

SNOMED CT

Coping with concept inactivation

Presenter: Dr Jeremy Rogers, IHTSDO Consultant Terminologist

Date: 10th October 2013, IHTSDO Showcase, Washington DC

Outline

- Recap : how code queries are executed
- The 'inactive content' problem
- Two-part solution
 - Substitutions table
 - 'Role Inclusion Closure' table

Querying in Ye Olde Worlde

(READ2, ICPC, ICD, OPCS...)

- Direct lexical comparison of ConceptID

C1... Other endocrine gland diseases E Respiratory tract
C10.. Diabetes mellitus E02 Plastic operations on nose
C10E4 Unstable type 1 diabetes mellitus E02.5 Reduction rhinoplasty

- Pros

- Doesn't require any external reference table to compute
- Quick to execute

- Cons

- Stuck with a monohierarchy (and duplicate codes)
- Can't move concepts if initially put in the wrong place
- Hierarchy gets full : can't put in the right place!
- etc

New World Querying

(CTV3, SNOMED...)

- Identifiers are meaningless
 - to fix probs with meaningful IDs
 - but therefore lexical comparison of IDs won't work
- Hierarchy stored in IS-A table, not in codes:

PARENT	CHILD
X40Gd	X40J1
X40J1	X40J3
X40J3	C10..
C10..	X40J4
X40J4	Xa4g7

- But IS-A table too slow for subsumption check
 - Recursive calls : **Xa4g7** *subtype-of* **X40Gd** ?
 - Need derivative : 'Transitive Closure' table

'Transitive Closure' Table?

- IS-A relationship is logically 'transitive':
 - $\forall a,b,c : a \text{ IS-A } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c$
- 'Closure' table : list of all relationships inferrable from IS-A axioms

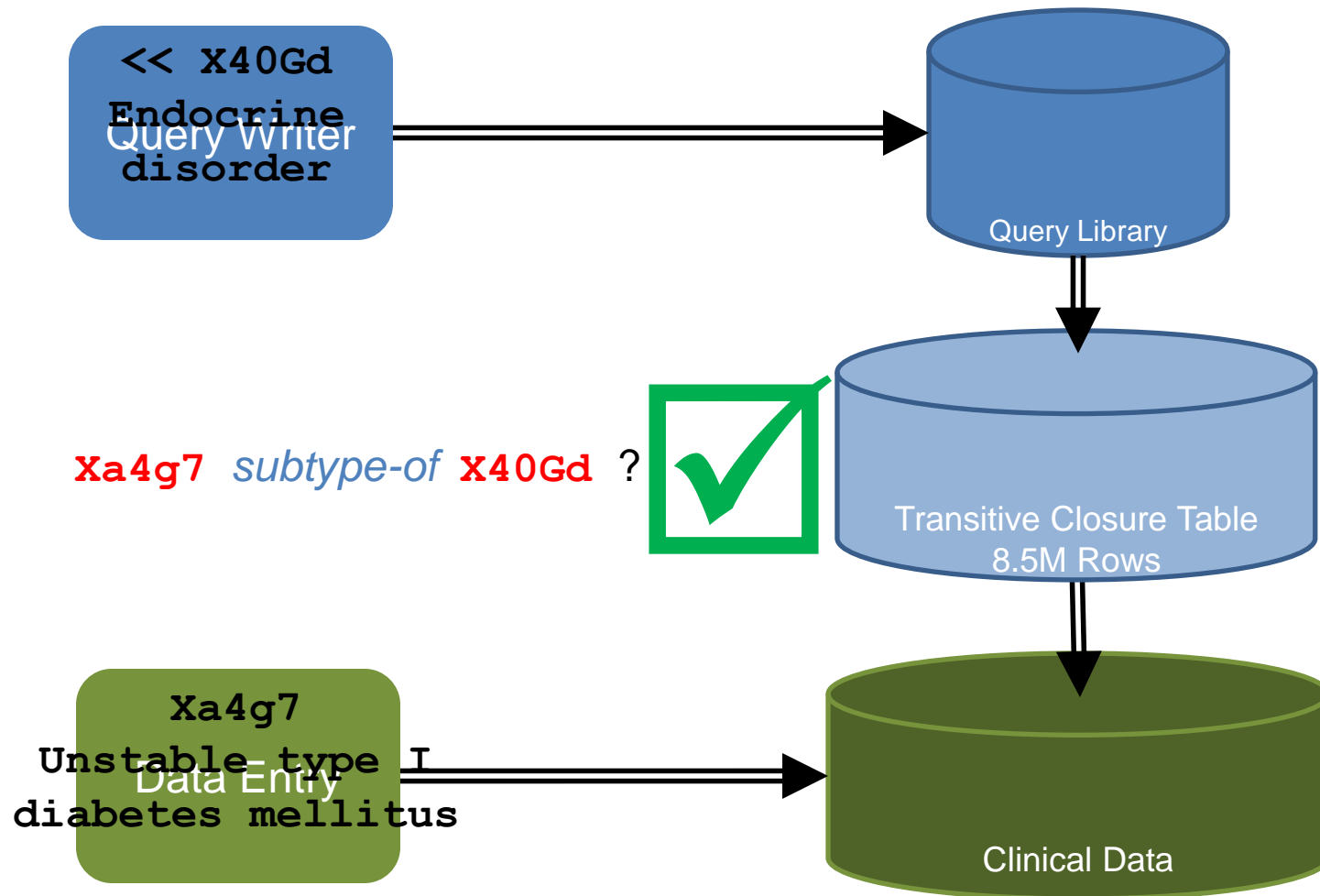
PARENT	CHILD	SUPERTYPE	SUBTYPE
X40Gd	X40J1	X40Gd	X40Gd
X40J1	X40J3	X40Gd	X40J1
X40J3	C10..	X40Gd	C10..
C10..	X40J4	X40Gd	X40J4
X40J4	Xa4g7	X40Gd	Xa4g7
		X40J1	X40J1
		X40J1	C10..
		X40J1	X40J4
		X40J1	Xa4g7
		X40J3	X40J3
		X40J3	C10..
		X40J3	X40J4
		X40J3	Xa4g7
		C10..	C10..
		C10..	X40J4
		C10..	Xa4g7
		X40J4	X40J4
		X40J4	Xa4g7
		Xa4g7	Xa4g7



Similar closure tables can be built for any other transitive relationship e.g. partOf, causedBy, follows etc

Typical Reporting Architecture

(example from CTV3)



NB Some graph databases, or graph-optimised RDBMS servers, include SQL transitivity extensions such as CONNECT BY that mean you don't have to *explicitly* build a transitive closure table

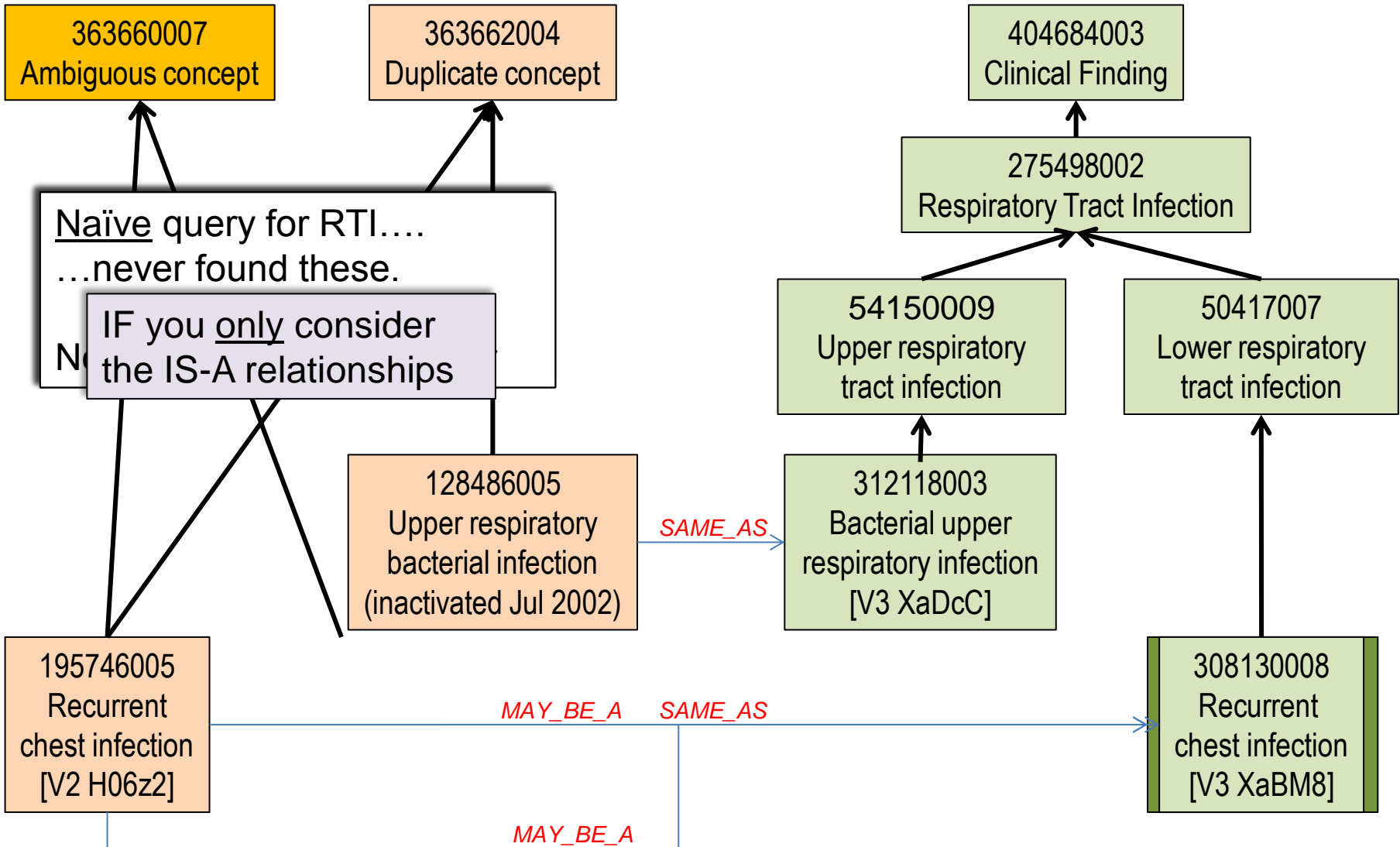
Outline

- Recap : how code queries are executed
- The 'inactive content' problem
- Two-part solution
 - Substitutions table
 - 'Role Inclusion Closure' table

The 'inactive content' problem

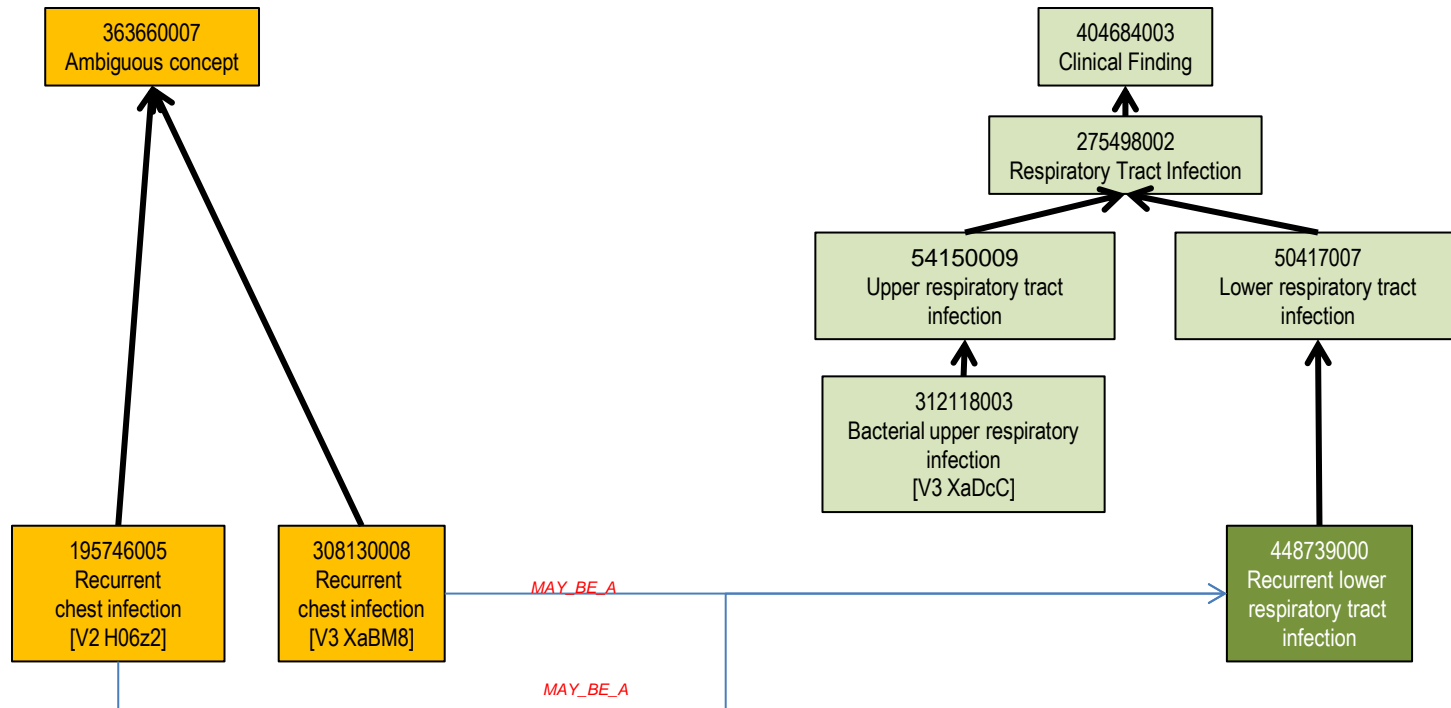
- Classical SNOMED CT querying - including by means of 'transitive closure' tables - ONLY considers IS-A relationships
- In combination with how SNOMED CT represents concept inactivation 'out of the box', this is not sufficient

Bacterial upper respiratory infection



Summary of Result

Inactive content in your data



TC 2011a TC 2011b

252744	250791
2421	522
2257	358
2000	2000
0	100

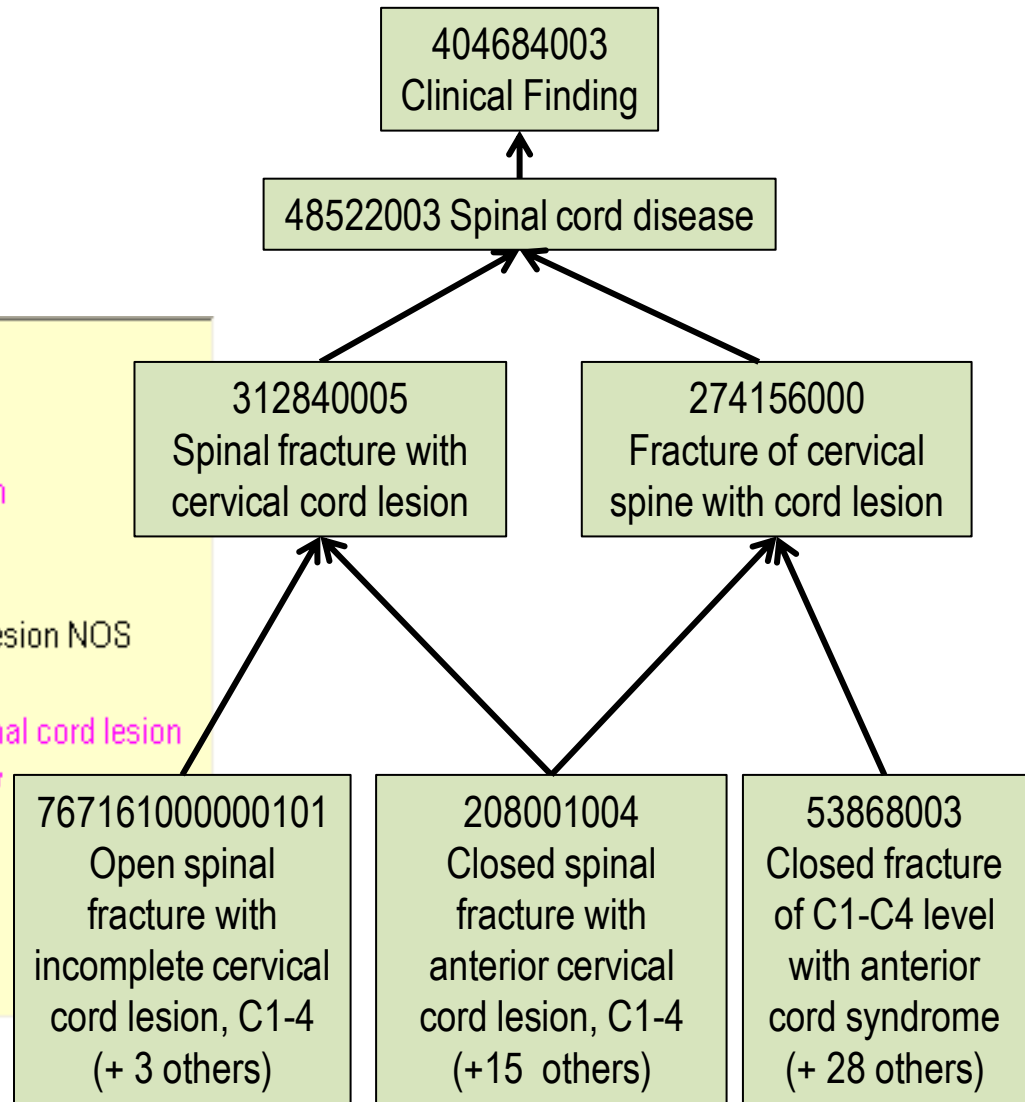
- 404684003 Clinical Finding
- 275498002 Respiratory Tract Infection
- 50417007 Lower respiratory tract infection
- 308130008 Recurrent chest infection (inactivated Oct 2011)
- 448739000 Recurrent lower respiratory tract infection (added current Oct 2011)

Inactive content in your Queries

Before October 2011...

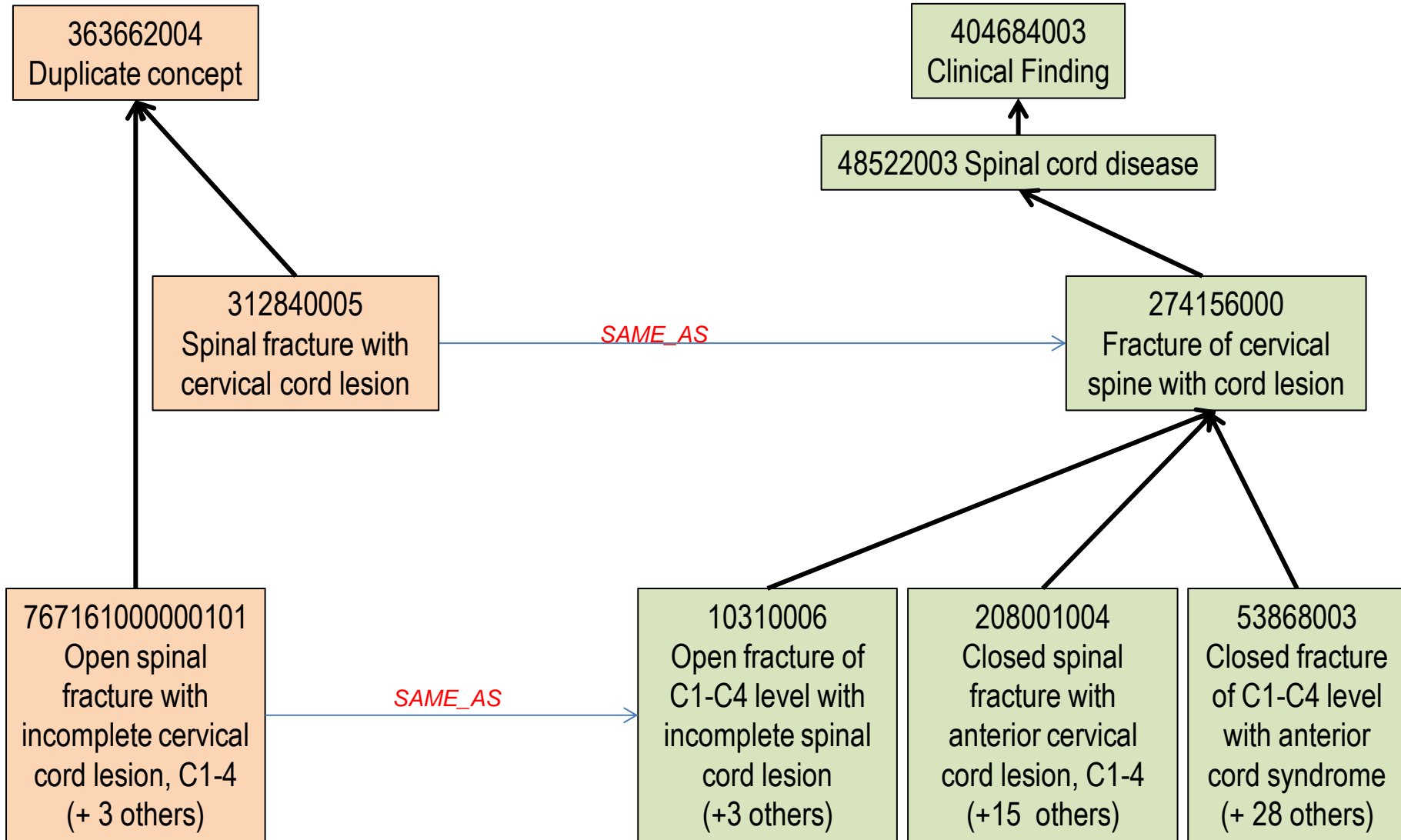
363662004
Duplicate concept

- + 50448004 Fracture of vertebral column
- + 48522003 Spinal cord disease
- + 269078005 Fracture of spine with spinal cord lesion
- + 208082008 Closed fracture of coccyx with spinal cord lesion
- + 274156000 Fracture of cervical spine with cord lesion
- + 274158004 Fracture of lumbar spine with cord lesion
- x 687331000000100 [UK] Fracture of spine with spinal cord lesion NOS
- + 274157009 Fracture of thoracic spine with cord lesion
- x 263179007 Fracture of transverse process of spine with spinal cord lesion
- + 1734006 Fracture of vertebral column with spinal cord injury
- + 208088007 Open fracture of coccyx with spinal cord lesion
- + 312840005 Spinal fracture with cervical cord lesion
- + 312842002 Spinal fracture with lumbar cord lesion
- + 312841009 Spinal fracture with thoracic cord lesion



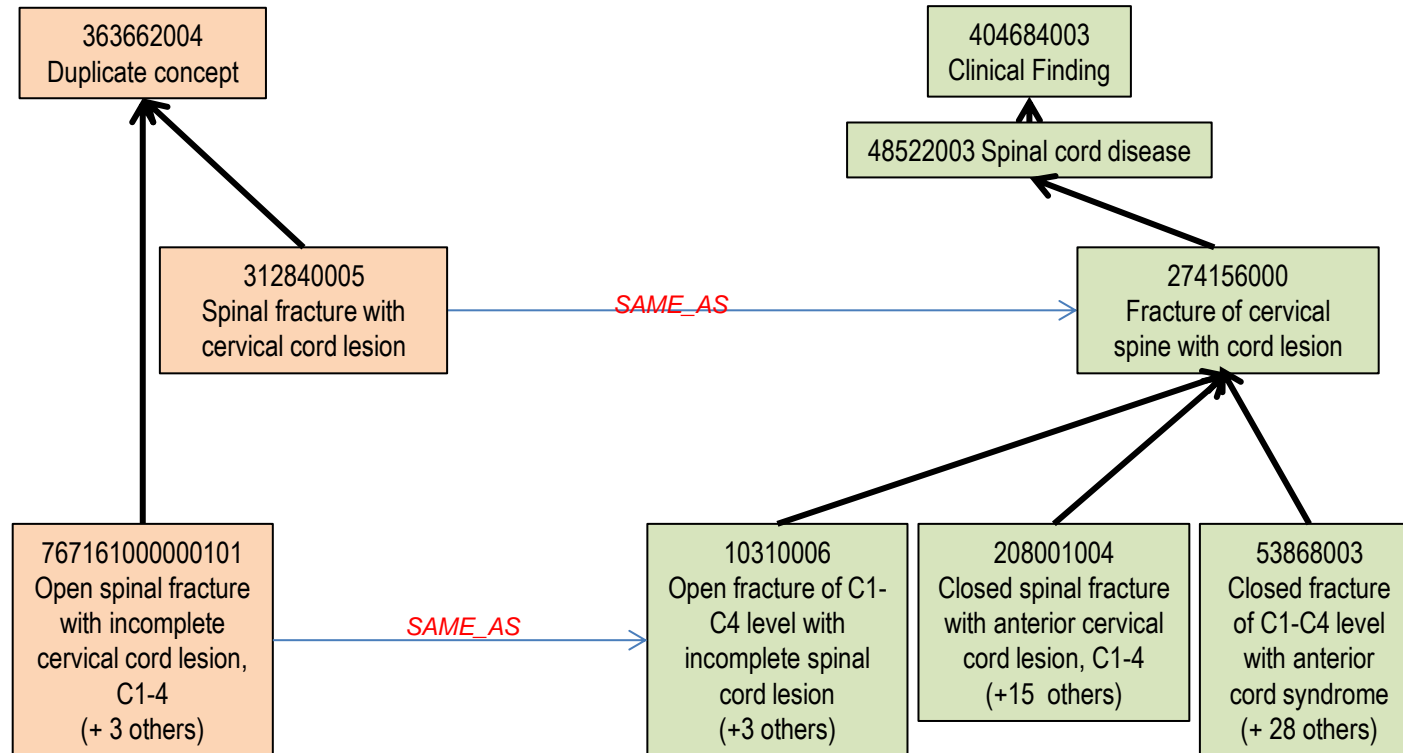
Inactive content in your Queries

After October 2011...



Summary of Result

Inactive content in your Queries



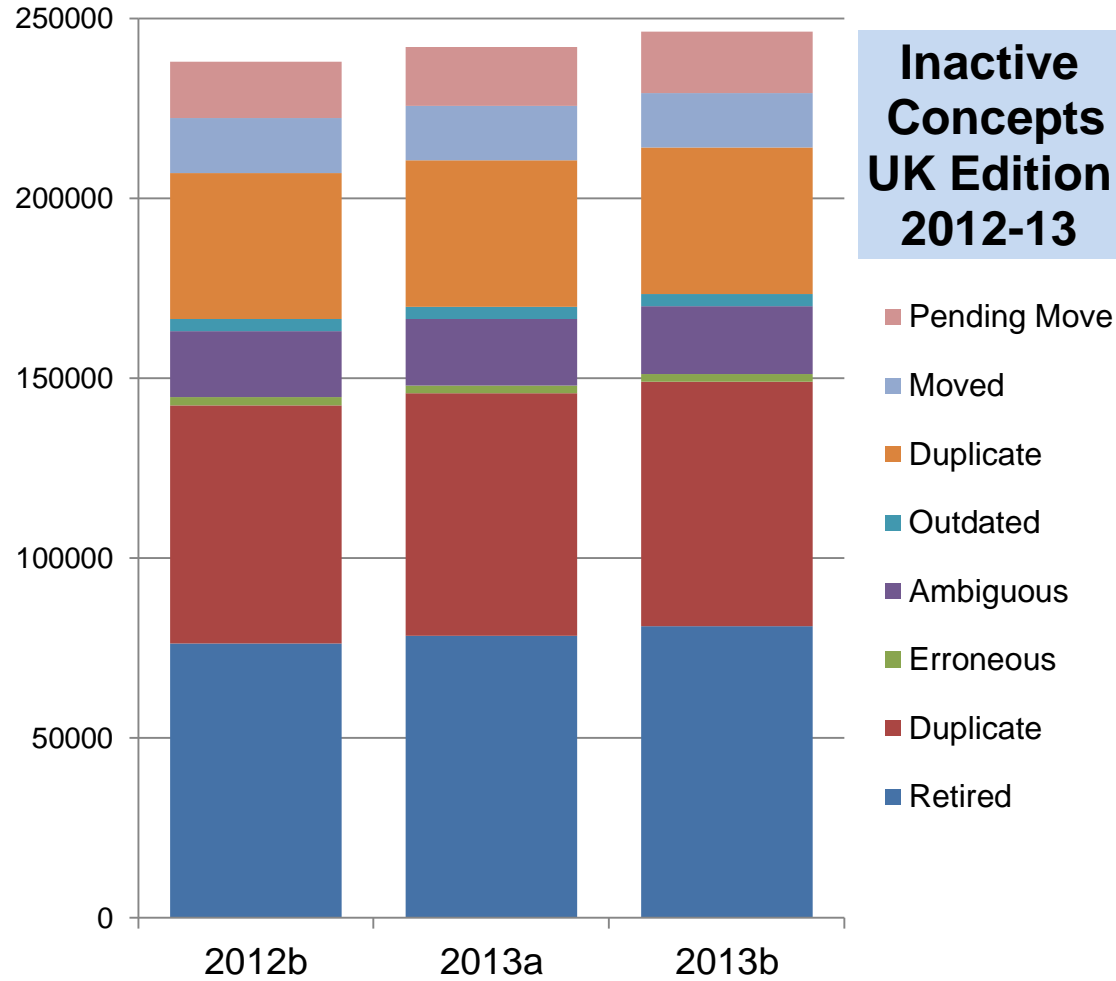
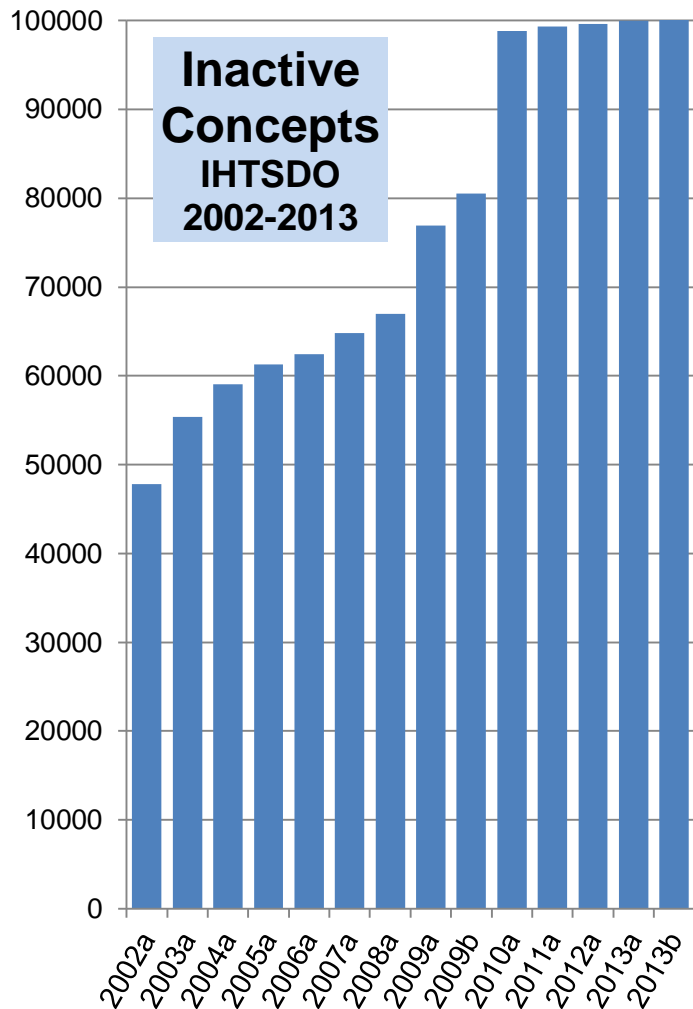
TC 2011a TC 2011b

1200	100
500	445
400	386

312840005 Spinal fracture with cervical cord lesion (duplicate Oct 2011)
 274156000 Fracture of cervical spine with cord lesion
 125609005 Open fracture of cervical region with spinal cord injury

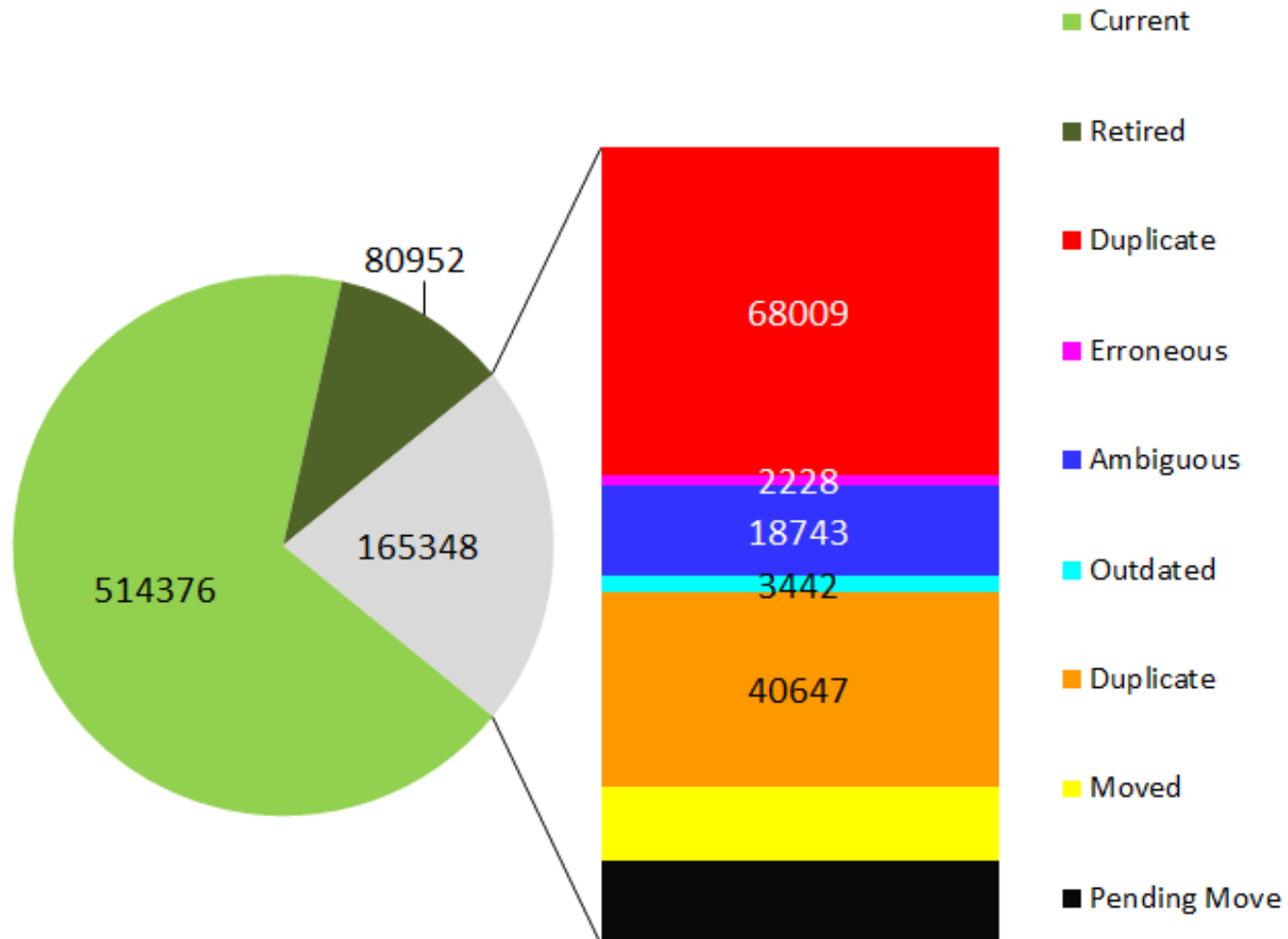
Concept Inactivations so far..

(UK Edition [International, Drug plus Clinical] Releases)



Concept Inactivations so far..

(UK Edition [International, Drug plus Clinical] October 2013 Release)



Concept inactivation so far in...

UMLS Core Problem List

5191 concepts in original 200908 release

64 are now inactive (1.24% of codes)

Account for 0.75% combined 'frequency'

398175007 Male erectile disorder (disorder)
190392008 Type II diabetes mellitus - poor control (disorder)
399221001 Bleeding from vagina (disorder)
38511004 Fetus OR newborn affected by premature rupture of membranes (disorder)
64756007 Previous cesarean section (disorder)
201836008 Localized, primary osteoarthritis of the lower leg (disorder)
267782008 Cellulitis and abscess of leg (disorder)
198881004 Pregnancy complications (disorder)
6408001 Finding of nocturia (finding)
102574007 Edema of leg (finding)
102550009 Leg cramp (finding)
69124005 Complete abortion (disorder)
268808004 Fetal or neonatal effect of breech delivery and extraction (disorder)
33282003 Stenosis of esophagus (disorder)
238402004 Cellulitis of leg (disorder)
277675000 Blind (finding)
192839001 Essential tremor (finding)
106190000 Allergic state (disorder)
190371008 Type I diabetes mellitus - poor control (disorder)
91588005 Closed fracture of metacarpal bone (disorder)
18001006 Fetal or neonatal effect of multiple pregnancy (disorder)
4946002 Threatened premature labor (disorder)
16863000 Incomplete abortion (disorder)
37757003 Communicable disease contact (finding)
201441006 Scleroderma (disorder)
200665006 Cellulitis and abscess of arm (disorder)
128079007 Reflex sympathetic dystrophy (disorder)
276328002 Telangiectasia (disorder)

65599008 Fetal or neonatal effect of oligohydramnios (disorder)
23294000 Sports injury (morphologic abnormality)
37472003 Fluid volume deficit (finding)
192840004 Benign familial tremor (finding)
284480000 Cellulitis of arm (disorder)
60535003 Adenomatous polyp of cervix (disorder)
275918005 Unstable diabetes mellitus (disorder)
199516000 Known or suspected fetal abnormality (disorder)
66215008 Fetal or neonatal effect of polyhydramnios (disorder)
58193001 Diplegic cerebral palsy (disorder)
386633000 Orchiectomy (procedure)
68983007 Fetal or neonatal effects of maternal complication of pregnancy (disorder)
300889000 Swelling of arm (finding)
417162001 Nasolacrimal duct obstruction (disorder)
267821008 Hypertrophic cicatrix (disorder)
76226003 Tattoo (disorder)
410064000 Non-traumatic subdural hematoma (disorder)
73890002 Fetal or neonatal effect of delivery by vacuum extractor (disorder)
89600009 Secondary cardiomyopathy (disorder)
262951009 Traumatic subdural hematoma (disorder)
262954001 Traumatic subarachnoid intracranial hemorrhage (disorder)
201937003 Traumatic arthropathy of the lower leg (disorder)
73009009 Bloodshot eye (finding)
249782009 Bowing of leg (finding)
61628006 Drug withdrawal syndrome in newborn (disorder)
47874006 Sprain of arm (disorder)
281638009 Hepatitis B contact (finding)
41006004 Depression (finding)
371330000 Fatty liver (disorder)
131016008 Increased thyroid stimulating hormone level (finding)
166829003 Serum cholesterol borderline (finding)
78431007 Influenza due to Influenza virus, type A, human (disorder)
416103000 Elevated erythrocyte sedimentation rate (finding)
50047001 Compound dental caries (disorder)
63079007 Closed traumatic dislocation of hip joint (disorder)
64333001 Preinfarction angina (disorder)

Concept inactivation so far in...

Data from one ED department

- 408,831 episodes
 - 1 code per episode reason for attendance
 - 38 months of data Oct-2008 to Dec-2011
- Code now inactive for 1216 episodes (0.3%)
- 81 codes involved (out of 12,069 used)

Top 20 most frequently used inactive codes:

220	<i>episodes of</i>	267040003	Leg swelling (finding)
130		77299006	Olecranon bursitis (disorder)
118		41006004	Depression (finding)
95		238402004	Cellulitis of leg (disorder)
55		300889000	Swelling of arm (finding)
53		425406006	Hematoma of leg (disorder)
45		211832003	Partial thickness burn of arm (disorder)
39		198881004	Pregnancy complications (disorder)
39		285321000000107	Haematoma of leg (disorder)
38		284480000	Cellulitis of arm (disorder)
34		90821003	Complication related to pregnancy (disorder)
32		211828009	Superficial burn of arm (disorder)
29		47874006	Sprain of arm (disorder)
20		267782008	Cellulitis and abscess of leg (disorder)
18		281842005	Fracture tibial plateau (disorder)
18		102550009	Leg cramp (finding)
15		402261004	Chemical burn (disorder)
14		1508000	Intracerebral hemorrhage (disorder)
13		162345005	Blocked ear (finding)
12		37324003	Superficial injury of leg with infection (disorder)

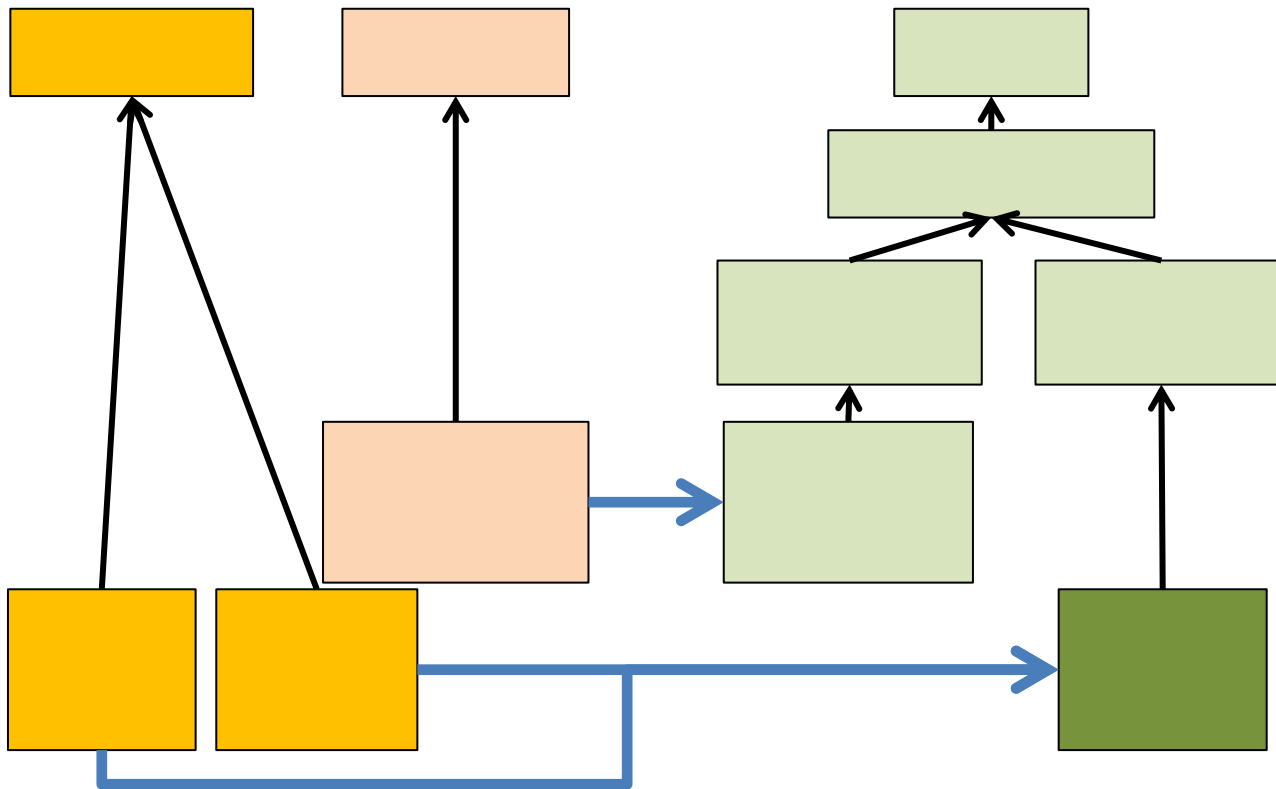
Outline

- Recap : how code queries are executed
- The 'inactive content' problem
- **Two-part solution**
 - Substitutions table
 - 'Role Inclusion Closure' table

Two-part Solution

Part 1: Substitutions Table

- Need to compute the blue arrows



Building a Substitutions Table

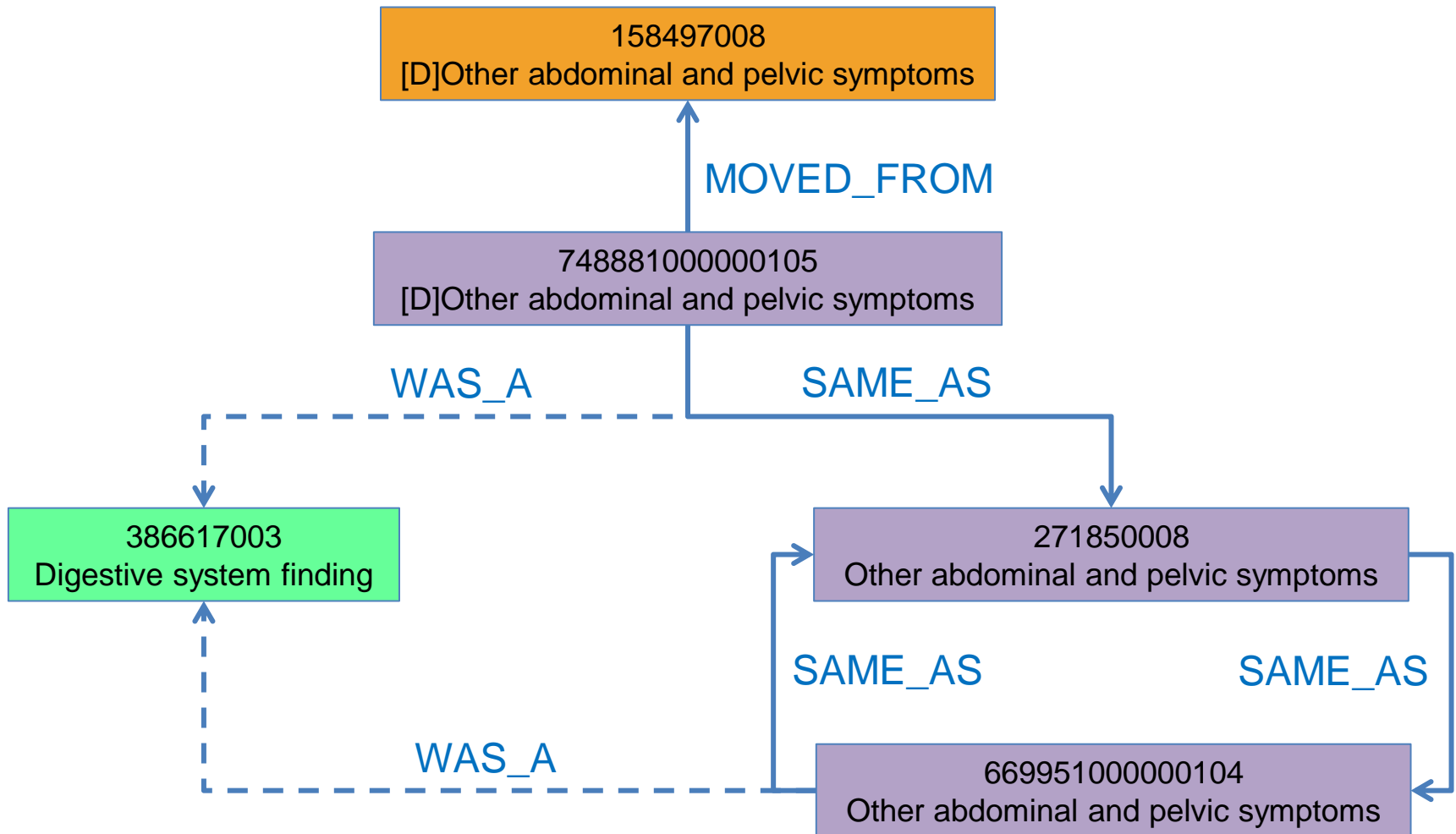
- Theoretically, simple lookup of
 - RF1 history relations from sct1_relationships
 - RF2 association reference sets
- In practice, not that simple

Building a Substitutions Table

- Not 1:1 (and often 1:0)
- Necessarily recursive if extensions involved
 - 158497008 [D]Other abdominal and pelvic symptoms
 - 170251007 Child 6 month examination NEC (procedure)
- SAME_AS is cyclic for ‘limited status’
- Not all 1:∞ is explicitly flagged
 - 667611000000104 Myalgia and myositis unspecified (disorder)
- Special handling in some jurisdictions
- Unclear semantics for WAS_A, MAY_BE

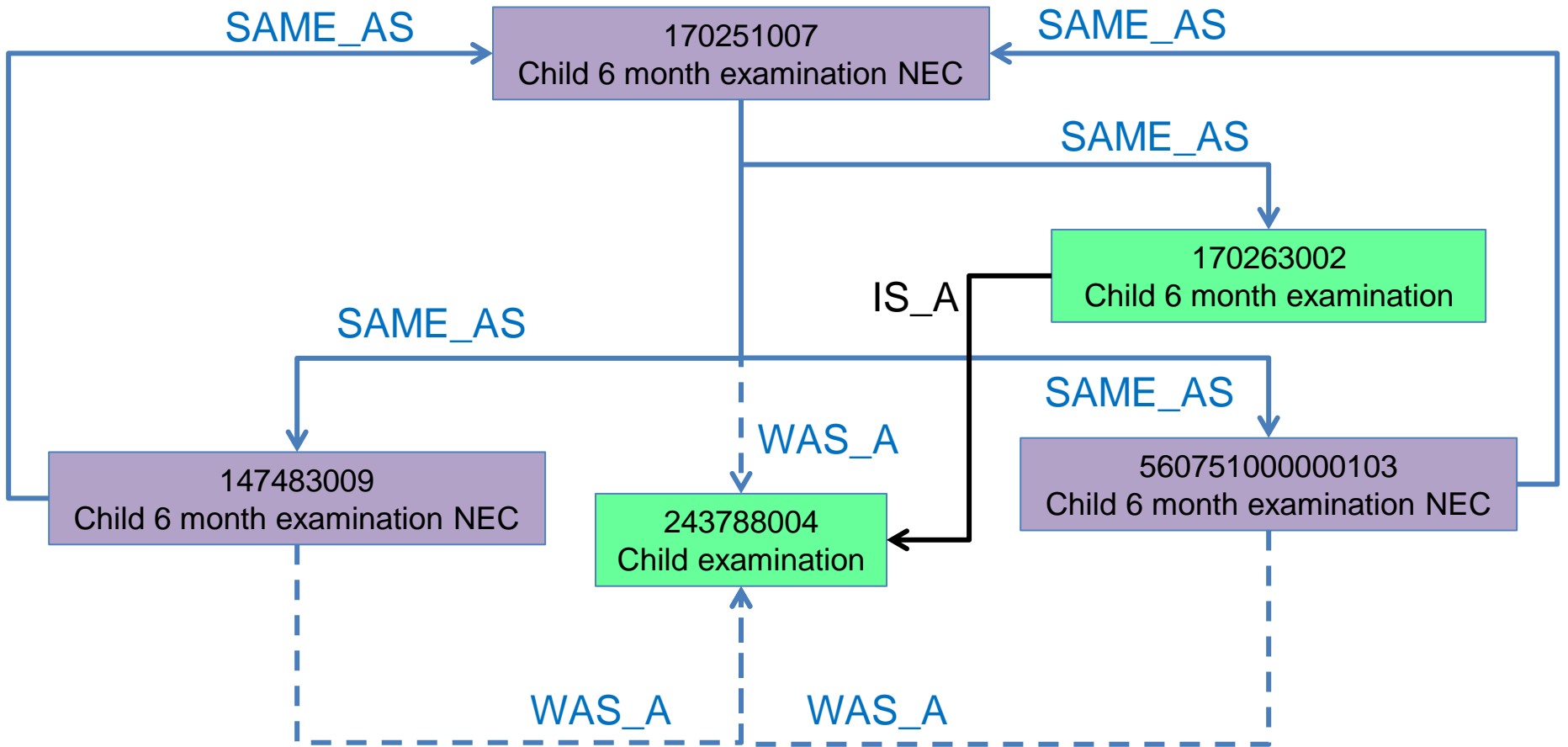
SNOMED CT Historical Relations

Recursive and cyclic



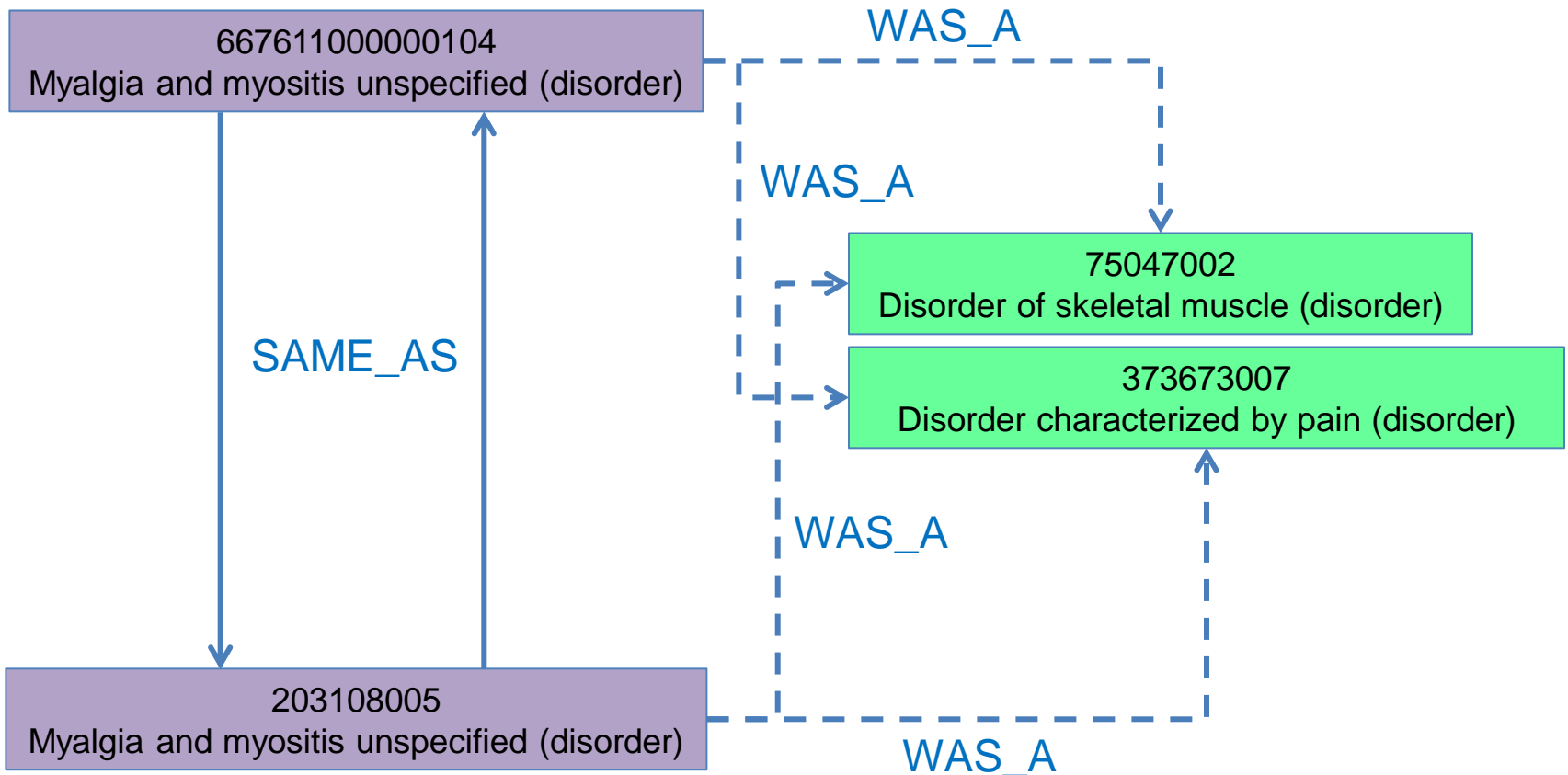
SNOMED CT Historical Relations

Recursive and cyclic



```
+ 243788004 Child examination
+ 170263002 Child 6 month examination
x 410627007 Well child visit, 6 month
```

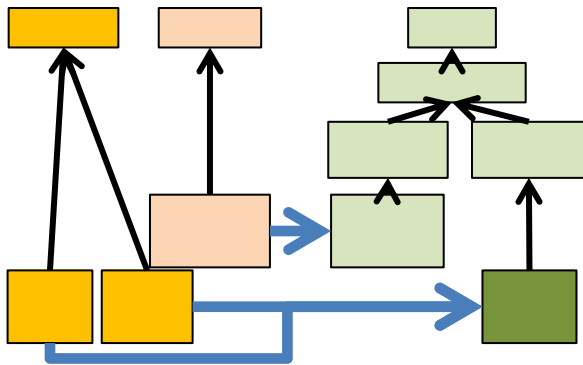
Unflagged ambiguity



Two-part Solution

Part 1: Substitutions Table

- Need to compute the blue arrows



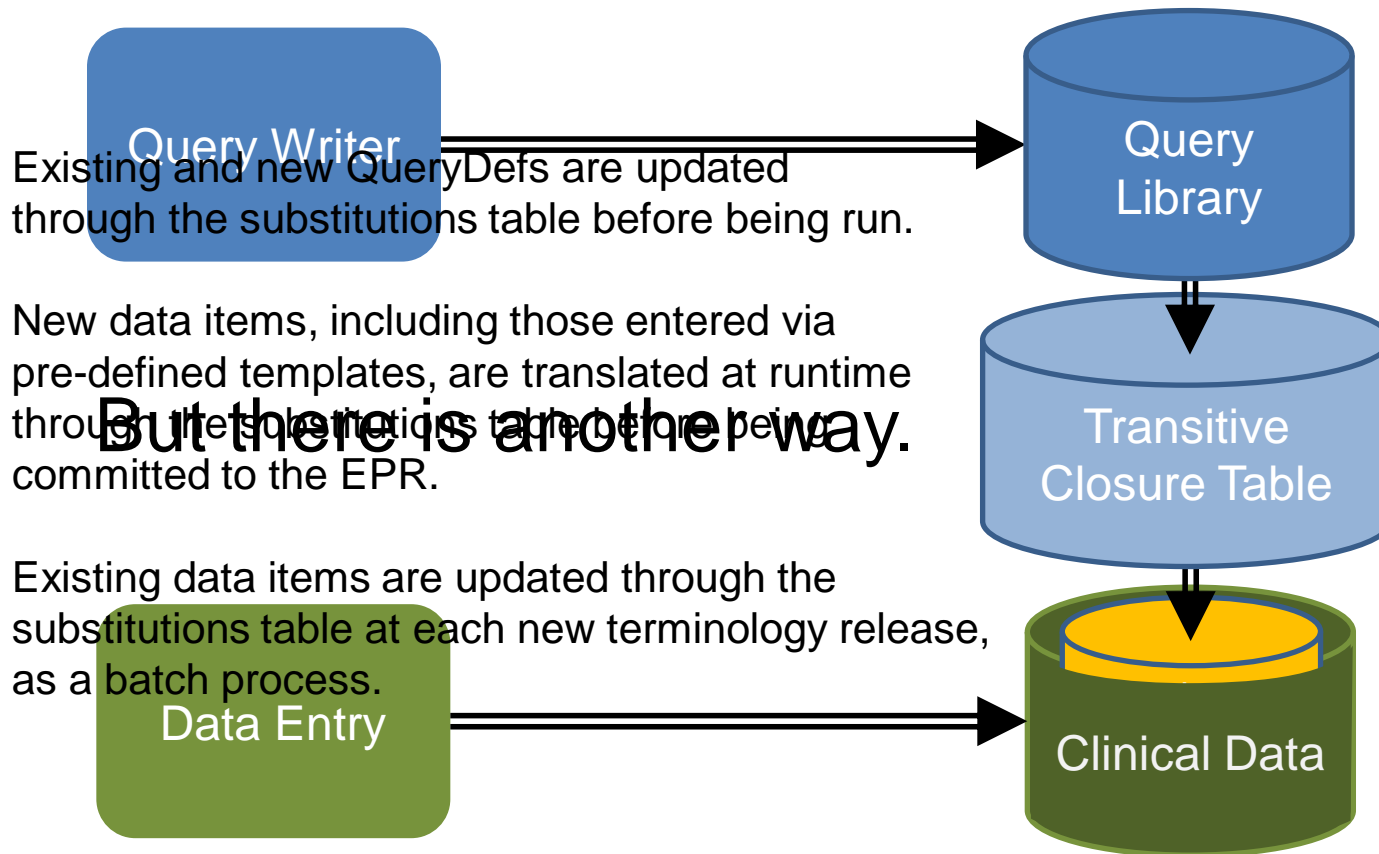
- Fiddly algorithm (though 2 minutes to execute)
- UKTC Product: UK Substitutions Table

Outline

- Recap : how code queries are executed
- The 'inactive content' problem
- Two-part solution
 - Substitutions table
 - 'Role Inclusion Closure' table

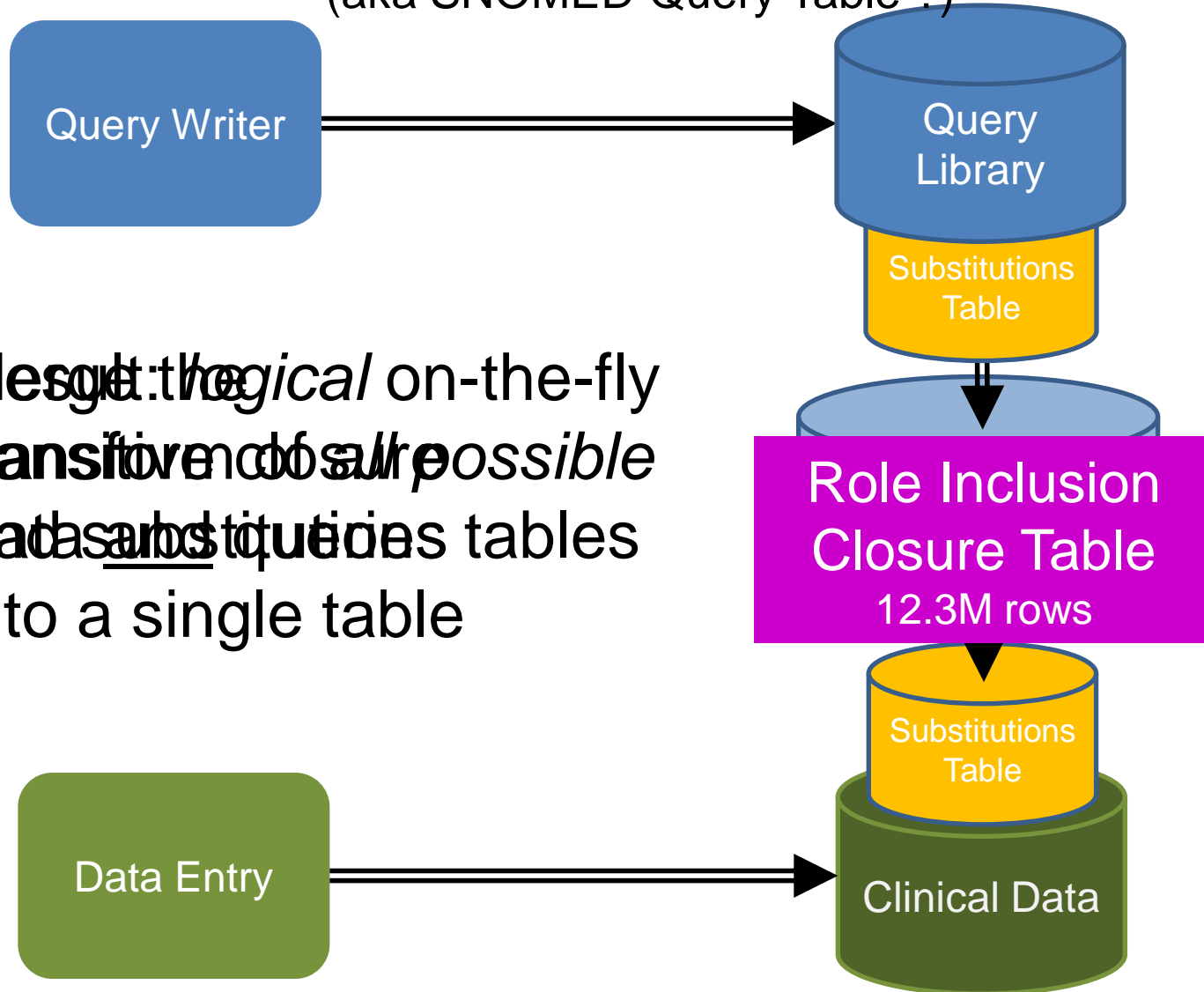
Using a Substitutions Table

One possible architecture..



'Role Inclusion Closure' Table

(aka SNOMED Query Table ?)

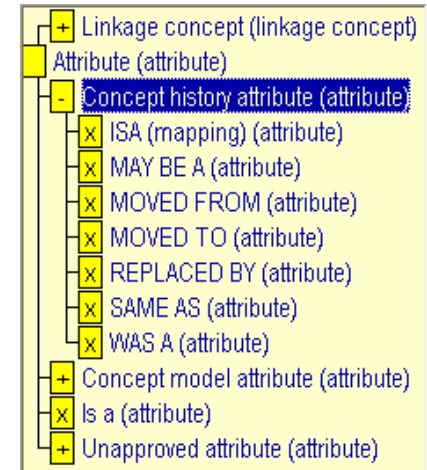


Result: the logical on-the-fly transformation of all possible data and substitutions tables into a single table

R Query Table ?

- Combinatorial logic of IS-A + history roles

- $\forall a,b,c : a \text{ IS-A } b, b \text{ SAME-AS } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ SAME-AS } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : b \text{ MOVED-FROM } a, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ IS-A } b, b \text{ MOVED-FROM } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ IS-A } b, b \text{ MOVED-FROM } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ WAS-A } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ IS-A } b, b \text{ WAS-A } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ REPLACED-BY } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c$
- $\forall a,b,c : a \text{ IS-A } b, c \text{ REPLACED-BY } b \Rightarrow a \text{ IS-A } c$



- $\forall a,b,c : a \text{ MAY-BE-A } (b \Delta c), b \text{ IS-A } d, c \text{ IS-A } e \Rightarrow a \text{ MAY-BE-A } d, a \text{ MAY-BE-A } e, a \text{ IS-A } (d \Delta e) ?$
- More than (single) role transitivity; now 'role inclusion'
- 'Closure' table : list of all descendents inferrable from set of IS-A, SAME-AS, WAS-A, MOVED-FROM, REPLACED-BY and MAY-BE-A axioms

Query Table: How it works

PARENT	CHILD	SUPERTYPEID	SUBTYPEID	SUPERTYPEID	SUBTYPEID	SUPERTYPEID	SUBTYPEID	SUPERTYPEID	SUBTYPEID
362969004	17346000	111553002	111553002	190323008	73211009	302864000	73211009	47481000000105	46635009
17346000	73211009	111553002	46635009	190362004	190362004	362969004	111553002	47481000000105	47481000000105
73211009	46635009	111553002	73211009	190362004	46635009	362969004	154671004	50201000000100	46635009
		154671004	154671004	190382000	190382000	362969004	154673001	50201000000100	50201000000100
		154671004	46635009	190382000	46635009	362969004	154703006		
		154671004	73211009	190382000	73211009	362969004	17346000		
		154673001	154673001	190383005	190383005	362969004	190322003		
		154673001	46635009	190383005	46635009	362969004	190323008		
		154703006	154703006	190383005	73211009	362969004	190362004		
		154703006	17346000	190420007	190420007	362969004	190382000		
		154703006	362969004	190420007	46635009	362969004	190383005		
		154703006	46635009	190420007	73211009	362969004	190420007		
		154703006	73211009	190567009	17346000	362969004	190567009		
		17346000	111553002	190567009	190567009	362969004	191024000		
		17346000	154671004	190567009	362969004	362969004	191044006		
		17346000	154673001	190567009	46635009	362969004	267467004		
		17346000	17346000	190567009	73211009	362969004	267469001		
		17346000	190322003	191024000	17346000	362969004	267486007		
		17346000		191024000	191024000	362969004	302864000		
		17346000		191024000	362969004	362969004	362969004		
		17346000		191024000	46635009	362969004	43521000000100		
		17346000		191024000	73211009	362969004	46635009		
		17346000	190420007	191044006	191044006	362969004	47481000000105		
		17346000	191044006	191044006	46635009	362969004	50201000000100		
		17346000	267467004	191044006	73211009	362969004	62163100000105		
		17346000	46635009	267467004	267467004	362969004	65800100000101		
		17346000	43521000000100	267467004	46635009	362969004	65805100000100		
		17346000	46635009	267467004	73211009	362969004	73211009		
		17346000	47481000000105	267469001	267469001	362969004	75521003		
		17346000	50201000000100	267469001	46635009	43521000000100	43521000000100		
		17346000	62163100000105	267486007	17346000	43521000000100	46635009		
		17346000	65800100000101	267486007	267486007	46635009	154673001		
		17346000	65805100000100	267486007	362969004	46635009	190322003		
		17346000	73211009	267486007	46635009	46635009	190362004		
		190322003	190322003	267486007	73211009	46635009	267469001		
		190322003	46635009	302864000	17346000	46635009	43521000000100		
		190322003	73211009	302864000	302864000	46635009	46635009		
		190323008	190323008	302864000	362969004	46635009	47481000000105		
		190323008	46635009	302864000	46635009	46635009	50201000000100		

Substitution Table

IF 191044006
SAME-AS
73211009 THEN
everything that's
known about
73211009
is also true of
191044006.
It has the same
descendents.
And the same
ancestors.
The role inclusion
closure for all
substitutes is a
MUCH bigger table.

Standard
Hierarchy
Table

Standard
Transitive
Closure
Table

Query Table Advantages

- No change to typical supplier architecture
 - Just use a bigger table where TC should be
- Single transparent fix for inactive content in data *and* in queries *and* in bound forms

Outline

- Recap : how code queries are executed
- The 'inactive content' problem
- Two-part solution
 - Substitutions table
 - 'Role Inclusion Closure' table
- **Demo**
- **Caveats**

SNOMED Query Table Demo

Open Source MS Access VBA Application

Simulated data:

30,000 'patients'

180,000 coded episodes in GP distro

108k	Findings/Disorders
------	--------------------

54k	Procedures/Therapies
-----	----------------------

19k	Others
-----	--------

850 planted inactive codes

Caveats

- Full RI Table is 'huge'
 - 33M rows vs 8.5M for standard TC table
 - But only 12M if inactivation before 2
- Combinatorial semantics require
 - SAME-AS and MOVED-FROM s controversial; true identifier equi
 - REPLACED-BY also identifier ec
 - WAS-A : same ancestors, but no
 - Old descendents to be retrieved from
 - MAY-BE-A with...
 - only one substitute ? (43%)
 - > 1 nominated substitute (max = 23
 - IS-A lowest common subsumer?

169122008

CTV3: 57... Diagnostic nuclear medicine (& various isotope studies)

Parents

this concept Is a Ambiguous concept

Historical

this concept MAY BE A Brain isotope studies

this concept MAY BE A CSF isotope study

this concept MAY BE A Cystographic isotope studies

this concept MAY BE A Nuclear medicine diagnostic procedure

this concept MAY BE A Nuclear medicine imaging procedure

this concept MAY BE A Nuclear medicine procedure

this concept MAY BE A Placenta isotope study

this concept MAY BE A Plasma radioiron turnover rate

this concept MAY BE A Radioisotope function study of liver

this concept MAY BE A Radioisotope joint imaging

this concept MAY BE A Radioisotope scan of bone

this concept MAY BE A Radioisotope scan of lymphatic system

this concept MAY BE A Radioisotope scan of spleen

this concept MAY BE A Radioisotope study of liver

this concept MAY BE A Radionuclide studies in hematology

this concept MAY BE A Radionuclide study of heart

this concept MAY BE A Radionuclide study of lung

this concept MAY BE A Radionuclide study of thorax

this concept MAY BE A Radionuclide urinary tract study

this concept MAY BE A Renal isotope studies

this concept MAY BE A Skull isotope studies

this concept MAY BE A Thyroid imaging

this concept MAY BE A Vitamin B12 isotope studies

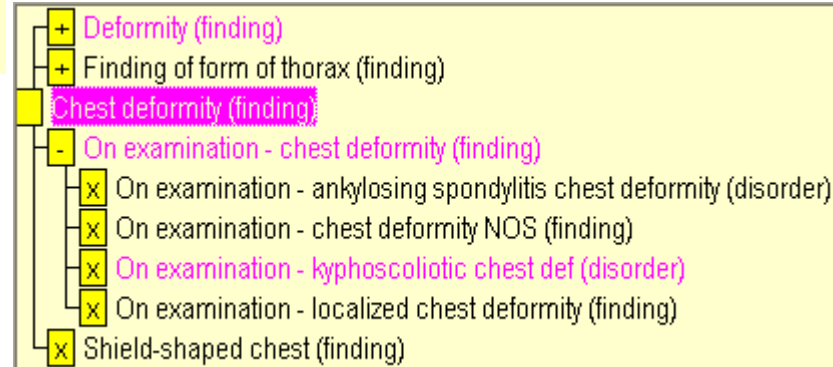
Conceptual Drift

SNOMED 2009

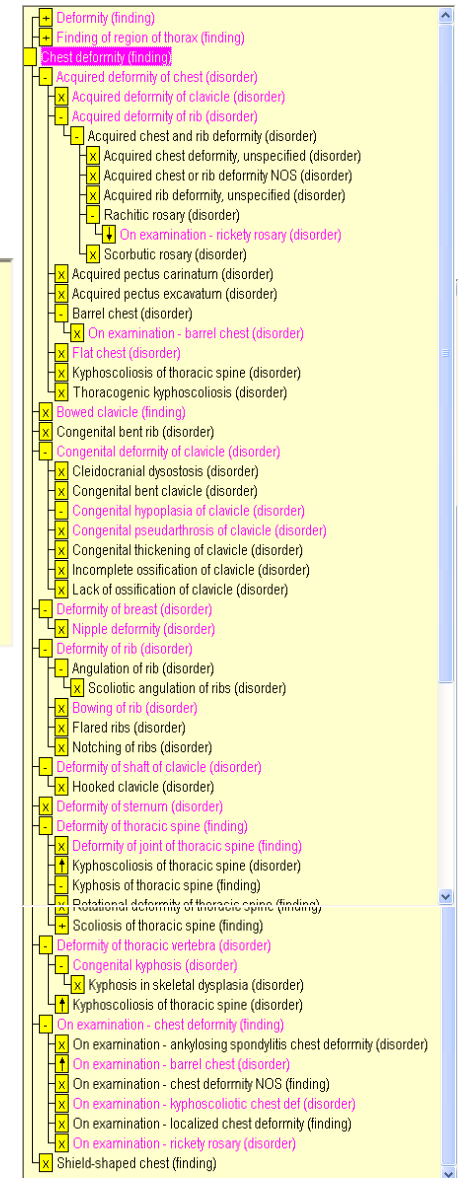
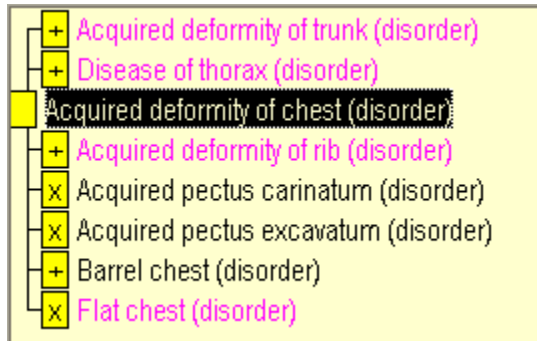
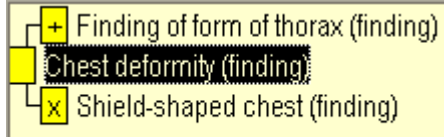
CTV3 (Jan 2002 – Oct 2010)

- X76gD Form of thorax
- Xa870 Chest deformity
- X76gJ Shield-shaped chest
- XM01u Pectus carinatum
- XM01w Pectus excavatum
- XM01x Barrel chest

SNOMED 2006



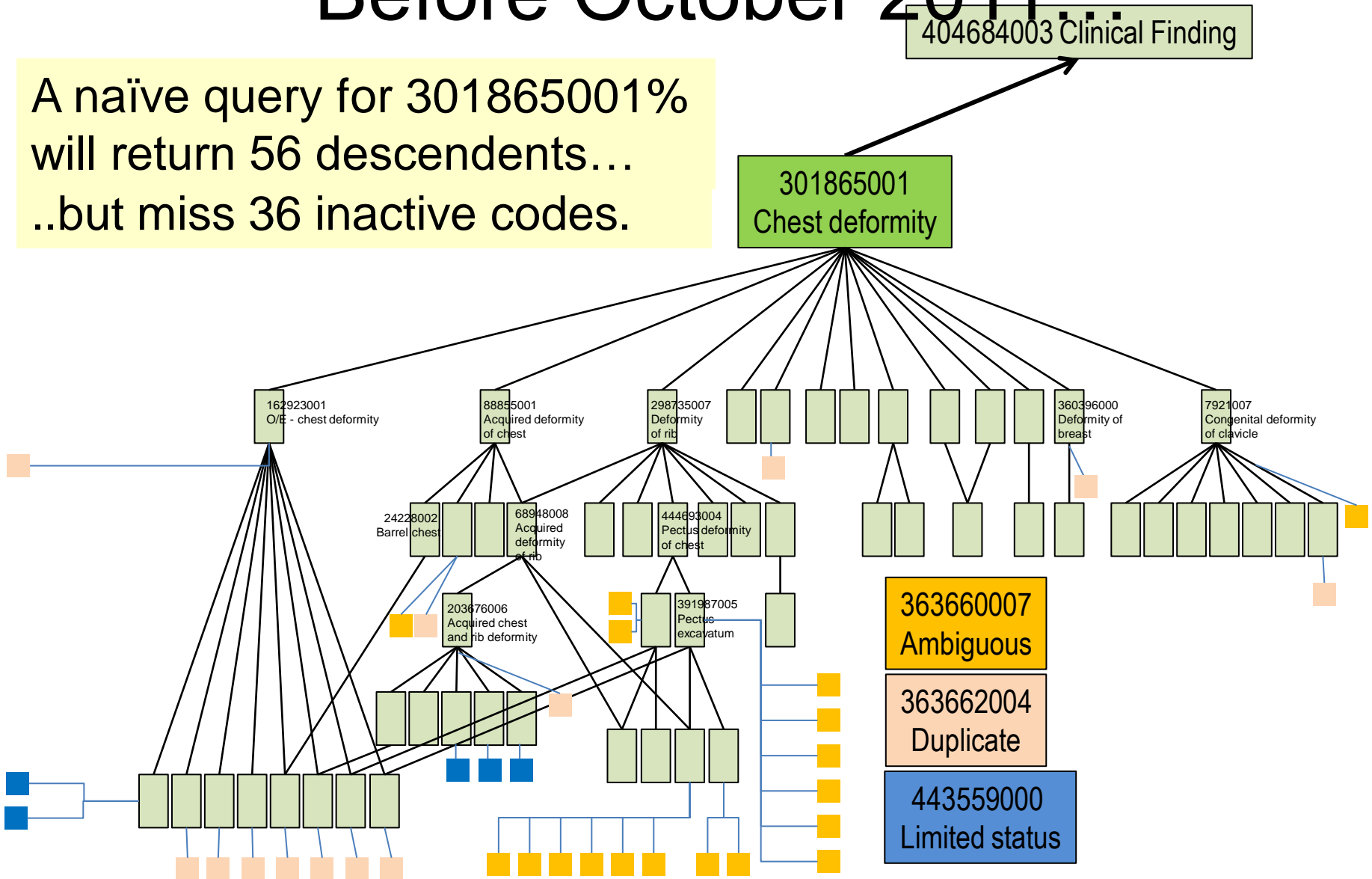
SNOMED 2002



What happens if the inactive concept is in the QueryDef?

Before October 2011

A naïve query for 301865001% will return 56 descendents...
..but miss 36 inactive codes.



404684003 Clinical Finding

301865001
Chest deformity

162923001
O/E - chest deformity

88835001
Acquired deformity of chest

298735007
Deformity of rib

360396000
Deformity of breast

7921007
Congenital deformity of clavicle

24228002
Barrel chest

68948008
Acquired deformity of rib

44493004
Pectus deformity of chest

203676006
Acquired chest and rib deformity

391987005
Pectus excavatum

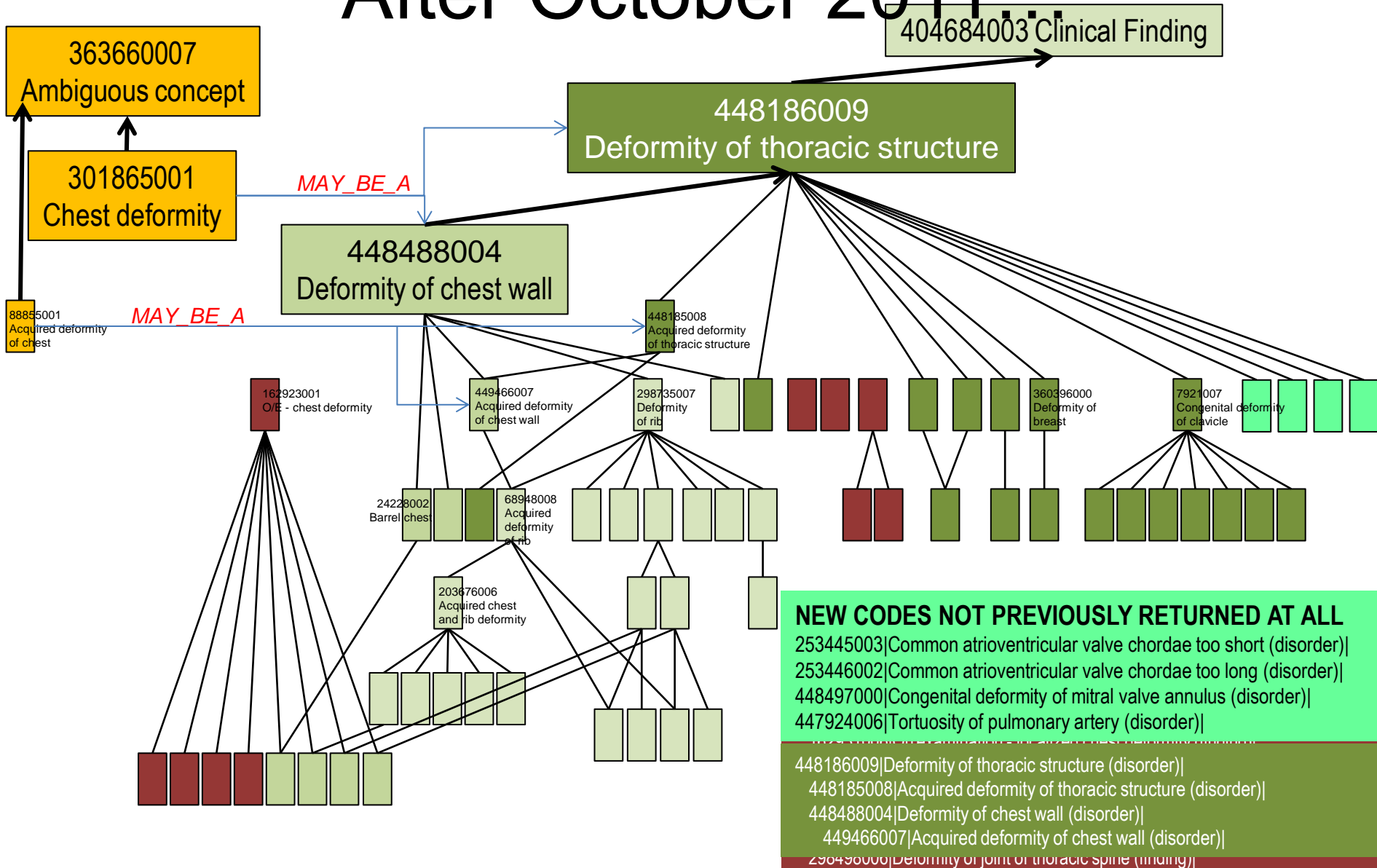
363660007
Ambiguous

363662004
Duplicate

443559000
Limited status

What happens if the inactive concept is in the QueryDef?

After October 2011



Thank You!

jeremy.rogers@hscic.gov.uk