of agreed harmonized model for Observable entities and publication of content for IHTSDO and LOINC communities of use
- Publication of harmonized observables content suitable for deployment by SNOMED CT or LOINC communities of use

6.1.3.3 Design Weakness

- Reproducibility of the harmonized model for Observables is weak, requiring central control of authoring and template development for classes of AP/MP observables

6.1.3.4 Design Risks

Description of risk	Importance	Mitigation plan

7 Recommendation

7.1.1 Detailed design final specification

See chapter 6

7.1.2 Iteration plan

We plan for continuous quality improvement consisting of content review and revision of modelling with each CAP work sheet. iPaLM SIG and the Observables project team will be the primary review agents.

8 Quality program criteria

8.1 Quality metrics

8.1.1 Quality metric 1

Component	Characteristic and Description		Metric	Target	Result
Logic definitions of concepts in Observables	**Char:**	sufficiently defined	- Proportion sufficiently defined - Numerator: count of those defined. - Denominator: count of all concepts under <concept nnnnn>	95%	
	Descr:	Concept logic definitions should be “defined” not “primitive”			

8.1.2 Quality metric 2

Component	Characteristic and Description		Metric	Target	Result
Fully specified names in <domain>	**Char:**	Adherence to terming guidelines	- Proportion meeting guidelines, based on manual review	100%	
	Descr:	The fully specified name should adhere to terming guidelines listed in the editorial guide, sections <list sections>			

8.2 Use case scenarios

8.2.1 Scenario One: EHR deployment

8.2.1.1 Expected Setting: Epic electronic health record

8.2.1.2 Data capture requirement

Clinical pathologist will electronically ‘fill out’ the CAP cancer worksheet in COPATH® tailored to employ structured data recording. Structured AP and MP data will pass by HL7 interface to the Epic® EHR where it will be stored in context as structured flowsheet data reporting the pathologists observations.

8.2.1.3 Data retrieval requirement

The clinical oncologist and the primary care physician will access the AP and MP data in context in the EHR for record review and decision making.

8.2.2 Scenario Two: Clinical research test case

8.2.2.1 Expected Setting: Department of pathology

8.2.2.2 Data capture requirement

Clinical pathologist will electronically 'fill out' the CAP cancer worksheet in COPATH® tailored to employ structured data recording. Structured AP and MP data will pass by HL7 interface to the pathology data warehouse where it will reside along with other clinical data

8.2.2.3 Data retrieval requirement

The clinical pathologist has developed an IRB approved research project to assess the utility of 104785321000004109|Status of tumor budding from excised carcinoma (observable entity)| in the staging of colorectal carcinoma. 100 retrospective cases will be encoded using the CAP work sheet data for this cancer. The research data warehouse will serve to retrieve relevant case material for the pathologist' study.

8.2.3 Scenario Three: Research data warehouse deployment

8.2.3.1 Expected Setting: Tissue biobank within enterprise data warehouse

8.2.3.2 Data capture requirement

Clinical pathologist will electronically 'fill out' the CAP cancer worksheet in COPATH® tailored to employ structured data recording. Structured AP and MP data will pass by HL7 to the Neo4j graph database to support the tissue biobank. HL7 messages will also be processed and organized into staging data for i2b2. Staging data will be periodically extracted and ported to the i2b2 data warehouse for clinical research purposes.

8.2.3.3 Data retrieval environment

Query functionality of the graph database will be used to retrieve research test cases as proposed by the consultant pathology team. Query functionality of i2b2 will be employed to select clinical cancer cases from the data warehouse for research projects to be specified by the Greater Plains Collaborative for clinical research.

8.2.4 Scenario Four: International collaboration on cancer epidemiology

8.2.4.1 Data capture requirement

Expanded SNOMED CT and LOINC metadata employing the AP/MP terminology will be developed for distribution to the international community of use of i2b2. Cancer case data recorded with the COPATH® software and ported to the i2b2 data warehouse will be stored on this i2b2 platform.

8.2.4.2 Data retrieval requirement

ICCR collaborators define cancer research questions which can answered with the synoptic data reported to i2b2. A project which will query AP/MP data sets across i2b2 platforms in two or more countries will demonstrate the ability to retrieve multi-national data sets for clinical research.

9 Project Resource Estimates

Estimate project size; Forecast project velocity and duration
Evaluate risks; Establish costs and articulate value; Plan deployment; Outline project lifecycle

9.1 Scope of construction phase

Based upon historical analysis of pathology reports at UNMC, 95% of synoptic pathology reports address 35 different CAP work sheets. These 35 AP work sheets along with relevant biomarker work sheets (approximately 10 in number) will constitute the development working set. The project is budgeted for 24 months to completion and employs personnel with skill sets as follows:

- Consultant terminologist 0.5 FTE X 24 months
- Certified Implementation specialist 0.5 FTE X 24 months
- Anatomic pathology system analyst 0.25 FTE X 24 months
- EHR build analyst 0.25 FTE X 6 months; 0.05 FTE X 18 months
- EHR Interface analyst 0.10 FTE X 12 months
- Database programmer/analyst 0.50 FTE X 1 year
- Consultant physicians; ICCR pathology, CAP pathology, UNMC oncology 0.05 FTE X 8 X 24 months

Work packages:

- Project management
- Nebraska publication website with integrated browser and publishing security and tracking
- Semantic analysis of CAP synoptic worksheets; terminology modelling and valueset development
- ICCR collaboration and consensus
- Observables project collaboration and consensus
- COPATH anatomic pathology work sheet build and implementation
- HL7 V 2.X outbound interface tailoring and implementation
- EPIC system analysis; inbound interface tailoring and implementation
- Tissue biobank; inbound interface tailoring and implementation; tools development
- Clinician liaison for use case assessment and documentation
- Terminology documentation and publication
- HL7 CDA implementation guide for synoptic reports
- HL7 FHIR mapping templates from SNOMED CT Observables to FHIR datasets for AP and MP; trial implementation in Epic

9.2 Projection of remaining overall project resource requirements

9.2.1 Expected project resource requirement category

Major; Project manager assigned at Nebraska site

9.2.2 Expected project impact and benefit

Updated view of impact and benefit, organized by stage if the project is to be staged

9.2.3 Indicative resource estimates for construction, transition and maintenance:

Construction phase:

35 cancer worksheets account for 95% of cancer case volume at UNMC

Construction and transition phase: 600-800 new concepts; approximately 350-450 Observables

Maintenance phase: 500-1000 new ‘frequent usage’ concept requests in 1st 3 years

10 Bibliography

[1]Slack, J.M.W. Genes-A Very Short Introduction. Oxford University Press 2014.

[2]Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). Molecular Biology of the Cell (Fourth ed.). New York: Garland Science. ISBN 978-0-8153-3218-3.

[3] College of American Pathologists. Cancer Protocol Templates. http://www.cap.org/web/oracle/webcenter/portalapp/pagehierarchy/cancer_protocol_templates

[4] Batts KP, Jurgen L. An Update on Terminology and Reporting. American Journal of Surgical Pathology: December 1995.

[5] https://app1.unmc.edu/pathmicro/umls/logon.cfm?r=Publication_Files_20170331.zip

[6]Campbell WS, Campbell JR, West WW, McClay JC, Hinrichs SH. Semantic analysis of SNOMED CT for a post-coordinated database of histopathology findings. J Am Med Inform Assoc. 2014 Sep-Oct;21(5):885-92. doi: 10.1136/amiajnl-2013-002456. Epub 2014 May 15.

[7] Campbell WS, Karlsson D, Vreeman DJ, Lazenby AJ, Talmon GA, Campbell JR. A computable pathology report for precision medicine: Extending and observables ontology unifying SNOMED CT and LOINC. JAMIA; 0(0), 2017, 1-8. https://academic.oup.com/jamia/article/doi/10.1093/jamia/ocx097/4157681/A-computable-pathology-report-for-precision?guestAccessKey=ee8ed0e2-8e9f-44bd-a467-bfbcb3e3a2c0

[8]Campbell JR, Talmon G, Cushman-Vokoun A, Karlsson D, Scott Campbell W. An Extended SNOMED CT Concept Model for Observations in Molecular Genetics. AMIA Annu Symp Proc. 2017 Feb 10;2016:352-360.

Appendix A First review

12/18/2016
Review comments by Matthew Cordell
“Document meets most of the completion criteria, however I don't think the problem/solution is clear to those unfamiliar with the topic.
I think the document would benefit from at least a single example of the content that is problematic and how bits relate. LOINC/SNOMED CT/Protocols/Genetics are all mentioned, but not clear how the components relate. I have some idea, particularly after seeing the presentation in Wellington.
I'd suggest either a single example of current problem and proposed solution (to illustrate to readers the issue). Or include the presentation as an appendix/supplementary document (since it goes into a lot more detail). Currently the problem and solution are discussed in an almost abstract sense - at least for those unfamiliar with the topic/issue.
Otherwise the document is good, and no additional changes required.”