SNOMED-CT

Petr Křemen

Ontologies and Semantic Web
Winter 2023

(Have You noticed … ?)

- COVID vaccination certificates refer to
  - code of COVID-19 disease/agens
    - 840533007
  - code of COVID-19 vaccine
    - 1119349007

Vaccination certificate
Certifikát o provedené vakcinaci

Disease or agent targeted *(Cílená nemoc nebo agens)*
SARS-CoV-2 (ICD 11 XN109, SNOMED CT 840533007)

Vaccine/prophylaxis *(Vakcína/Profylaxe)*
mRNA vakcína proti onemocnění COVID-19
COVID-19 mRNA Vaccine, Severe acute respiratory syndrome coronavirus 2 mRNA only vaccine product(SNOMED CT 1119349007)

Vaccine medicinal product *(Vakcína)*
Comirnaty

Vaccine marketing authorisation holder or manufacturer
(Výrobce nebo držitel rozhodnutí o registraci vakcíny)
BioNTech Manufacturing GmbH

Number in a series of vaccinations/doses and the overall number of doses in the series
(Pořadové číslo dávky/počet dávek)
1/2

Date of vaccination *(Datum vakcinace)*
2021-05-25

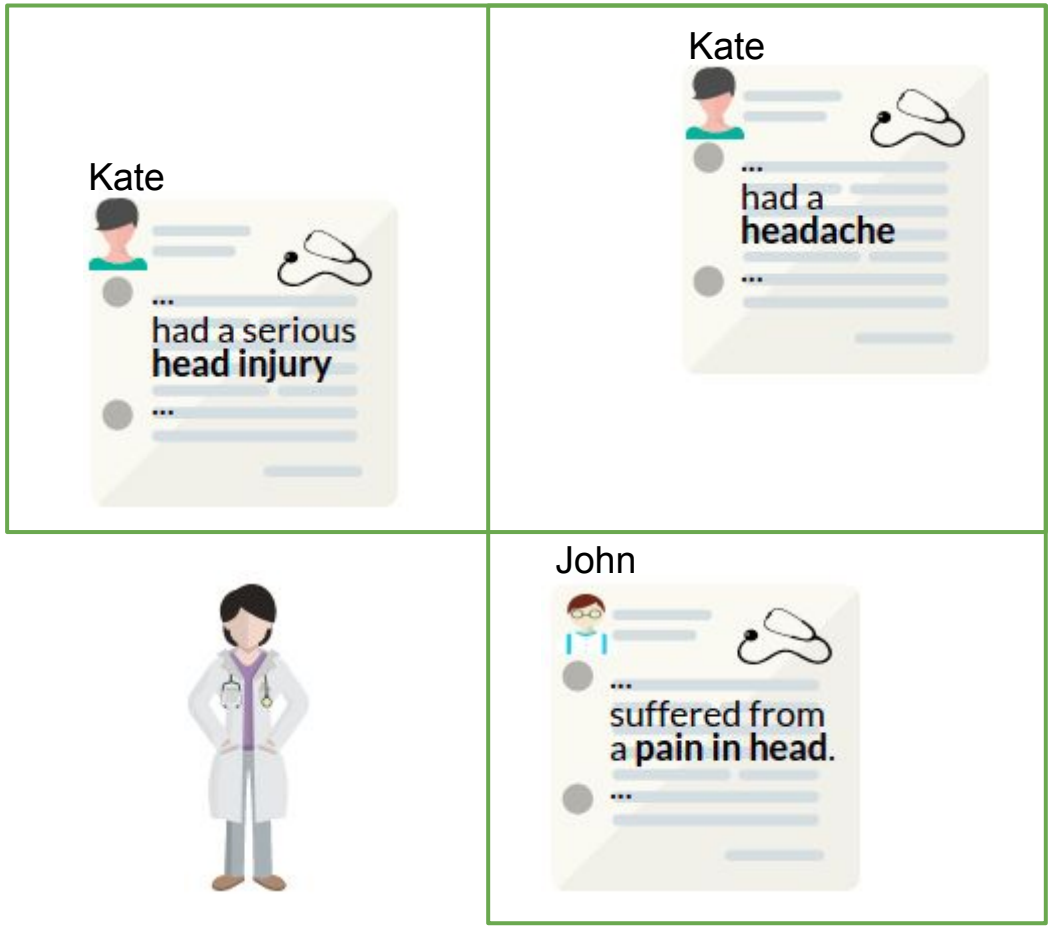
Member State of vaccination *(Členský stát)*
CZ

Certificate issuer*(Vydavatel certifikátu)*
Ministry of Health of the Czech Republic / Ministerstvo zdravotnictví České republiky

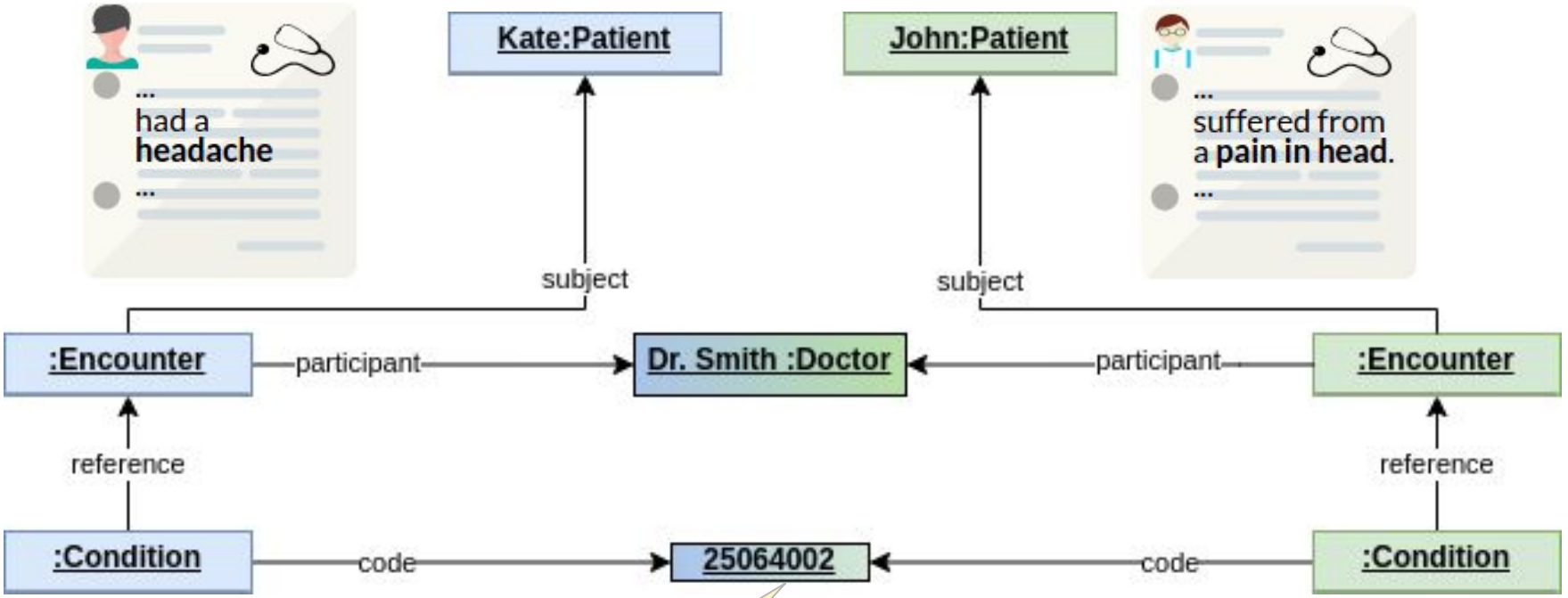
https://koronavirus.mzcr.cz/wp-content/uploads/2021/08/EU-Digital-COVID-Certificate.pdf

Healthcare Data Integration

- Which patients had some problems with head manifested by a pain?
- What is the Kate's history of problems related to head?



Which patients had some problems with head manifested by a pain?

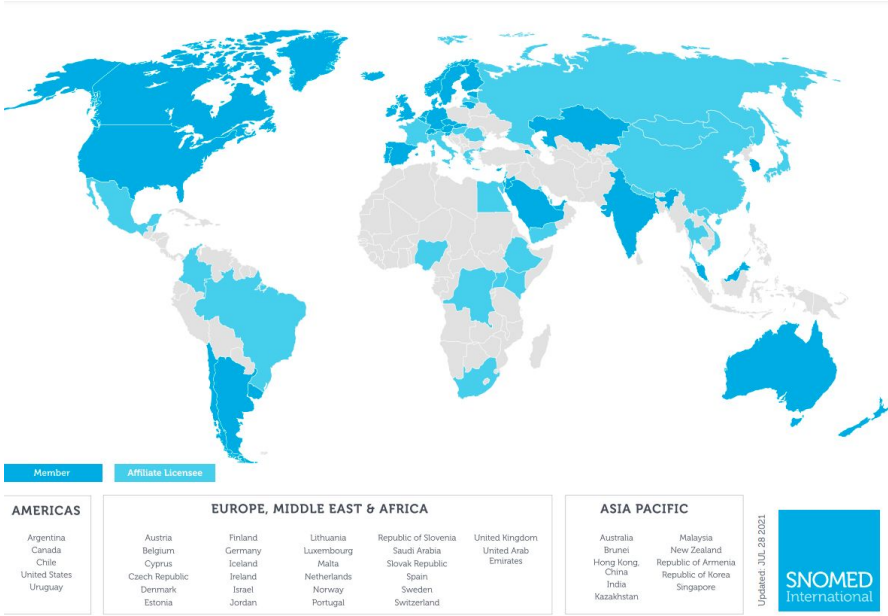


Headache (finding)

- SCTID: 25064002
- synonym ‘Headache’
- synonym ‘Pain in head’
- ...

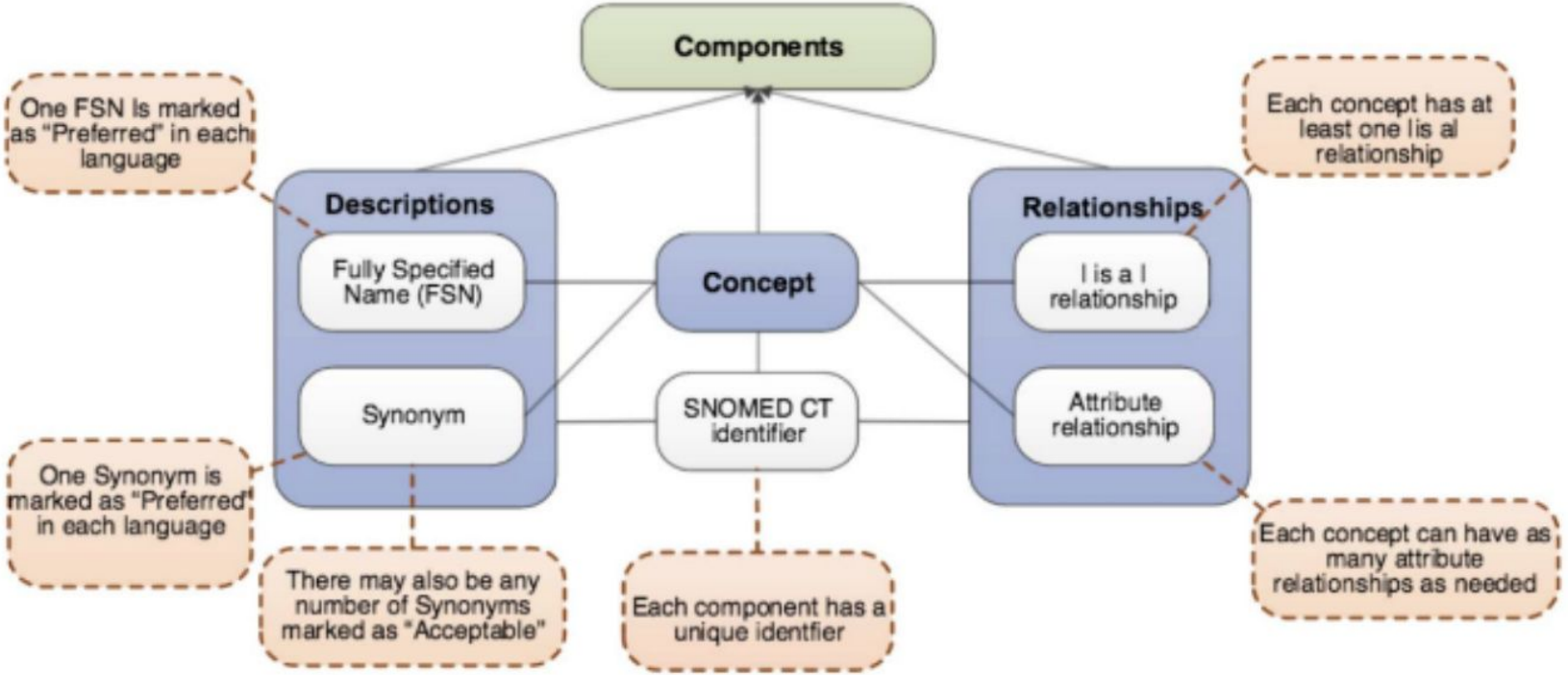
About SNOMED-CT

- **S**ystematized **NO**menclature of **MED**icine – **C**linical **T**erms
- multiple inheritance clinical terminology
- developed, curated and distributed by SNOMED International
  - 40 member states (including Czech Republic)
- as of july 2022 - international edition
  - 350k+ clinical concepts



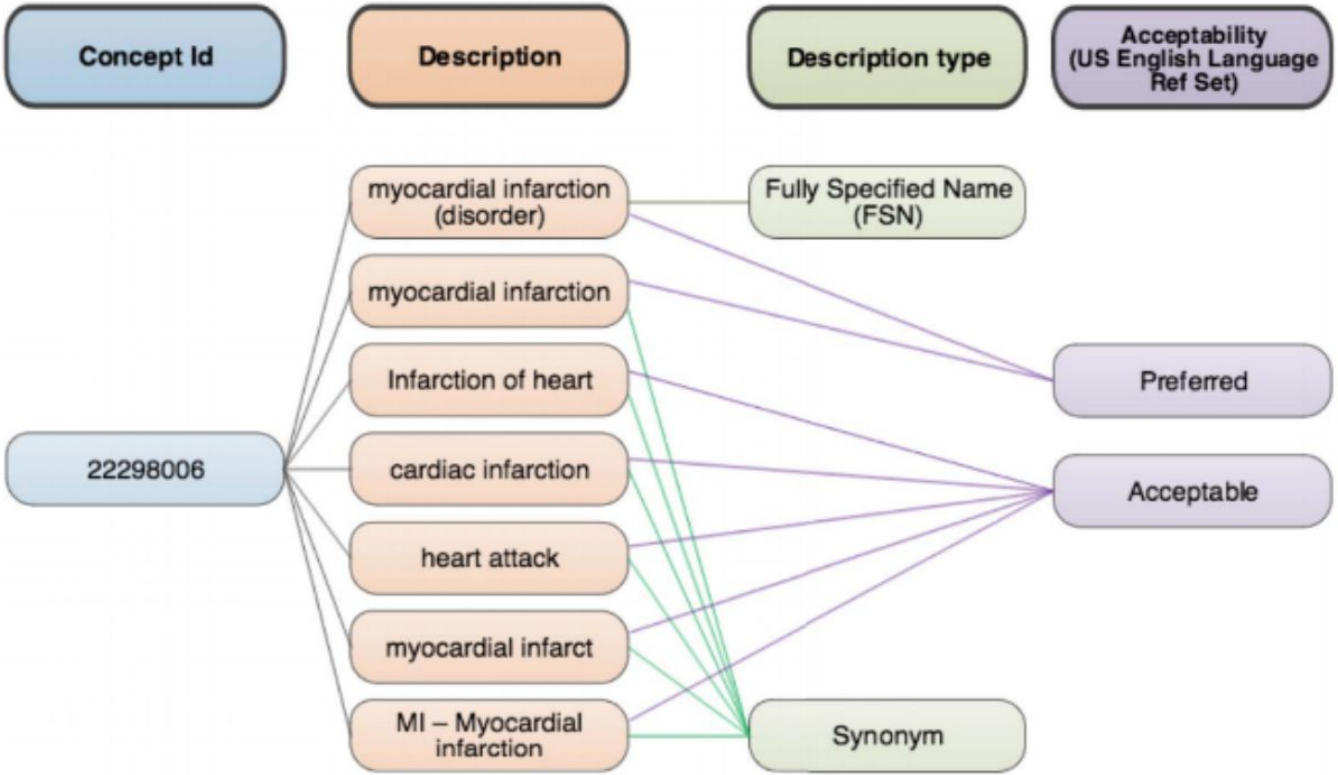
https://www.snomed.org/SNOMED/media/SNOMED/members-flag-header-images/Member%20Map/SNOMED_International_Member_Map_07282021_outlines.pdf?ext=.pdf

SNOMED-CT components



https://www.snomed.org/snomed-ct/five-step-briefing

SNOMED-CT descriptions



https://www.snomed.org/snomed-ct/five-step-briefing

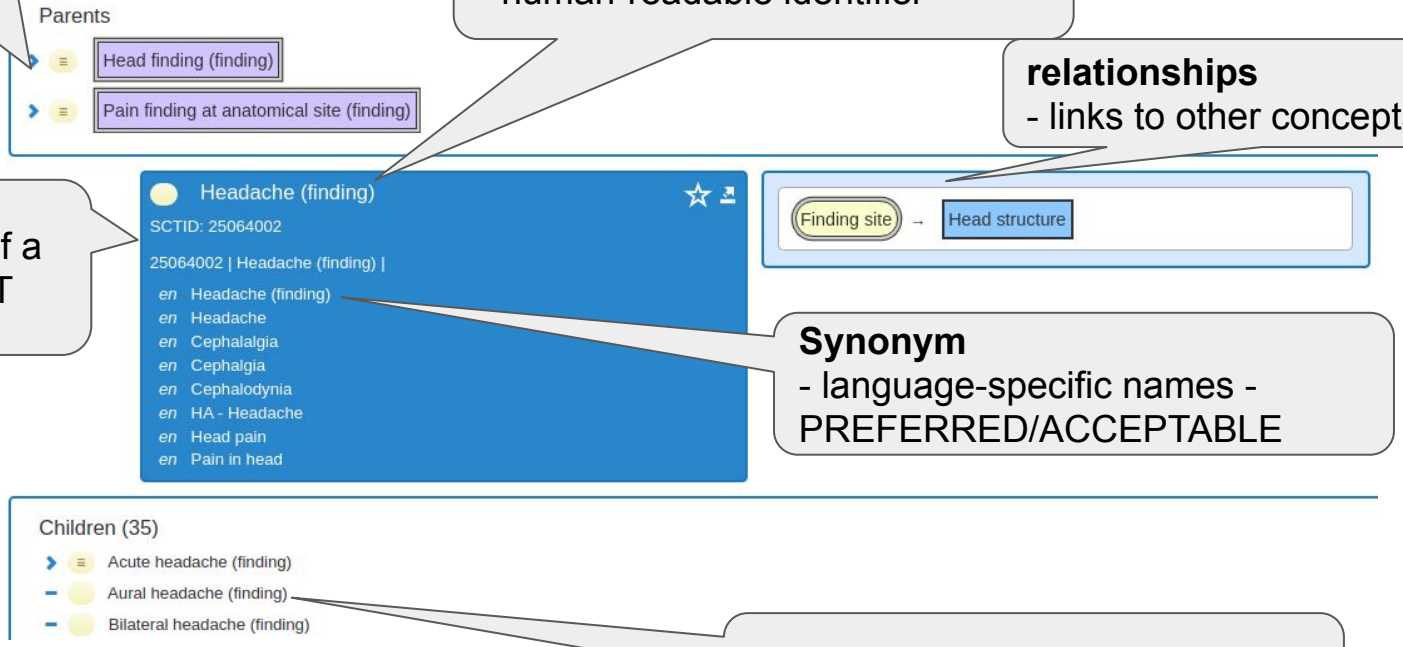
SNOMED-CT concept structure

super concepts
(generalizations)

FSN (fully specified name)
- human-readable identifier

relationships
- links to other concepts

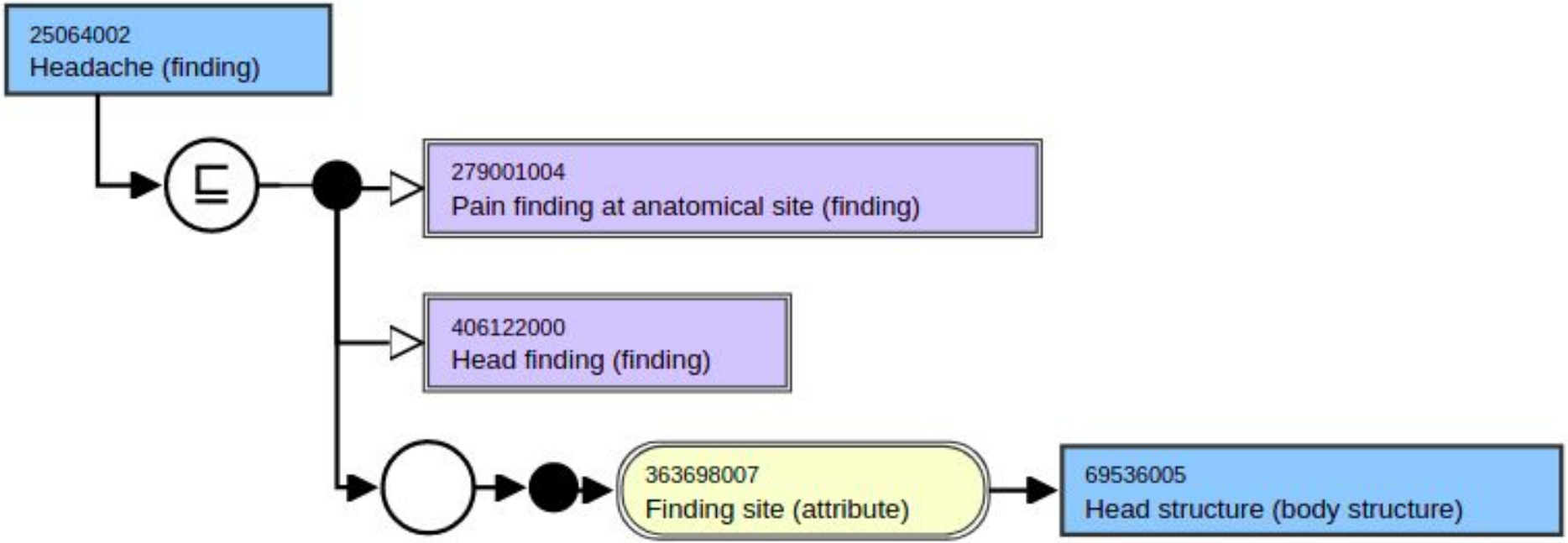
SCTID:
- unique ID of a SNOMED-CT concept



Synonym
- language-specific names - PREFERRED/ACCEPTABLE

sub concepts
(specializations)

SNOMED-CT concept diagram



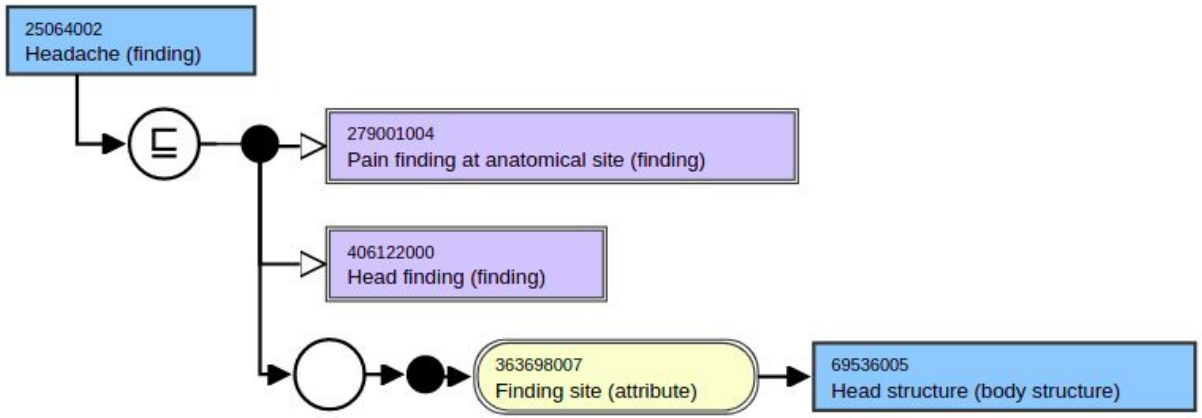
SNOMED-CT in OWL

“Headache (finding)” ⊑ ∃ FindingSite . HeadStructure

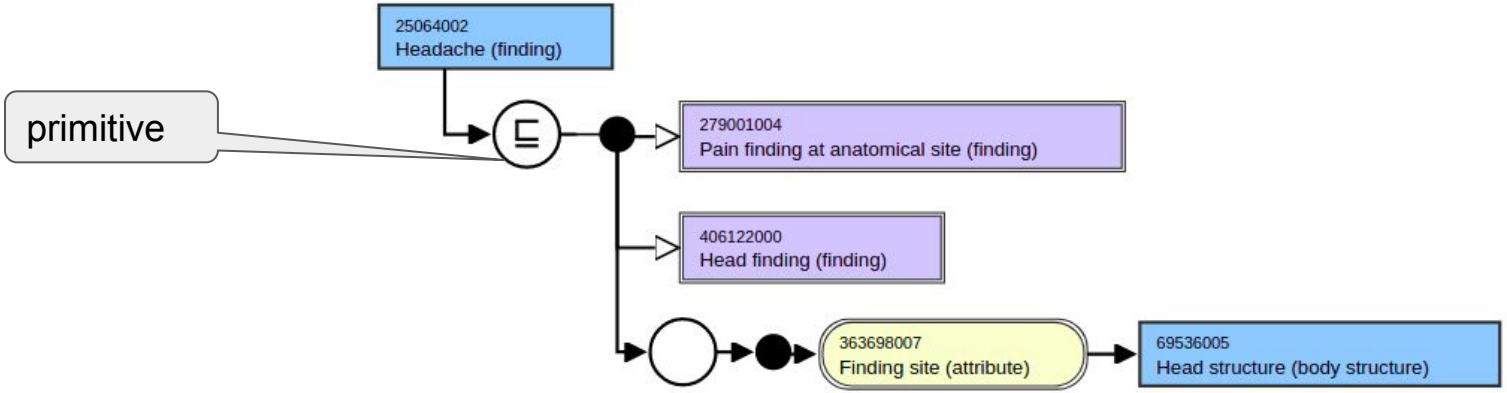
“Headache (finding)” ⊑ “Head finding (finding)”

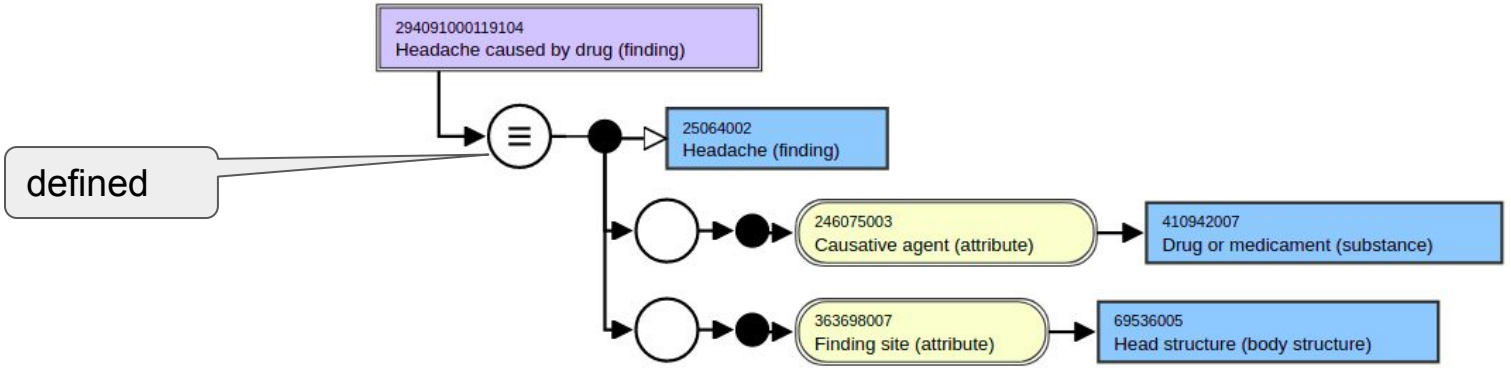
“Headache (finding)” ⊑ “Pain finding at anatomical site (finding)”

...



Defined vs. Primitive Concepts





Description Logic Intermezzo

- SNOMED-CT - expressible in description logics EL++
- EL++ is expressible in the OWL 2 EL profile
- EL++ reasoners are used to classify SNOMED-CT

  (e.g. the ELK reasoner)

- reasoning in EL++ is polynomial and does not require satisfiability checking

Description Logic EL++

Name	Syntax	Semantics
top	$\top$	$\Delta^{\mathcal{I}}$
bottom	$\bot$	$\emptyset$
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
concrete domain	$p(f_1,\ldots,f_k)$ for $p \in \mathcal{P}^{\mathcal{D}_j}$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y_1,\ldots,y_k \in \Delta^{\mathcal{D}_j} : f_i^{\mathcal{I}}(x) = y_i \text{ for } 1 \le i \le k \wedge (y_1,\ldots,y_k) \in p^{\mathcal{D}_j}\}$
GCI	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
RI	$r_1 \circ \cdots \circ r_k \sqsubseteq r$	$r_1^{\mathcal{I}} \circ \cdots \circ r_k^{\mathcal{I}} \subseteq r^{\mathcal{I}}$

Normalized TBox

- $A \sqsubseteq B$
- $A_1 \sqcap A_2 \sqsubseteq B$
- $A \sqsubseteq \exists r.B$
- $\exists r.A \sqsubseteq B$

Normalization takes a general TBox and creates a normalized TBox (preserving classification results):

$\exists r.A \sqcap \exists r.\exists s.A \sqsubseteq A \sqcap B$

can be normalized to

$\exists r.A \sqsubseteq B_1$ $\quad B_1 \sqcap B_2 \sqsubseteq B_0$ $\quad \exists s.A \sqsubseteq B_3$ $\quad \exists r.B_3 \sqsubseteq B_2$ $\quad B_0 \sqsubseteq A$ $\quad B_0 \sqsubseteq B$

Reasoning in EL++

rule	if	then	note
CR1	{}	$A \sqsubseteq A$	
CR2	{}	$A \sqsubseteq \top$	
CR3	$\{A_1 \sqsubseteq A_2, A_2 \sqsubseteq A_3\}$	$A_1 \sqsubseteq A_3$	
CR4	$\{A \sqsubseteq A_1, A \sqsubseteq A_2, A_1 \sqcap A_2 \sqsubseteq B\}$	$A \sqsubseteq B$	convexity
CR5	$\{A \sqsubseteq \exists r.A_1, A_1 \sqsubseteq B_1, \exists r.B_1 \sqsubseteq B\}$	$A \sqsubseteq B$	

Reasoning in EL++ - Example

$\{ A \sqsubseteq \exists r.A, \quad \exists r.B \sqsubseteq B_1, \quad T \sqsubseteq B, \quad A \sqsubseteq B_2, \quad B_1 \sqcap B_2 \sqsubseteq C \}$

$\models \quad A \sqsubseteq T \qquad (CR2)$

$\models \quad A \sqsubseteq B \qquad (CR3)$

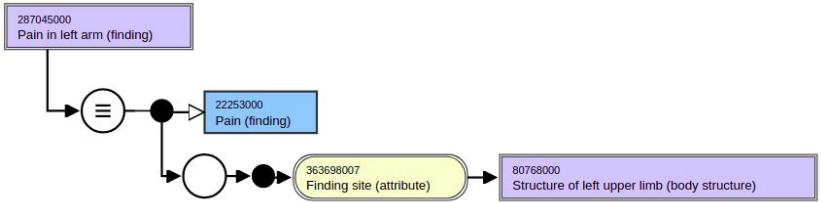
$\models \quad A \sqsubseteq B_1 \qquad (CR4)$

$\models \quad A \sqsubseteq C \qquad (CR5)$

End of intermezzo.

Classification of SNOMED-CT

- for defined concepts infers their children
- `287045000 | Pain in left arm |`
- before classification - 0 children
- after classification - 11 direct children (26 total)



- ≡ Chronic pain of left upper limb (finding)
- ≡ Joint pain in left hand (finding)
- ≡ Pain of amputation stump of left upper limb (finding)
- ≡ Pain of left acromioclavicular joint (finding)
- ≡ Pain of left forearm (finding)
- ≡ Pain of left shoulder blade (finding)
- ≡ Pain of left sternoclavicular joint (finding)
- ≡ Pain of left upper arm (finding)
- ≡ Pain of left upper limb co-occurrent and due to ischemia (disorder)
- ≡ Repetitive motion disorder of left hand (disorder)
- ≡ Repetitive motion disorder of left shoulder (disorder)

Reference Sets (RefSets)

- a subset of components (concepts, descriptions or relationships), optionally equipped with additional attributes
- use-cases
  - Ordered lists of components
  - Sets of associations between components
  - Mapping between SNOMED CT concepts and other systems codes, classifications, or knowledge resources.
  - Language translations to concepts

Reference Sets (RefSets)

Description file
(International release)

conceptId	id	typeId	term
22298006	751689013	900000000000003001 | Fully specified name |	Myocardial infarction (disorder)
22298006	37436014	900000000000013009 | Synonym |	Myocardial infarction
22298006	37442013	900000000000013009 | Synonym |	Cardiac infarction
22298006	37443015	900000000000013009 | Synonym |	Heart attack
22298006	37441018	900000000000013009 | Synonym |	Infarction of heart
22298006	1784872019	900000000000013009 | Synonym |	MI - Myocardial infarction
22298006	1784873012	900000000000013009 | Synonym |	Myocardial infarct

900000000000509007 \| United States of America English language reference set\|

refsetId	referencedComponentId	acceptabilityId
900000000000509007	751689013	900000000000548007 | Preferred |
900000000000509007	37436014	900000000000548007 | Preferred |
900000000000509007	37442013	900000000000549004 | Acceptable|
900000000000509007	37443015	900000000000549004 | Acceptable|
900000000000509007	37441018	900000000000549004 | Acceptable|
900000000000509007	1784872019	900000000000549004 | Acceptable|
900000000000509007	1784873012	900000000000549004 | Acceptable|

https://confluence.ihtsdotools.org/display/DOCRFSPG/5.8.+Language+Reference+Set

Content

- combined metadata with data (like in RDF)
- **SNOMED CT Model Component** is healthcare agnostic

Expression Constraint Language (ECL)

- language to express concept model/to query SNOMED-CT

Symbol	Name	Version	Notes
|	Pipe	1.0	Used on either side of a concept's term for human readability
*	Any	1.0	Retrieves all concepts in the substrate
^	Member of	1.0	Retrieves all (active) members of a reference set identified by a specified reference set concept
<	Descendant of	1.0	Retrieves all descendants (subtypes) of the specified concept *excluding* the concept itself
<<	Descendant or self of	1.0	Retrieves all descendants (subtypes) of the specified concept *including* the concept itself
<!	Child of	1.1	Retrieves all children (immediate subtypes) of the specified concept *excluding* the concept itself
<<!	Child or self of	1.4	Retrieves all children (immediate subtypes) of the specified concept *including* the concept itself
>	Ancestor of	1.0	Retrieves all ancestors (supertypes) of the specified concept *excluding* the concept itself
>>	Ancestor or self of	1.0	Retrieves all ancestors (supertypes) of the specified concept *including* the concept itself
>!	Parent of	1.1	Retrieves all parents (immediate supertypes) of the specified concept *excluding* the concept itself
>>!	Parent or self of	1.4	Retrieves all parents (immediate supertypes) of the specified concept *including* the concept itself
AND	Conjunction	1.0	Retrieves the intersection of the results of each sub-expressions
OR	Disjunction	1.0	Retrieves the union of the results of each sub-expressions
MINUS	Exclusion	1.0	Retrieves the members of the first expression and excludes the members returned by the second expression
:	Refinement	1.0	Used before one or more attribute-value pairs to refine the set of concepts retrieved
[1..3]	Cardinality	1.0	Used to indicate the minimum and maximum number of occurrences of attributes or relationship groups
R	Reverse flag	1.0	Retrieves the set of attribute values (i.e. destination concepts) of a specified attribute for a specified set of concepts
.	Dot notation	1.1	Retrieves the set of attribute values (i.e. destination concepts) of a specified attribute for a specified set of concepts
/* */	Comment	1.1	Allows comments to be added within the text of an expression constraint
{{ }}	Description filter	1.5	Filters the result set, by matching only on concepts which have a description with a matching term, language, type, dialect and/or acceptability
{{ D }}	Description filter	1.5	Filters the result set, by matching only on concepts which have a description with a matching term, language, type, dialect and/or acceptability
{{ C }}	Concept filter	1.6	Filters the result set based on the definition status, module, effectiveTime and active status of each concept

https://confluence.ihtsdotools.org/display/DOCECL/Appendix+D+-+ECL+Quick+reference

Expression Constraint Language

- What are the types of pain that can occur in the left arm (including the concept itself)?

  ```
  << 287045000 | Pain in left arm |
  ```

- What are the types of pain that can occur in the left arm?

  ```
  < 287045000 | Pain in left arm |
  ```

- What are the finding sites of types of pain in left arm?

  ```
  << 287045000 | Pain in left arm | . 363698007 | Finding Site |
  ```

Expression Constraint Language

- What can occur on head?

```
< * : 363698007 | Finding Site | = 69536005 | Head |
```

- What can occur on head or its part?

```
< * : 363698007 | Finding Site | = << 69536005 | Head |
```

- What are the types of paralytic syndromes caused by a cardiovascular finding

```
<< 29426003 | Paralytic syndrome |: << 42752001 | Due to | = << 106063007 |
Cardiovascular finding|
```

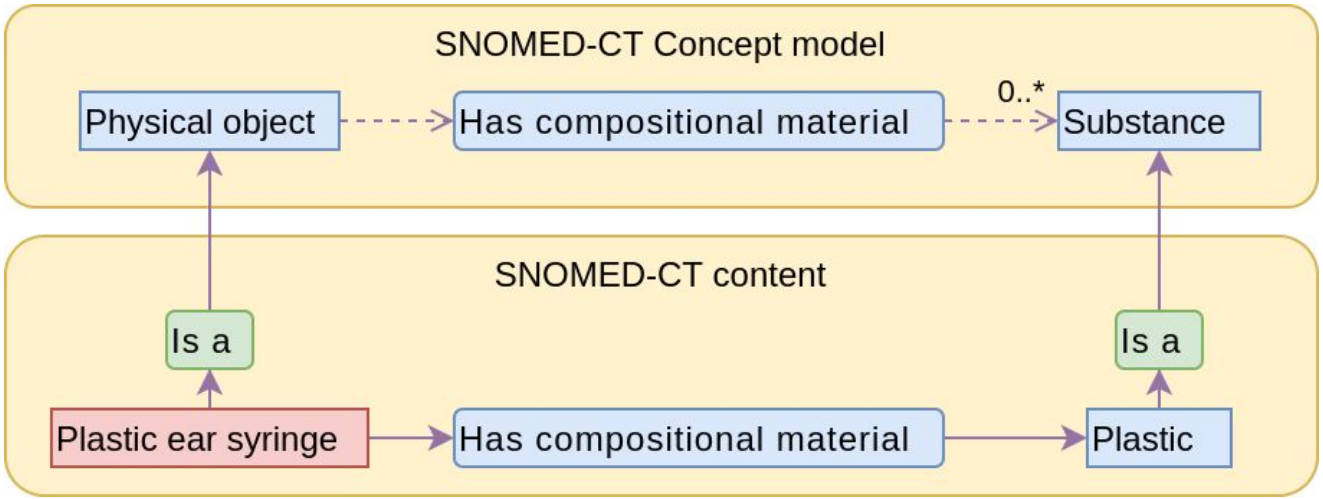
- Which concepts belong to the CTV3 map

```
< ^ 900000000000497000 | CTV3 simple map reference set |
```

Machine readable concept model (MRMC)

id	effectiveTime	active	moduleId	refsetId	referencedComponentId
b3a823ef-aba5-4cf5-be9a-6c9808ad58b9	20200731	1	900000000000012004	723561005	840560000

domainId	grouped	attributeCardinality	attributeInGroupCardinality	ruleStrengthId	contentTypeId
260787004	0	0..*	0..0	723597001	723596005



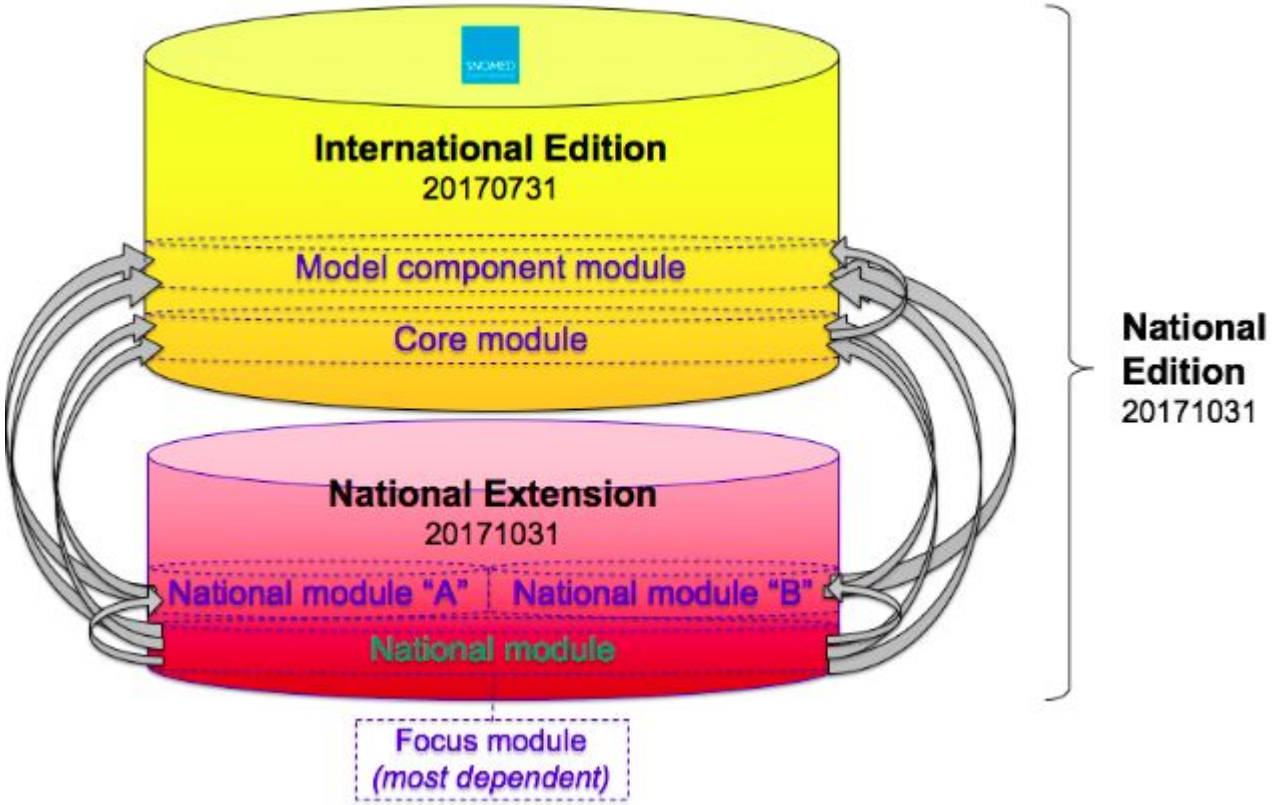
Distribution

- set of files in RF2 format
- snapshot
  - currently valid version
- delta
  - changes w.r.t last snapshot
- full
  - snapshot and all deltas

```
SnomedCT_InternationalRF2_PRODUCTION_20210131T120000Z.zip
Delta
      ...
Full
      ...
Snapshot
      Terminology
        sct2_Concept_Snapshot_INT_20210131.txt
        sct2_Description_Snapshot-en_INT_20210131.txt
        sct2_Identifier_Snapshot_INT_20210131.txt
        sct2_Relationship_Snapshot_INT_20210131.txt
        sct2_sRefset_OWLExpressionSnapshot_INT_20210131.txt
        sct2_StatedRelationship_Snapshot_INT_20210131.txt
        sct2_TextDefinition_Snapshot-en_INT_20210131.txt
      Refset
        Content
            der2_cRefset_AssociationSnapshot_INT_20210131.txt
            der2_cRefset_AttributeValueSnapshot_INT_20210131.txt
            der2_Refset_SimpleSnapshot_INT_20210131.txt
        Language
            der2_cRefset_LanguageSnapshot-en_INT_20210131.txt
        Map
            der2_iissccRefset_ExtendedMapSnapshot_INT_20210131.txt
            der2_sRefset_SimpleMapSnapshot_INT_20210131.txt
        Metadata
            der2_cciRefset_RefsetDescriptorSnapshot_INT_20210131.txt
            der2_ciRefset_DescriptionTypeSnapshot_INT_20210131.txt
            der2_cissccRefset_MRCMAttributeDomainSnapshot_INT_20210131.txt
            der2_cRefset_MRCMModuleScopeSnapshot_INT_20210131.txt
            der2_ssccRefset_MRCMAttributeRangeSnapshot_INT_20210131.txt
            der2_ssRefset_ModuleDependencySnapshot_INT_20210131.txt
            der2_sssssssRefset_MRCMDomainSnapshot_INT_20210131.txt
```

Editions and Extensions

- each country can create its own extension to the international edition
  - custom concepts
  - custom refsets
  - translations



https://confluence.ihtsdotools.org/display/DOCEXTPG/4.4+Editions

Reference

1. *SNOMED International, https://www.snomed.org/, cit. 8.12.2021*
2. *FHIR, profile Condition https://www.hl7.org/fhir/condition.html , cit 8.12.2021*
3. *Franz Baader, Sebastian Brandt, Carsten Lutz: Pushing the EL Envelope. IJCAI 2005: 364-369*