

# Data coding - The role of SNOMED CT in sharing structured data.

## Interoperability in Practice

### 3<sup>rd</sup> February 2014

Presented by  
Edward Cheetham

# Outline

---

- Scope
  - Packaging and distribution
  - Information model overlap
  - Detailed model design and term binding
  
  - Note - assumes some familiarity with SNOMED CT – this presentations concentrates on those features relevant to data sharing/communication
-

# Scope

---

- What to represent using SNOMED CT?
    - “Clinical” vs. Demographic” vs. “Structural”
      - Plenty of boundary cases
    - Analysis – value in having access to SNOMED CT structure/organising principles
    - Publication
      - Value in having access to SNOMED CT publication mechanisms [vs. release cycle]
      - Prior content in SNOMED CT or precursors
-

# Scope

---

- How to represent in SNOMED CT?
  - Reuse of available content versus adding new content for specific activities
    - Clinical observations & procedures
      - Same concepts (or precursor standard representations) used in source systems
      - Probably more stable meaning independent of recording/communication exercise
    - vs.
    - Document types or sections
      - Risk of code reuse obscuring changes to underlying definitions ('personal preferences')
-

# Scope

---

- How to represent in SNOMED CT?
    - “The code / value problem”
  - Nominalised statements
    - Biopsy of right kidney (as .code)
  - Replicating structured data collection
    - Site of biopsy (as .code/question)
    - Right kidney (as .value/answer)
    - Note .code/.value desirable for ‘property observations’ such as lab test results
  - Replicating document sections
    - Family history, past history...
    - Semantics must be preserved for analysis
-

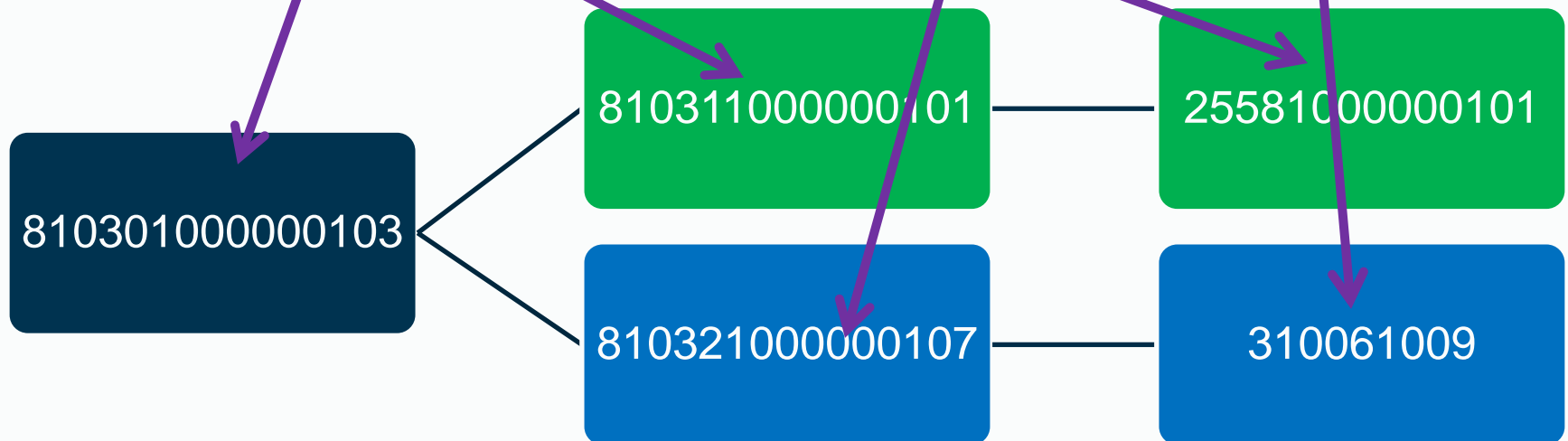
# Scope

---

- How to represent in SNOMED CT?
  - Faithful to published defining model
    - Allergy:causative\_agent=substance
  - Vs.
  - Local extension to defining model
    - Allows combinatorial expressivity where exhaustive enumeration would be problematic
    - Clinical Document Architecture (CDA)  
ClinicalDocument.code
      - Document Type
      - Care Setting
-

# ClinicalDocument.code

Discharge summary report (25581000000101) from a Gynaecology service (310061009) would be represented as  
<code code="{810301000000103:  
810311000000101=25581000000101,810321000000107=310061009"}.  
..>



Risk – this approach is not supportive of equivalence detection  
Note – SCT ‘compositional grammar’ serialisation syntax (SCG)

# Packaging and distribution

---

- Complete national release of SNOMED CT requires processing/integration of many data files
    - ‘Core’ files
      - Concepts [“the ideas”]
      - Relationships [“how they relate to one another”]
      - Descriptions [“how people refer to the ideas”]
    - Other files
      - Subsets / reference sets
      - Classification cross maps
      - Developer tools, stated relationships, history data...
    - International, UK Clinical, UK Drugs
    - 6 monthly release cycle [4 weekly for drugs]
    - RF1 -> RF2 (successor format)
-



# Packaging and distribution

---

- All the files have value. However...
  - A straightforward communication use case should not require all the files to be considered to 'get going'.
  - If SNOMED CT is being used as a source of value sets for a communication specification, there is a case for making the value sets available in a 'simple form'
    - notably isomorphic/comparable with non-SNOMED CT message value sets (e.g. ActStatus)
  - Such an approach is popular, but has caveats
-

# Packaging and distribution

- Value set simplification

```
<?xml version="1.0" encoding="UTF-8" ?>
- <vocabulary name="Family History" type="Realm Concept" setId="55861000000130" version="14" id="1331000000135"
  memberCount="540" status="Active" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
- <concept code="57177007">
  <displayName lang="en" type="PT">Family history with explicit context</displayName>
  <displayName lang="en" type="FSN">Family history with explicit context (situation)</displayName>
  <displayName lang="en" type="SY">FH</displayName>
  <displayName lang="en" type="SY">FH - Family history</displayName>
  <displayName lang="en" type="SY">Family history of</displayName>
</concept>
- <concept code="134439009">
  <displayName lang="en" type="PT">FH: premature coronary heart disease</displayName>
  <displayName lang="en" type="FSN">Family history: premature coronary heart disease (situation)</displayName>
  <displayName lang="en" type="SY">Family history: premature coronary heart disease</displayName>
</concept>
- <concept code="160250007">
  <displayName lang="en" type="PT">No family history of malignancy</displayName>
  <displayName lang="en" type="FSN">No family history of malignancy (situation)</displayName>
</concept>
- <concept code="160252004">
  <displayName lang="en" type="PT">No family history of cardiovascular accident or stroke</displayName>
  <displayName lang="en" type="FSN">No family history of cardiovascular accident or stroke (situation)</displayName>
</concept>
- <concept code="160266009">
  <displayName lang="en" type="PT">No family history of</displayName>
  <displayName lang="en" type="FSN">No family history of clinical finding (situation)</displayName>
  <displayName lang="en" type="SY">No family history of clinical finding</displayName>
  <displayName lang="en" type="SY">No relevant FH: family history</displayName>
```

# Packaging and distribution

---

- By contrast:
  - Earlier specifications used less direct reference to SNOMED CT-based value sets:
  - Name the SubsetOriginalId and require developers to craft their own value set:
    - Subset/RefSet itself (via SSOld->SSId)
    - Descriptions table(s)
    - +/-NHS Realm Description Subset/RefSet table(s)
-

# Packaging and distribution

---

- Create a table (PT) of preferred terms (RF2):
- “For any **conceptId**, give me the **active preferred term** as specified by the **NHS RDR**”

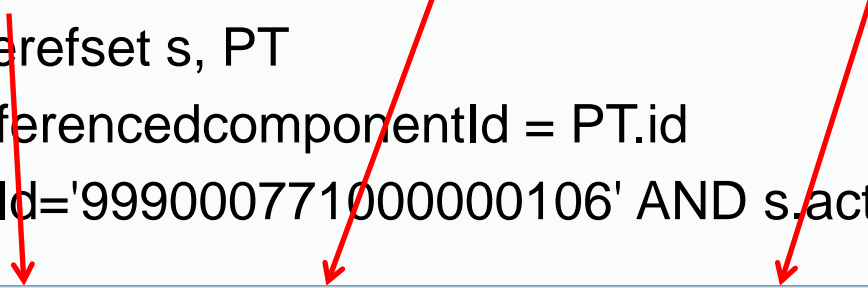
```
SELECT Descriptions.conceptId, Descriptions.term
FROM Descriptions, NHSRDR, Concepts
WHERE
    Descriptions.id = NHSRDR.referencedComponentId
    AND NHSRDR.acceptabilityId = '900000000000548007'
    AND Descriptions.typeId = '900000000000013009'
    AND Descriptions.conceptId = Concepts.ConceptId
    AND Descriptions.active = 1 and NHSRDR.active = 1;
```

---

# Packaging and distribution

- For RefSets (e.g. the 'Family history simple reference set' – refsetId='999000771000000106' ):

```
SELECT s.refsetId, s.referencedcomponentId, PT.term  
FROM simplerefset s, PT  
WHERE s.referencedcomponentId = PT.id  
AND s.refsetId='999000771000000106' AND s.active = 1;
```

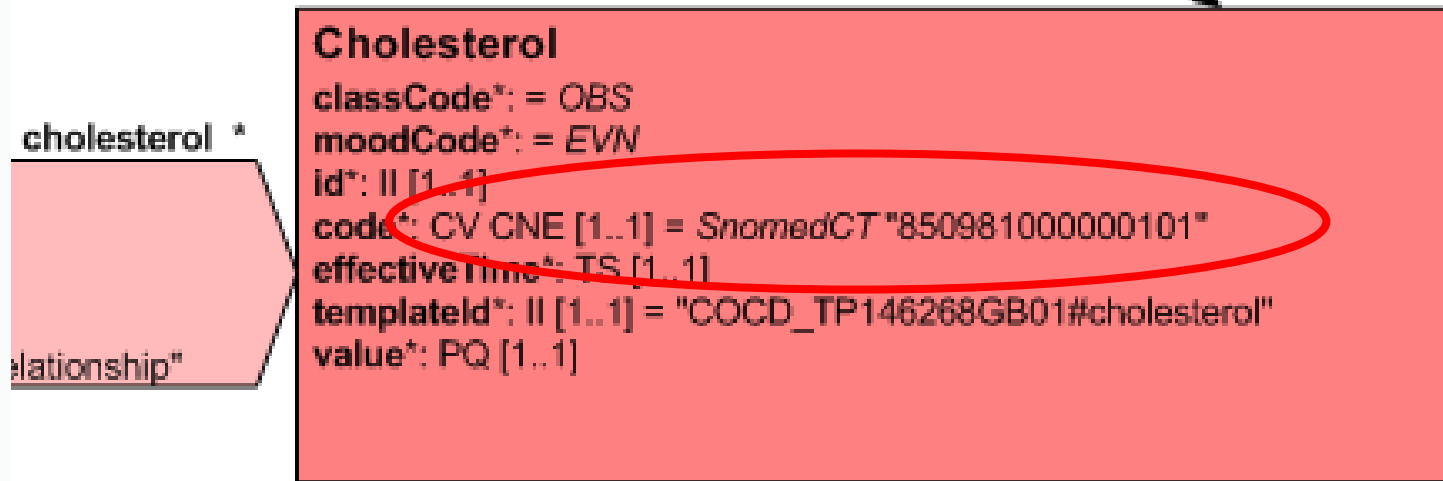


refsetId	referencedcomponentId	term
999000771000000106	430679000	Family history of diabetes mellitus type 2
999000771000000106	443861000	Educating parent of subject
999000771000000106	266891008	FH: CNS disorder
999000771000000106	288249001	Maternal medical problem
999000771000000106	160407004	FH: Eczema
999000771000000106	429958001	Family history of conduction disorder of the heart
999000771000000106	160311006	FH: Obesity
999000771000000106	430568004	Family history of dissection of aorta
999000771000000106	444161008	Maternal history of insulin dependence
999000771000000106	160333008	FH: Suicide
999000771000000106	266970001	Family history of chronic medical disorder
999000771000000106	160421002	FH: Congenital RS anomaly
999000771000000106	160352002	FH: Ear disorder

# Packaging and distribution

---

- Similarly, earlier specifications did not reference single codes directly (indirection via SubsetId).
- Recent specifications have used fixed codes, e.g. within telehealth templates:



- Risk – SNOMED CT changes. Changes to fixed codes can be disruptive to schema/specification stability
-

# Packaging and distribution

---

- Caveats (“All the files have value”):
  - Need for authoritative source of data term quality

```
<concept code="83901003">
<displayName lang="en" type="PT">Sj&#246;gren&apos;s syndrome</display
<displayName lang="en" type="FSN">Sj&#246;gren&apos;s syndrome (disor
<displayName lang="en" type="SY">Sicca syndrome</displayName>
<displayName lang="en" type="SY">Sjogren&apos;s syndrome</displayNam
<displayName lang="en" type="SY">Sjogrens syndrome</displayName>
<displayName lang="en" type="SY">Sj&#246;gren syndrome</displayName>
<displayName lang="en" type="SY">Sj&#246;gren&apos;s disease</display
</concept>
```

- Subsumption testing
-

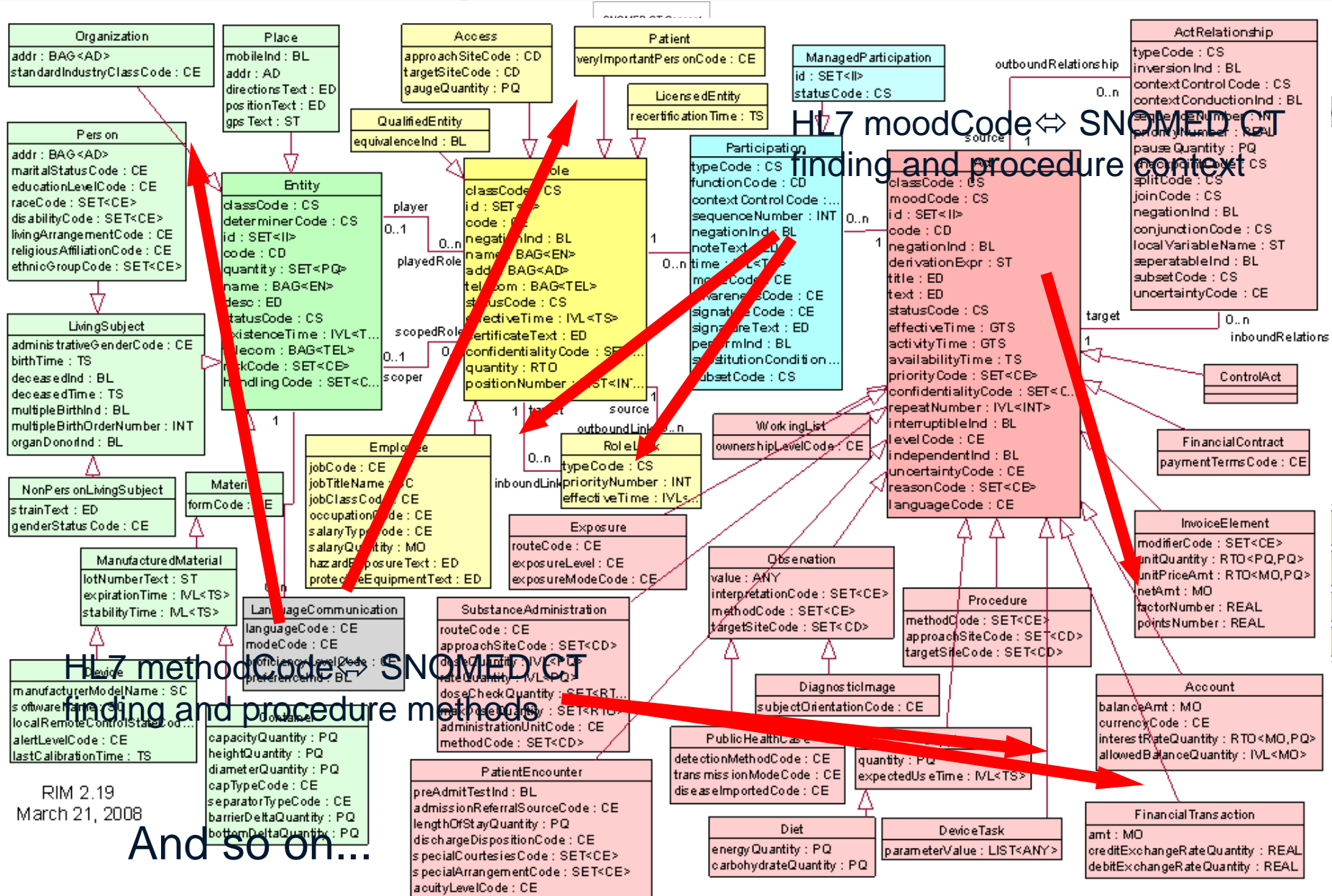
# Information model overlap

---

- HL7 Version 3 Implementation Guide: TermInfo - Using SNOMED CT in CDA R2 Models, Release 1
    - Joint HL7/IHTSDO initiative
    - Initial work 2005-9, now undergoing revision
    - Guidance/rules on how to use SCT and CDA together
    - Extensive document, most relevant section
      - “Guidance on Overlaps between RIM and SNOMED CT Semantics”
-



# Model overlaps



RIM 2.19  
March 21, 2008

And so on...

# TermInfo

---

- Guidance on how to use SCT and CDA together
    - Preferred patterns where semantics overlap
      - Some flexibility, but cannot support all options
    - Specific guidance on how to use SCT in Observation classes (? Use `.code` or `.value` or `.interpretationCode`)
-

# Detailed model design and term binding

---

- Multiple initiatives to improve and standardise clinical model design
    - CIMI, openEHR & ISO13606 archetypes, NHS LRA, ISO TS 13972
  - Recurring theme/notion of *terminology binding*
    - “Linking of information model components to one or more concepts in a terminology”
    - Simplest case – value sets already discussed
-

# Detailed model design and term binding

---

- If models (including communication) rely heavily on SNOMED CT for expressivity, then a constraint language is needed.
  - Various prior attempts
    - Terminfo – extension to SCG
    - IHTSDO Query specification
  - Most recently
    - IHTSDO Terminology binding syntax
    - Primary form based on SCG – transformable to other formalisms:

404684003 | clinical finding |

<< 404684003 | clinical finding |

< 404684003 | Clinical finding |

:( 246454002 | Occurrence | =

(255398004 | Childhood | OR 3658006 | Infancy |)

)

AND (246075003 | Causative agent | = !<409822003 | Bacteria |)

---

# Summary

---

- To meet its requirements, SNOMED CT is a large and changeable product
  - During design – if considering SCT – what and how?
  - For many ‘simple’ communication use cases, not all this complexity is needed
    - Product simplification can help implementers ‘get started’, but ultimately deeper features will need to be handled
  - With HL7, detailed work has been undertaken to help determine the relative contribution of SCT and RIM attributes
  - Where designs need complex/model-based SCT contributions, emergent binding syntax provides a formalism for specification.
-