Analyzing SNOMED CT's Historical Data: Pitfalls and Possibilities.

AMIA 2016 Annual Symposium
Chicago, Illinois, November 12 - 16, 2016

Werner CEUSTERS[a,b,c,d], MD and Jonathan P BONA[e], PhD

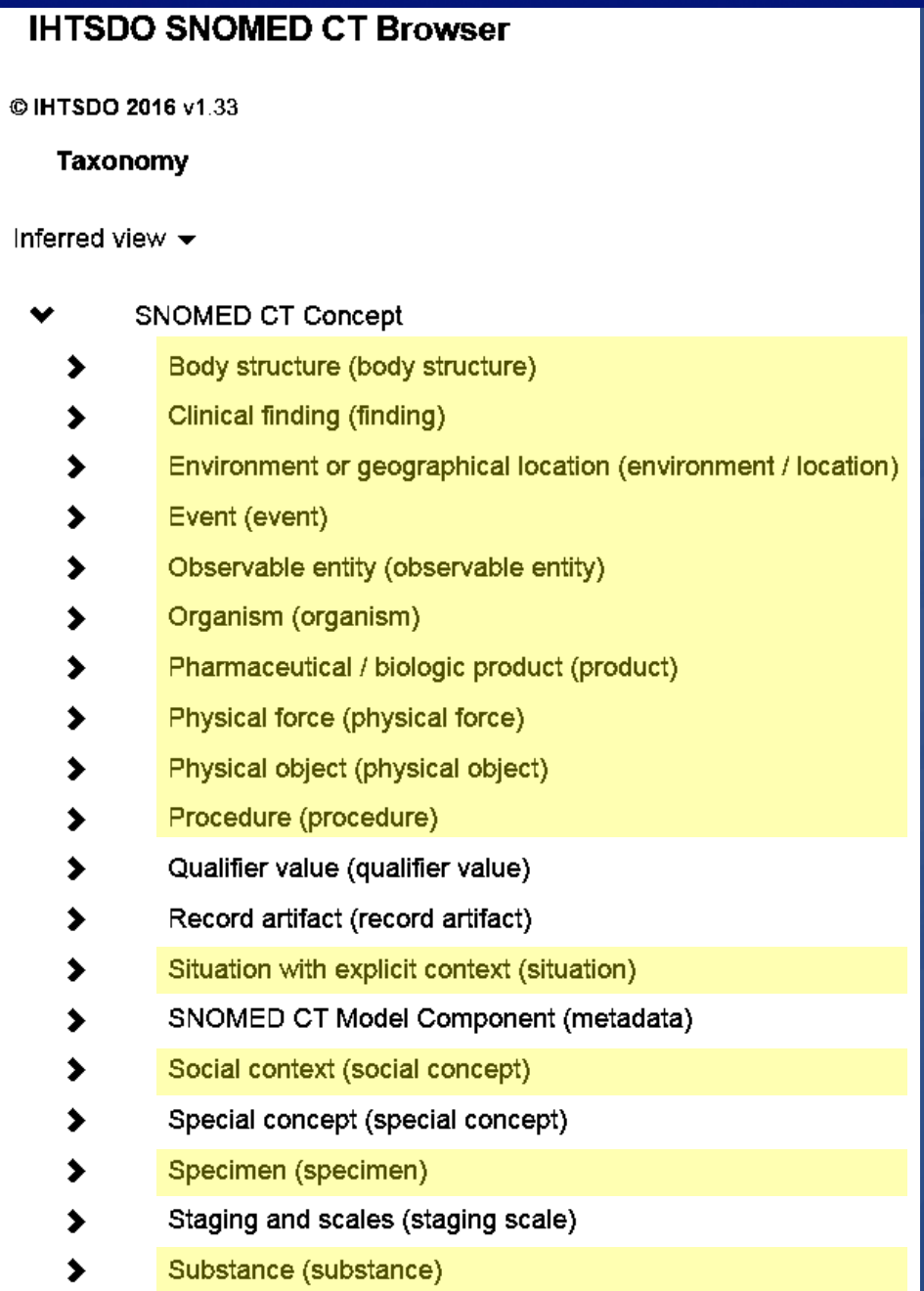
Ontology Research Group, Center of Excellence in Bioinformatics and Life Sciences[a];

UB Institute for Healthcare Informatics[b],

Departments of Biomedical Informatics[c] and Psychiatry[d],

University at Buffalo, NY, USA

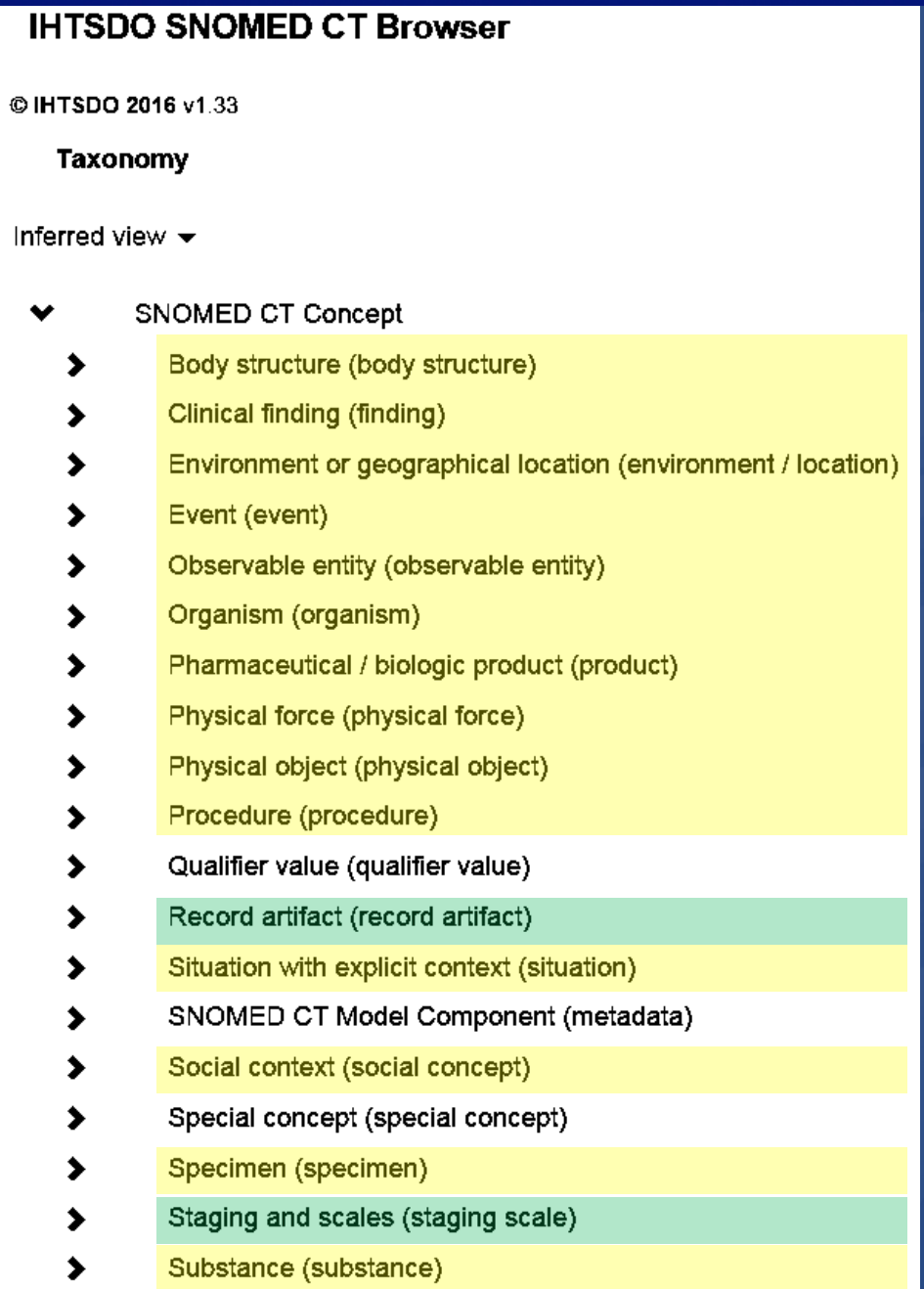
Department of Biomedical Informatics[e]
University of Arkansas for Medical Science, AR



SNOMED CT's top categories

denote (roughly):

A. Entities on the side of the patient

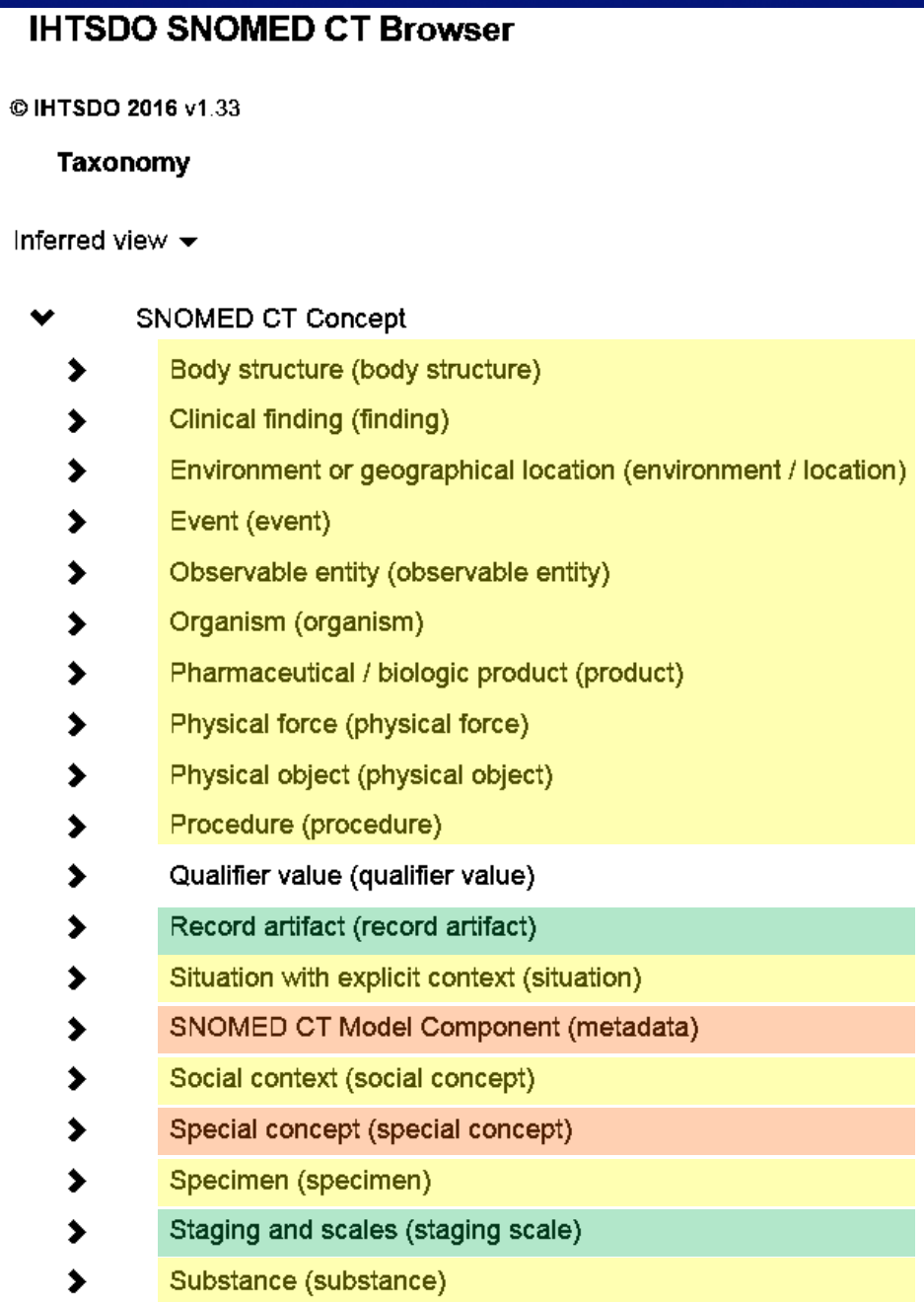


SNOMED CT's top categories

denote (roughly):

A. Entities on the side of the patient

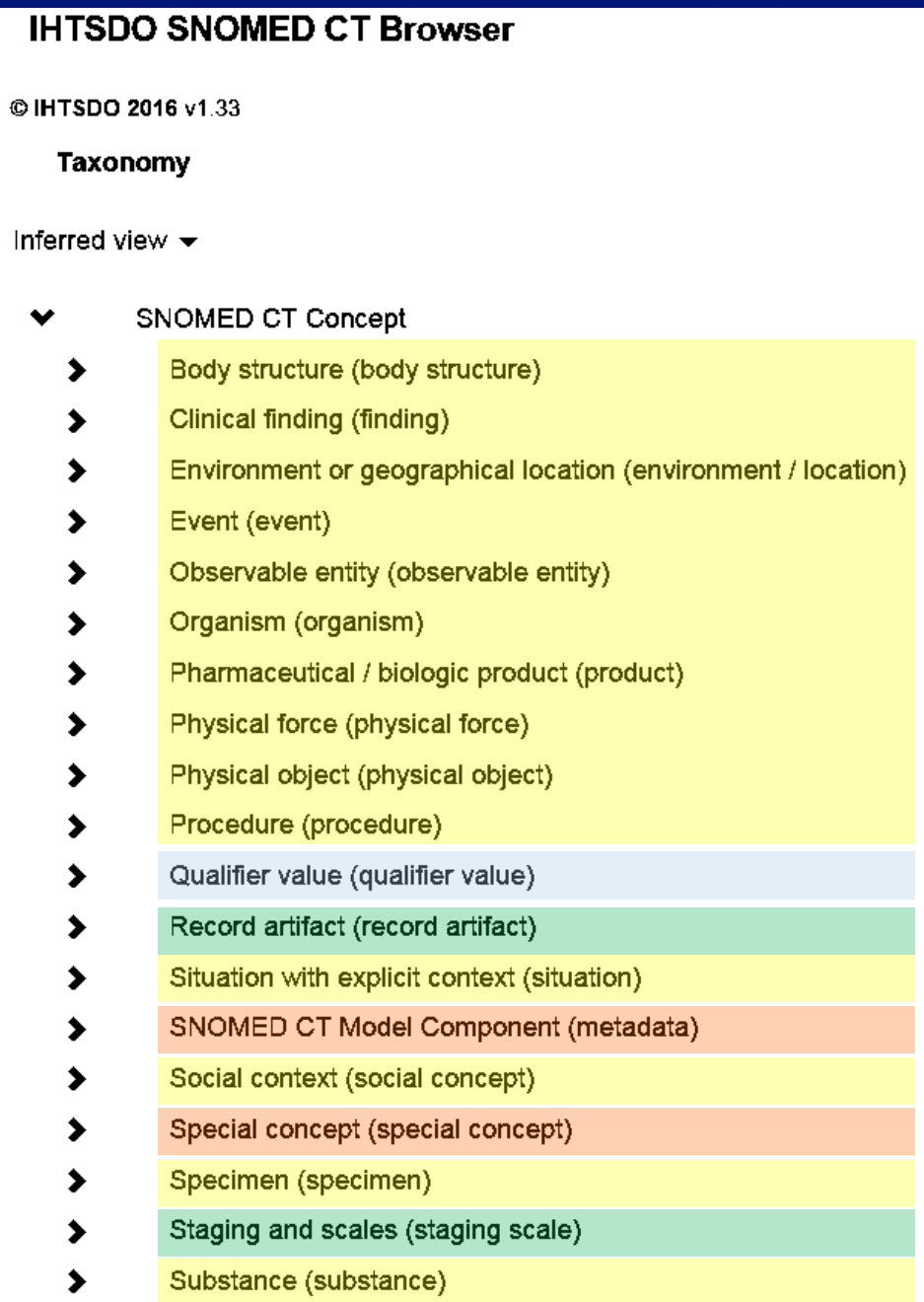
B. Entities to describe entities on the side of the patient



SNOMED CT's top categories

denote (roughly):

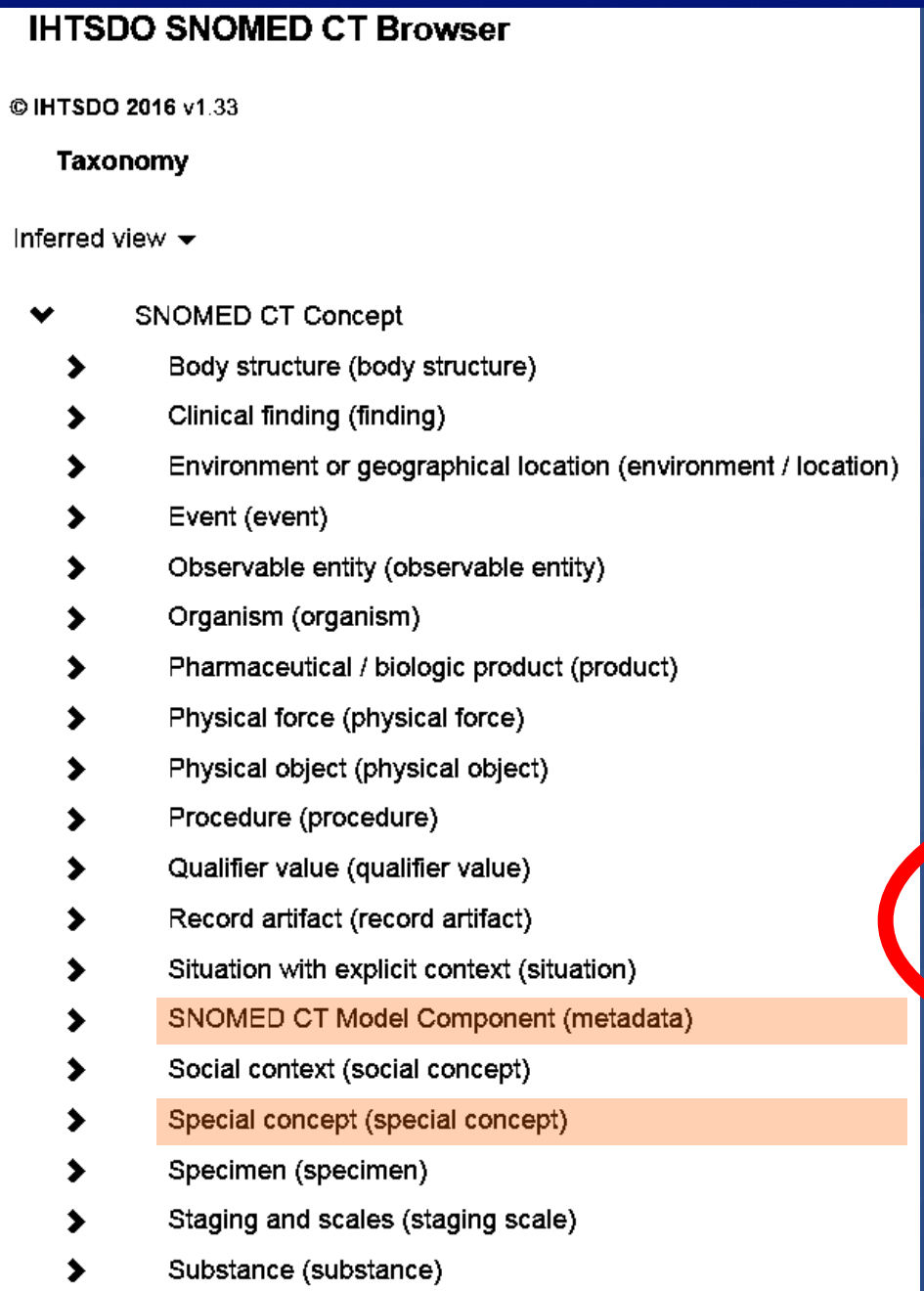
A. Entities on the side of the patient
B. Entities to describe entities on the side of the patient
C. Entities on the side of SNOMED CT as descriptive tool



SNOMED CT's top categories

denote (roughly):

A. Entities on the side of the patient
B. Entities to describe entities on the side of the patient
C. Entities on the side of SNOMED CT as descriptive tool
D. A hodge podge of A and B



SNOMED CT's top categories

- Describe
  - The structure of SNOMED CT
  - Changes introduced in any of the SNOMED CT components

C. Entities on the side of SNOMED CT as descriptive tool

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

Changes in concepts are enormous and not straightforward.

Research based on the January 2016 version of SNOMED CT

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

Whereas in 2002 14.7% of total concepts were inactive, it was already 24.8% in January 2016.

Research based on the January 2016 version of SNOMED CT

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

All versions exhibit activations and inactivations, but only few versions come with more inactivations than activations.

Research based on the January 2016 version of SNOMED CT

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

Of the 278,183 active concepts in 2002, only **66.7%** was (still) active in January 2016.

Research based on the January 2016 version of SNOMED CT

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

Of the 47,833 inactive concepts in 2002, **160** became re-activated between 2002 and 2016.

Research based on the January 2016 version of SNOMED CT

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

Of all inactivated concepts, **651** became reactivated.

Research based on the January 2016 version of SNOMED CT

	From Concept file		From 2016 'snapshot' file	
Version	Activated In	Deactivated in	Active since	Inactive since
20020131	278183	47833	185564	47673
20020731	7485	1747	9950	1725
20030131	7806	5834	7151	5569
20030731	8124	1905	7684	1888
20040131	4573	2008	3899	1990
20040731	4599	1580	7462	1471
20050131	2817	897	6699	891
20050731	1722	438	4148	433
20060131	2422	718	6687	710
20060731	2125	1035	2295	1029
20070131	3036	1363	4717	1351
20070731	2325	907	4153	902
20080131	2073	1071	3403	1070
20080731	5126	889	11747	884
20090131	3745	9221	4214	9221
20090731	1337	3657	4773	3655
20100131	1744	18293	3702	18292
20100731	1157	289	2288	288
20110131	1906	211	3363	210
20110731	1969	189	2162	187
20120131	319	108	340	108
20120731	794	281	1124	279
20130131	1673	114	1841	112
20130731	1124	304	1307	304
20140131	2229	2466	2593	2466
20140731	12610	349	12580	348
20150131	3287	1131	3671	1131
20150731	4458	399	5309	399
20160131	3115	726	4620	726
SUMS	373883	105963	319446	105312

SNOMED CT concept (in)activations

A total of **153,830** activations and deactivations effectuated since the first version, resulted in:

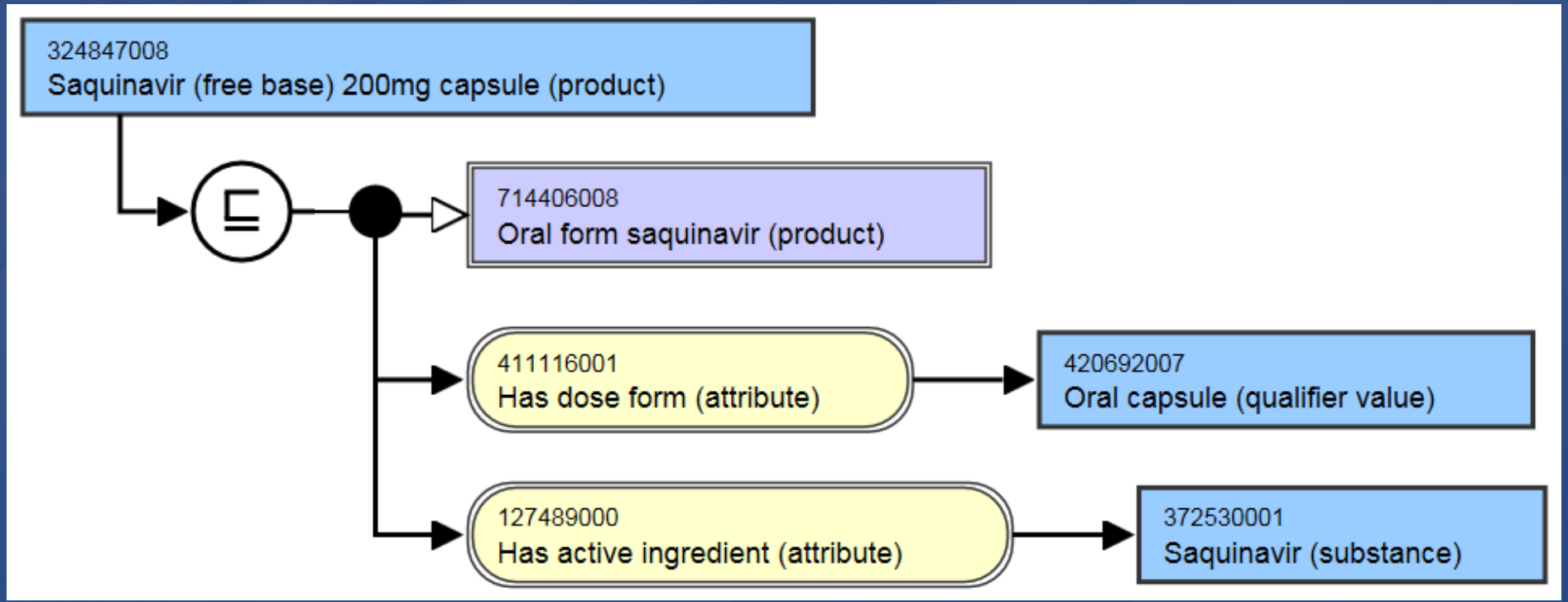
- an increase of **41,263** active concepts (→ a growth of merely **14.8%**), with
- an efficiency of only **26.9%**.

Research based on the January 2016 version of SNOMED CT

ConceptID	Version	Active
324847008	20020131	1
324847008	20030131	0
324847008	20040131	1
324847008	20040731	0
324847008	20050131	1
324847008	20050731	0
324847008	20070131	1

Several concepts exhibit multiple (in)activations:

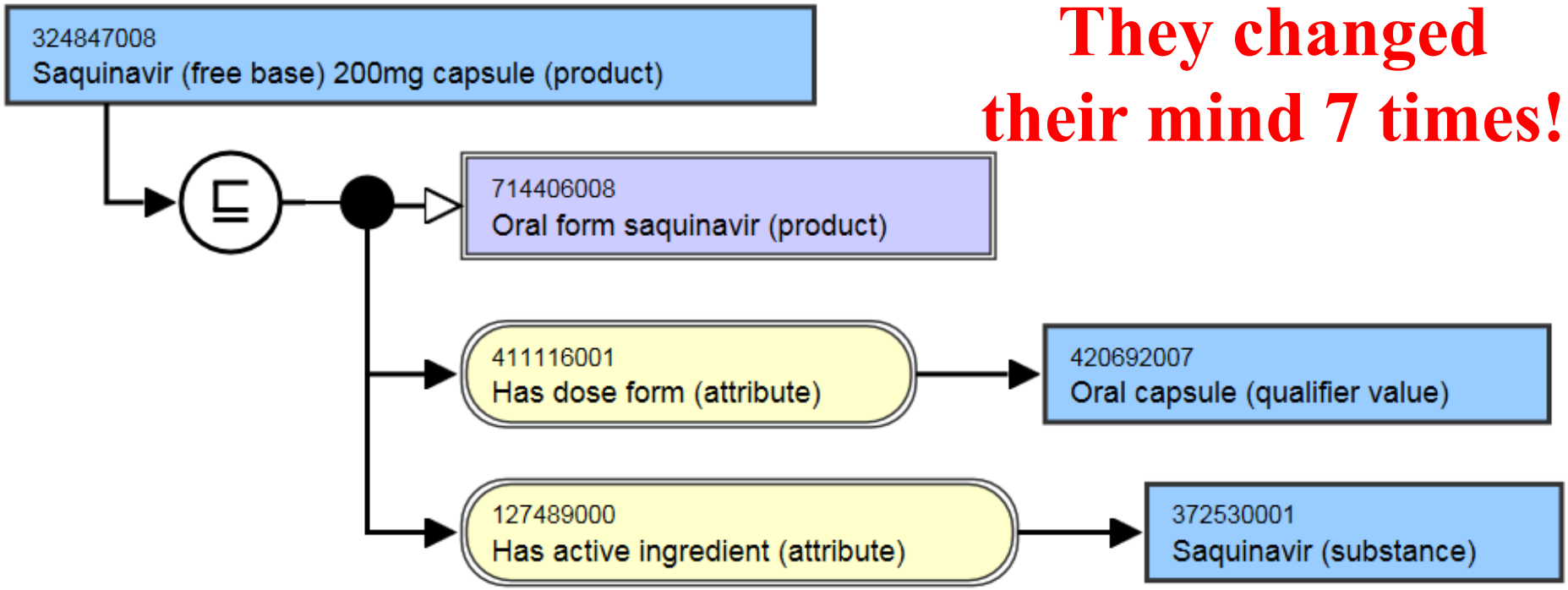
e.g. the history of Saquinovir



ConceptID	Version	Active
324847008	20020131	1
324847008	20030131	0
324847008	20040131	1
324847008	20040731	0
324847008	20050131	1
324847008	20050731	0
324847008	20070131	1

Several concepts exhibit multiple (in)activations:

e.g. the history of Saquinovir



They changed their mind 7 times!

History of '*Saquinavir 200mg capsule*'

ConceptID1	RelationshipType	ConceptID2	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11
324847008	SAME AS	375690004		Y	Y							
324848003	SAME AS	375690004		Y	Y							
375690004	SAME AS	324848003				Y	Y	Y	Y	Y	Y	
324847008	SAME AS	324848003					Y					
324847008	SAME AS	324848003							Y	Y	Y	
324845000	MAY BE A	422836001										Y
324845000	MAY BE A	324847008										Y
324848003	REPLACED BY	422836001										Y
375690004	REPLACED BY	422836001										Y

Saquinavir 200mg capsule
Saquinavir 200mg capsule (product)
Saquinavir 200mg capsule (substance)
Saquinavir (free base) 200mg capsule
Saquinavir (free base) 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (free base) 200mg capsule (substance)
Saquinavir mesylate 200mg capsule
Saquinavir 200mg capsule
Saquinavir (free base) 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir (as mesylate) 200mg capsule (substance)
Saquinavir (as mesylate) 200mg capsule
Saquinavir (as mesylate) 200mg capsule (product)
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (as mesylate) 200mg capsule

ConceptID	Version	Active
324847008	20020131	1
324847008	20030131	0
324847008	20040131	1
324847008	20040731	0
324847008	20050131	1
324847008	20050731	0
324847008	20070131	1

Research question

- Do assertions which are about changes in SNOMED CT exhibit patterns that would allow the detection of mistakes in assertions about external reality that have thus far not been discovered?

- In other words:

  what can we learn about SNOMED CT's mistakes committed in the past to detect still existing mistakes and prevent new ones?

Focus of research presented here

1. Networks of this sort

ConceptID1	RelationshipType	ConceptID2	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11
324847008	SAME AS	375690004		Y	Y							
324848003	SAME AS	375690004		Y	Y							
375690004	SAME AS	324848003				Y	Y	Y	Y	Y	Y	
324847008	SAME AS	324848003					Y					
324847008	SAME AS	324848003							Y	Y	Y	
324845000	MAY BE A	422836001										Y
324845000	MAY BE A	324847008										Y
324848003	REPLACED BY	422836001										Y
375690004	REPLACED BY	422836001										Y

Saquinavir 200mg capsule
Saquinavir 200mg capsule (product)
Saquinavir 200mg capsule (substance)
Saquinavir (free base) 200mg capsule
Saquinavir (free base) 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (free base) 200mg capsule (substance)
Saquinavir mesylate 200mg capsule
Saquinavir 200mg capsule
Saquinavir (free base) 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir (as mesylate) 200mg capsule (substance)
Saquinavir (as mesylate) 200mg capsule
Saquinavir (as mesylate) 200mg capsule (product)
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (as mesylate) 200mg capsule

Focus of research presented here

ConceptID1	RelationshipType	ConceptID2	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11
324847008	SAME AS	375690004		Y	Y							
324848003	SAME AS	375690004		Y	Y							
375690004	SAME AS	324848003				Y	Y	Y	Y	Y	Y	
324847008	SAME AS	324848003					Y					
324847008	SAME AS	324848003							Y	Y	Y	
324845000	MAY BE A	422836001										Y
324845000	MAY BE A	324847008										Y
324848003	REPLACED BY	422836001										Y
375690004	REPLACED BY	422836001										Y

Saquinavir 200mg capsul
Saquinavir 200mg capsu e (product)
Saquinavir 200mg capsu e (substance)
Saquinavir (free base) 200 g capsul
Saquinavir (free base) 200mg capsul (product)
Saquinavir mesylate 200mg capsule
Saquinavir (free base) 200mg capsu e (substance)
Saquinavir mesylate 200mg capsule
Saquinavir 200mg capsule
Saquinavir (free base) 200mg capsul
Saquinavir mesylate 200mg capsu e (product)
Saquinavir (as mesylate) 200mg apsule (substance)
Saquinavir (as mesylate) 200mg apsule
Saquinavir (as mesylate) 200mg ca sule (product)
Saquinavir mesylate 200mg capsule (p oduct)
Saquinavir mesylate 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (as mesylate) 200mg capsule

2. Changes in semantic tags

Focus 1: networks of modified SNOMED-concepts

1. Networks of this sort

ConceptID1	RelationshipType	ConceptID2	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11
324847008	SAME AS	375690004		Y	Y							
324848003	SAME AS	375690004		Y	Y							
375690004	SAME AS	324848003				Y	Y	Y	Y	Y	Y	
324847008	SAME AS	324848003					Y					
324847008	SAME AS	324848003							Y	Y	Y	
324845000	MAY BE A	422836001										Y
324845000	MAY BE A	324847008										Y
324848003	REPLACED BY	422836001										Y
375690004	REPLACED BY	422836001										Y

Input:

- *Historical Association Reference Sets* (HARS) and,
- *Component Inactivation Reference* Sets (CIRS).

Methodology:

- Build history profiles of concept modifications,
- Use Formal Concept Analysis (FCA) to calculate dependencies.

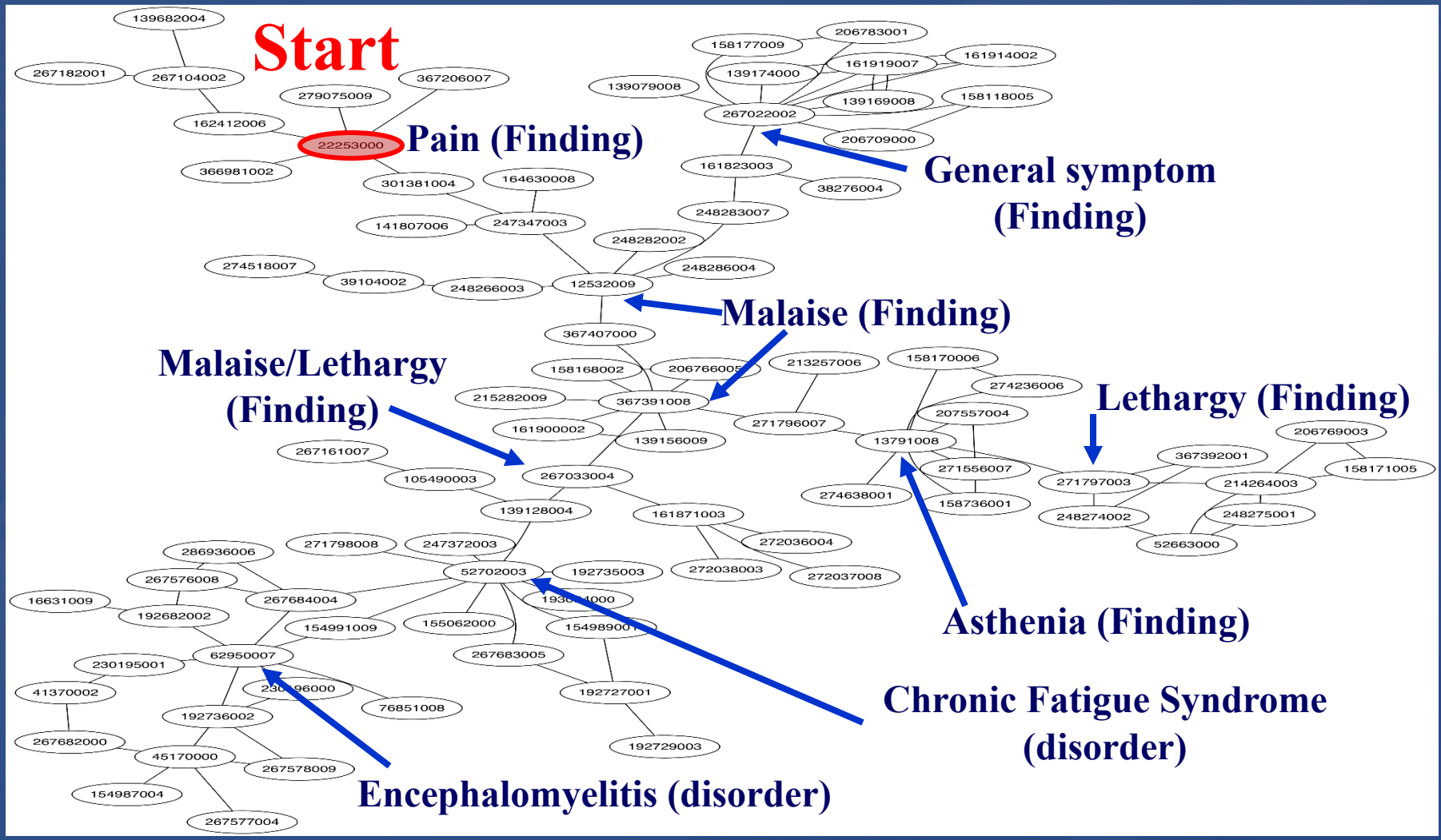
Historical association reference set types

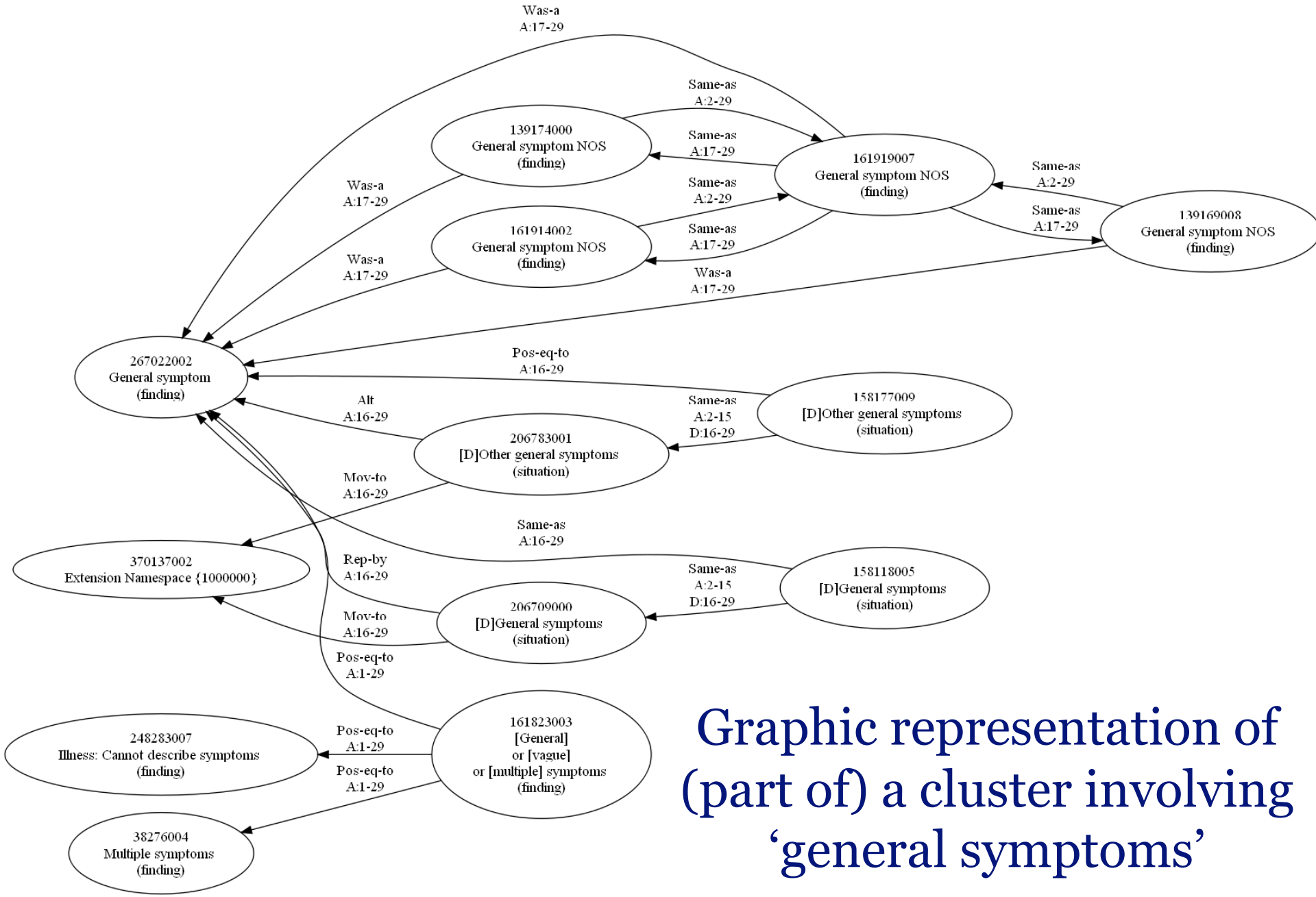
HARS name	Use
Possibly equivalent to (P)	From an ambiguous concept to one or more active concepts that represents one of the possible meanings of the inactive concept.
Moved to (T)	From a component to a namespace to which the component has been moved
Moved from (F)	From a namespace to the original component Identifier in its previous namespace.
Replaced by (R)	From an erroneous or obsolete inactive component to a single active replacement component.
Same as (S)	From a duplicate component to the active component that this component duplicates.
Was a (W)	From an inactive classification concept such as "not otherwise specified" to the active concept that was formerly its most proximal supertype.
Alternative (Z)	From an inactive classification concept derived from ICD-9 Chapter XVI 'Symptoms signs and ill-defined conditions' with the most similar active concept.
Refers to	From an inactive description which is inappropriate to the concept it is directly linked to but instead should refer to the concept referenced.

Component inactivation set types for concepts

CIRS value	Concept status and motivation
Duplicate (D)	inactive because it has the same meaning as another Concept
Outdated (O)	inactive because it is an outdated concept that is no longer used.
Ambiguous (A)	inactive because it is inherently ambiguous either because of an incomplete FSN or because it has several associated terms that are not regarded as synonymous or partial synonymous.
Erroneous (E)	inactive because it contains an error
Limited (L)	active prior to Jan 2010, inactive since then because of unstable meaning within SNOMED CT
moved to (M)	inactive because moved to another namespace.
Pending move	active but in the process of being moved to another namespace

Generation of clusters from HARS





Graphic representation of (part of) a cluster involving ‘general symptoms’

Construction of History Profiles for clusters

Row	Concept ID	Attribute	Value	History profile (one character per version)
1	324253001	Duplicate		AAYYNNNNNNNNNNNNNNNNNNNNNNNN
3	Azithromycin 200mg/5mL oral suspension (product)	Duplicated by	375558000	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
3		Duplicated by	375559008	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
4		Duplicated by	375948007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
5		Duplicated by	376025007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
6		Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYYY
7		Semantic tag	substance	YNNNNNNNNNNNNNNNNNNNNNNNNNNN
8		Is active		YYNNYYYYYYYYYYYYYYYYYYYYYYYYY
9		Same-as	375559008	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
10		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
11	375558000	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
12	Azithromycin dihydrate 200mg/5mL suspension (product)	Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
13		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
14		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
15		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
16	375559008	Duplicate		AAAYYYYYYYYYYYYYYYYYYYYYYYYYY
17	Azithromycin dihydrate 200mg/ 5 mL suspension (product)	Duplicated by	324253001	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
18		Is active		AYYNNNNNNNNNNNNNNNNNNNNNNNNN
19		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
20		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
21	375948007	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
22	Azithromycin dihydrate 200mg/5 mL suspension (product)	Duplicated by	324253001	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
23		Duplicated by	375558000	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
24		Duplicated by	375559008	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
25		Duplicated by	376025007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
26		Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
27		Is active		AYYYNNNNNNNNNNNNNNNNNNNNNNNN
28		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
29	376025007	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
30	Azithromycin dihydrate 200mg/5 mL powder (product)	Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
31		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
32		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
33		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN

Construction of History Profiles for clusters

Row	Concept ID	Attribute	Value	History profile (one character per version)
1	324253001	Duplicate		AAYYNNNNNNNNNNNNNNNNNNNNNNNN
3	Azithromycin 200mg/5mL oral suspension (product)	Duplicated by	375558000	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
3		Duplicated by	375559008	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
4		Duplicated by	375948007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
5		Duplicated by	376025007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
6		Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYYY
7		Semantic tag	substance	YNNNNNNNNNNNNNNNNNNNNNNNNNNN
8		Is active		YYNNYYYYYYYYYYYYYYYYYYYYYYYYY
9		Same-as	375559008	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
10		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
11	375558000	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
12	Azithromycin dihydrate 200mg/5mL suspension (product)	Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
13		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
14		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
15		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
16	375559008	Duplicate		AAAYYYYYYYYYYYYYYYYYYYYYYYYY
17	Azithromycin dihydrate 200mg/ 5 mL suspension (product)	Duplicated by	324253001	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
18		Is active		AYYNNNNNNNNNNNNNNNNNNNNNNNNN
19		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
20		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
21	375948007	Duplicate		
22	Azithromycin dihydrate 200mg/5 mL suspension (product)	Duplicated by	324253001	
23		Duplicated by	375558000	
24		Duplicated by	375559008	
25		Duplicated by	376025007	
26		Semantic tag	Product	
27		Is active		
28		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
29	376025007	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
30	Azithromycin dihydrate 200mg/5 mL powder (product)	Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
31		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
32		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
33		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN

Concept 375559008 not present in 1st version, but introduced as active in the 2nd version.

Construction of History Profiles for clusters

Row	Concept ID	Attribute	Value	History profile (one character per version)
1	324253001	Duplicate		AAYYNNNNNNNNNNNNNNNNNNNNNNNN
3	Azithromycin 200mg/5mL oral suspension (product)	Duplicated by	375558000	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
3		Duplicated by	375559008	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
4		Duplicated by	375948007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
5		Duplicated by	376025007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
6		Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYYY
7		Semantic tag	substance	YNNNNNNNNNNNNNNNNNNNNNNNNNNN
8		Is active		YYNNYYYYYYYYYYYYYYYYYYYYYYYYY
9		Same-as	375559008	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
10		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
11	375558000	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
12	Azithromycin dihydrate 200mg/5mL suspension (product)	Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
13		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
14		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
15		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
16	375559008	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
17	Azithromycin dihydrate 200mg/ 5 mL suspension (product)	Duplicated by	324253001	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
18		Is active		AYYNNNNNNNNNNNNNNNNNNNNNNNNN
19		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
20		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
21	375948007	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
22	Azithromycin dihydrate 200mg/5 mL suspension (product)	Duplicated by	324253001	
23		Duplicated by	375558000	
24		Duplicated by	375559008	
25		Duplicated by	376025007	
26		Semantic tag	Product	
27		Is active		
28		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
29	376025007	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
30	Azithromycin dihydrate 200mg/5 mL powder (product)	Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
31		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
32		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
33		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN

Concept 375559008 in the 3rd version recognized as being duplicated by 324253001

Construction of History Profiles for clusters

Row	Concept ID	Attribute	Value	History profile (one character per version)
1	324253001	Duplicate		AAYYNNNNNNNNNNNNNNNNNNNNNNNN
3	Azithromycin 200mg/5mL oral suspension (product)	Duplicated by	375558000	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
3		Duplicated by	375559008	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
4		Duplicated by	375948007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
5		Duplicated by	376025007	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
6		Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYYY
7		Semantic tag	substance	YNNNNNNNNNNNNNNNNNNNNNNNNNNN
8		Is active		YYNN
9		Same-as	375559008	AAYN
10		Same-as	375948007	AAAY
11	375558000	Duplicate		AAYY
12	Azithromycin dihydrate 200mg/5mL suspension (product)	Semantic tag	product	AYYY
13		Is active		AYNN
14		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
15		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
16	375559008	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
17	Azithromycin dihydrate 200mg/ 5 mL suspension (product)	Duplicated by	324253001	AAYNNNNNNNNNNNNNNNNNNNNNNNNN
18		Is active		AYYNNNNNNNNNNNNNNNNNNNNNNNNN
19		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
20		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
21	375948007	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
22	Azithromycin dihydrate 200mg/5 mL suspension (product)	Duplicated by	324253001	N
23		Duplicated by	375558000	N
24		Duplicated by	375559008	N
25		Duplicated by	376025007	N
26		Semantic tag	Product	Y
27		Is active		N
28		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
29	376025007	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
30	Azithromycin dihydrate 200mg/5 mL powder (product)	Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
31		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
32		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYYY
33		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN

Concept 324253001 deactivated in 3rd version and declared to be the same as 375559008

Concept 375559008 in the 3rd version recognized as being duplicated by 324253001

Certain profile stages (should) occur *in tandem*

Row	Concept ID	Attribute	Value	History profile (one character per version)
1	324253001	Duplicate		AAYYNNNNNNNNNNNNNNNNNNNNNNNN
3	Azithromycin 200mg/5mL oral suspension (product)	Duplicated by	375558000	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
3		Duplicated by	375559008	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
4		Duplicated by	375948007	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
5		Duplicated by	376025007	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
6		Semantic tag	product	A
7		Semantic tag	substance	Y
8		Is active		Y
9		Same-as	375559008	A
10		Same-as	375948007	A
11	375558000	Duplicate		A
12	Azithromycin dihydrate 200mg/5mL suspension (product)	Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
13		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNN
14		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
15		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
16	375559008	Duplicate		AAAYYYYYYYYYYYYYYYYYYYYYYYYY
17	Azithromycin dihydrate 200mg/ 5 mL suspension (product)	Duplicated by	324253001	AAYNNNNNNNNNNNNNNNNNNNNNNNN
18		Is active		AYYNNNNNNNNNNNNNNNNNNNNNNNN
19		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
20		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
21	375948007	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYY
22	Azithromycin dihydrate 200mg/5 mL suspension (product)	Duplicated by	324253001	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
23		Duplicated by	375558000	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
24		Duplicated by	375559008	AAAYNNNNNNNNNNNNNNNNNNNNNNNN
25		Duplicated by	376025007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN
26		Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
27		Is active		AYYYNNNNNNNNNNNNNNNNNNNNNNNN
28		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
29	376025007	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYYY
30	Azithromycin dihydrate 200mg/5 mL powder (product)	Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYYY
31		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNN
32		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
33		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNNN

It is not possible for a concept to duplicate one and be duplicated at the same time

Certain profile stages (should) occur *in tandem*

Row	Concept ID	Attribute	Value	History profile (one character per version)
1	324253001	Duplicate		AAYYNNNNNNNNNNNNNNNNNNNNNNN
3	Azithromycin 200mg/5mL oral suspension (product)	Duplicated by	375558000	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
3		Duplicated by	375559008	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
4		Duplicated by	375948007	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
5		Duplicated by	376025007	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
6		Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYY
7		Semantic tag	substance	YNNNNNNNNNNNNNNNNNNNNNNNNNN
8		Is active		YYNNYYYYYYYYYYYYYYYYYYYYYYYY
9		Same-as	375559008	AAYNNNNNNNNNNNNNNNNNNNNNNNN
10		Same-as	375948007	AAAYNNNNNNNNNNNNNNNNNNNNNNN
11	375558000	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYY
12	Azithromycin dihydrate 200mg/5mL suspension (product)	Semantic tag	product	AYYYYYYYYYYYYYYYYYYYYYYYYYY
13		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
14		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
15		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNN
16	375559008	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYY
17	Azithromycin dihydrate 200mg/ 5 mL suspension (product)	Duplicated by	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
18		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
19		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
20		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNN
21	375948007	Duplicate		AAAAYYYYYYYYYYYYYYYYYYYYYYYY
22	Azithromycin dihydrate 200mg/5 mL suspension (product)	Duplicated by	324253001	AAAYNNNNNNNNNNNNNNNNNNNNNNN
23		Duplicated by	375558000	AAYYNNNNNNNNNNNNNNNNNNNNNNN
24		Duplicated by	375559008	AAAYNNNNNNNNNNNNNNNNNNNNNNN
25		Duplicated by	376025007	AAYYNNNNNNNNNNNNNNNNNNNNNNN
26		Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYY
27		Is active		AYYYNNNNNNNNNNNNNNNNNNNNNNN
28		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
29	376025007	Duplicate		AAYYYYYYYYYYYYYYYYYYYYYYYYY
30	Azithromycin dihydrate 200mg/5 mL powder (product)	Semantic tag	Product	AYYYYYYYYYYYYYYYYYYYYYYYYYY
31		Is active		AYNNNNNNNNNNNNNNNNNNNNNNNNN
32		Same-as	324253001	AAAAYYYYYYYYYYYYYYYYYYYYYYYY
33		Same-as	375948007	AAYYNNNNNNNNNNNNNNNNNNNNNNN

If a concept is the same as another one, it is inactive.

Formal Concept Analysis (FCA)

- **Goal of FCA**: build lattice from data tables that represent binary relations between objects and attributes, thus tabulating pairs of the form 'object g has attribute m'.
- A formal concept is defined to be a pair (A, B), where
  - A is a set of objects (called the *extent*) and
  - B is a set of attributes (the *intent*)
- such that
  - the extent A consists of all objects that share the attributes in B, and dually
  - the intent B consists of all attributes shared by the objects in A.

https://en.wikipedia.org/wiki/Formal_concept_analysis

Easy example: FCA on the numbers 1 … 10

	composite	even	odd	prime	square
1			✓		✓
2		✓		✓	
3			✓	✓	
4	✓	✓			✓
5			✓	✓	
6	✓	✓			
7			✓	✓	
8	✓	✓			
9	✓		✓		✓
10	✓	✓			

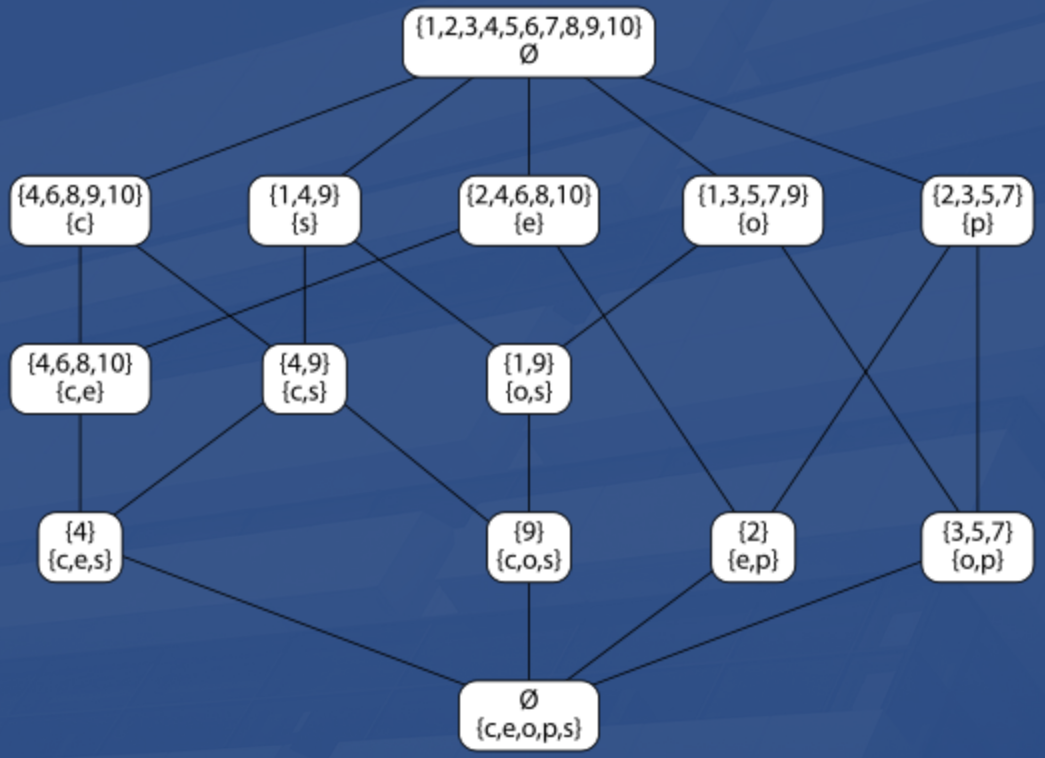
attributes

objects

https://en.wikipedia.org/wiki/Formal_concept_analysis

FCA lattice on the numbers 1 … 10

	composite	even	odd	prime	square
1			✓		✓
2		✓		✓	
3			✓	✓	
4	✓	✓			✓
5			✓	✓	
6	✓	✓			
7			✓	✓	
8	✓	✓			
9	✓		✓		✓
10	✓	✓			



https://en.wikipedia.org/wiki/Formal_concept_analysis

FCA for calculating ***valid*** implications

	composite	even	odd	prime	square
1			✓		✓
2		✓		✓	
3			✓	✓	
4	✓	✓			✓
5			✓	✓	
6	✓	✓			
7			✓	✓	
8	✓	✓			
9	✓		✓		✓
10	✓	✓			

- Whenever a number 1…10 is ***odd***, it is ***not even***, and vice versa

https://en.wikipedia.org/wiki/Formal_concept_analysis

FCA for calculating ***valid*** implications

	composite	even	odd	prime	square
1			✓		✓
2		✓		✓	
3			✓	✓	
4	✓	✓			✓
5			✓	✓	
6	✓	✓			
7			✓	✓	
8	✓	✓			
9	✓		✓		✓
10	✓	✓			

- Whenever a number 1…10 is ***composite*** and ***odd***, it is ***square***.

https://en.wikipedia.org/wiki/Formal_concept_analysis

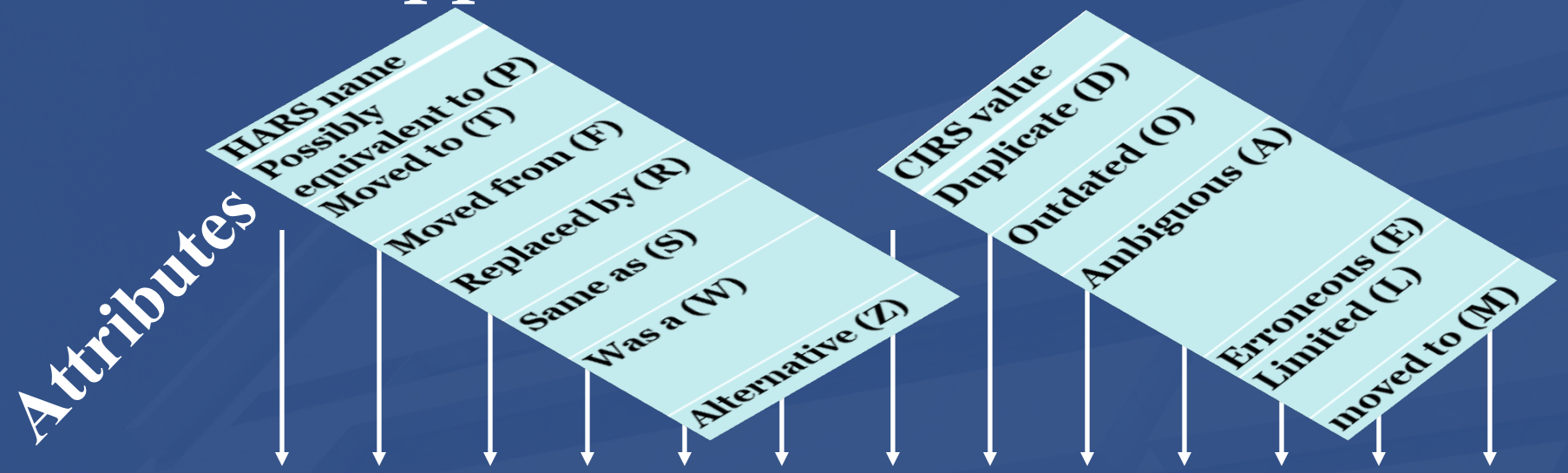
FCA for calculating ***approximate*** implications

	composite	even	odd	prime	square
1			✓		✓
2		✓		✓	
3			✓	✓	
4	✓	✓			✓
5			✓	✓	
6	✓	✓			
7			✓	✓	
8	✓	✓			
9	✓		✓		✓
10	✓	✓			

- 80% of the numbers 1…10 that are ***even*** are also ***composite***.

https://en.wikipedia.org/wiki/Formal_concept_analysis

FCA applied to HARS and CIRS (1)



SNOMED CT concepts

M	T	A	P	R	S	D	O	W	Z	L	F	E
1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0
1	1	0	0	0	1	1	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0
1	1	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	1	0	0	0	0	0
0	0	1	1	0	1	1	0	0	0	0	0	0

...

FCA applied to HARS and CIRS (2)

99,489 referenced concepts in HARS and CIRS reduce to 85 FCA-concepts

FCA-concepts

	M	T	A	P	R	S	D	O	W	Z	L	F	E
MT	1	1	0	0	0	0	0	0	0	0	0	0	0
AP	0	0	1	1	0	0	0	0	0	0	0	0	0
R	0	0	0	0	1	0	0	0	0	0	0	0	0
RS	0	0	0	0	1	1	0	0	0	0	0	0	0
DMTS	1	1	0	0	0	1	1	0	0	0	0	0	0
DS	0	0	0	0	0	1	1	0	0	0	0	0	0
DMT	1	1	0	0	0	0	1	0	0	0	0	0	0
OR	0	0	0	0	1	0	0	1	0	0	0	0	0
DAPS	0	0	1	1	0	1	1	0	0	0	0	0	0
ER	0	0	0	0	1	0	0	0	0	0	0	0	1
EW	0	0	0	0	0	0	0	0	1	0	0	0	1
PR	0	0	0	1	1	0	0	0	0	0	0	0	0
LW	0	0	0	0	0	0	0	0	1	0	1	0	0
APRS	0	0	1	1	1	1	0	0	0	0	0	0	0

… → 85

85 FCA-concepts in HARS and CIRS

1	2		3		4		5	6	7
R	MT	OW	DMT	ALP	DMTS	OAMT	DALPS	DAEPRS	DAEPRSW
O	AP	LR	DRS	DER	DAPS	DMPT	DELRS	DLMTSW	
L	RS	AR	ORS	OAP	APRS	ALPW	DERSW	DALPSW	
S	DS	OL	APR	DOS	DORS	ELRW	DAMPT	AELPRW	
D	OR	OA	DLS	AMT	OAPR	OLRW	AELPR		
A	ER	RW	ERS	DAS	DLSW	LMTR	DALSW		
W	EW	DR	DAP	MTZ	AEPR	LMTZ	LMTSW		
E	PR	DW	ERW	APS	DLRS	DERW	DLMTS		
F	LW		LMT	MTR	DERS	EMTR	DOERS		
			ELR	DFS	AMPT	OMTR	ALPSW		
			LSW	OER			ALMPT		
							DAPRS		

There are patterns in the 85 FCA-concepts

1	2		3		4		5	6	7
R	MT	OW	DMT	ALP	DMTS	OAMT	DALPS	DAEPRS	DAEPRSW
O	AP	LR	DRS	DER	DAPS	DMPT	DELRS	DLMTSW	
L	RS	AR	ORS	OAP	APRS	ALPW	DERSW	DALPSW	
S	DS	OL	APR	DOS	DORS	ELRW	DAMPT	AELPRW	
D	OR	OA	DLS	AMT	OAPR	OLRW	AELPR		
A	ER	RW	ERS	DAS	DLSW	LMTR	DALSW		
W	EW	DR	DAP	MTZ	AEPR	LMTZ	LMTSW		
E	PR	DW	ERW	APS	DLRS	DERW	DLMTS		
F	LW		LMT	MTR	DERS	EMTR	DOERS		
			ELR	DFS	AMPT	OMTR	ALPSW		
			LSW	OER			ALMPT		
							DAPRS		

Example: whenever a concept's history profile has a marker for it to be (or have been) a member of a 'moved-to' CIRS ('M'), it is also marked as being (or having been) a member of a 'moved-to' HARS ('T').

FCA applied to HARS and CIRS (3)

Positive implications from the Duquenne-Guigues base, i.e. those implications from which all other valid implications follow semantically:

< 19 > M ==> T;
< 19 > T ==> M;
< 9 > P S ==> A;
< 4 > A R S ==> P;
< 1 > M T A D ==> P;
< 3 > A R D ==> P S;
< 3 > P R D ==> A S;
< 2 > P O ==> A;
< 1 > A R O ==> P;
< 3 > D O ==> S;
< 2 > M T W ==> S L;
< 5 > P W ==> A;
< 3 > A D W ==> S;
< 1 > R O W ==> L;
< 2 > Z ==> M T;
< 1 > M T A L ==> P;
< 8 > P L ==> A;
< 2 > R S L ==> D;
< 9 > D L ==> S;
< 1 > R O L ==> W;
< 1 > O W L ==> R;
< 1 > S F ==> D;
< 1 > D F ==> S;
< 2 > A R L ==> P E;
< 1 > M T E ==> R;
< 5 > A E ==> P R;
< 5 > P E ==> A R;
< 7 > S E ==> R;
< 2 > A P R S E ==> D;
< 8 > D E ==> R;
< 2 > O E ==> R;
< 1 > R S O E ==> D;
< 5 > L E ==> R;
< 2 > A R W ==> P E;
< 2 > R S W ==> D E;
< 3 > R D W ==> E;

FCA applied to HARS and CIRS (3)

Positive implications from the Duquenne-Guigues base, i.e. those implications from which all other valid implications follow semantically:

< 19 > M ==> T;
< 19 > T ==> M;
< 9 > P S ==> A;
< 4 > A R S ==> P;
< 1 > M T A D ==> P;
< 3 > A R D ==> P S;
< 3 > P R D ==> A S;
< 2 > P O ==> A;
< 1 > A R O ==> P;
< 3 > D O ==> S;
< 2 > M T W ==> S L;
< 5 > P W ==> A;
< 3 > A D W ==> S;
< 1 > R O W ==> L;
< 2 > Z ==> M T;
< 1 > M T A L ==> P;
< 8 > P L ==> A;
< 2 > R S L ==> D;
< 9 > D L ==> S;
< 1 > R O L ==> W;
< 1 > O W L ==> R;
< 1 > S F ==> D;
< 1 > D F ==> S;
< 2 > A R L ==> P E;
< 1 > M T E ==> R;
< 5 > A E ==> P R;
< 5 > P E ==> A R;
< 7 > S E ==> R;
< 2 > A P R S E ==> D;
< 8 > D E ==> R;
< 2 > O E ==> R;
< 1 > R S O E ==> D;
< 5 > L E ==> R;
< 2 > A R W ==> P E;
< 2 > R S W ==> D E;
< 3 > R D W ==> E;

Implication formulation of *'whenever a concept's history profile has a marker for it to be (or have been) a member of a 'moved-to' CIRS ('M'), it is also marked as being (or having been) a member of a 'moved-to' HARS ('T')'.*

FCA applied to HARS and CIRS (3)

Positive implications from the Duquenne-Guigues base, i.e. those implications from which all other valid implications follow semantically:

< 19 > M ==> T;
< 19 > T ==> M;
< 9 > P S ==> A;
< 4 > A R S ==> P;
< 1 > M T A D ==> P;
< 3 > A R D ==> P S;
< 3 > P R D ==> A S;
< 2 > P O ==> A;
< 1 > A R O ==> P;
< 3 > D O ==> S;
< 2 > M T W ==> S L;
< 5 > P W ==> A;

< 3 > A D W ==> S;
< 1 > R O W ==> L;
< 2 > Z ==> M T;
< 1 > M T A L ==> P;
< 8 > P L ==> A;
< 2 > R S L ==> D;
< 9 > D L ==> S;
< 1 > R O L ==> W;
< 1 > O W L ==> R;
< 1 > S F ==> D;
< 1 > D F ==> S;
< 2 > A R L ==> P E;

< 1 > M T E ==> R;
< 5 > A E ==> P R;
< 5 > P E ==> A R;
< 7 > S E ==> R;
< 2 > A P R S E ==> D;
< 8 > D E ==> R;
< 2 > O E ==> R;
< 1 > R S O E ==> D;
< 5 > L E ==> R;
< 2 > A R W ==> P E;
< 2 > R S W ==> D E;
< 3 > R D W ==> E;

states that if a SNOMED CT concept has ever been annotated as being ambiguous, duplicate and enjoying a was-a association to some other SNOMED CT concept, then it is also the case that this concept has been annotated as having a same-as association.

FCA applied to HARS and CIRS (3)

Positive implications from the Duquenne-Guigues base, i.e. those implications from which all other valid implications follow semantically

< 19 >	M ==> T;	< 3 >	A D W ==> S;	< 1 >	M T E ==> R;
< 19 >	T ==> M;	< 1 >	R O W ==> L;	< 5 >	A E ==> P R;
< 9 >	P S ==> A;	< 2 >	Z ==> M T;	< 5 >	P E ==> A R;
< 4 >	A R S ==> P;	< 1 >	M T A L ==> P;	< 7 >	S E ==> R;
< 1 >	M T A D ==> P;	< 8 >	P L ==> A;	< 2 >	A P R S E ==> D;
< 3 >	A R D ==> P S;	< 2 >	R S L ==> D;	< 8 >	D E ==> R;
< 3 >	P R D ==> A S;	< 9 >	D L ==> S;	< 2 >	O E ==> R;
< 2 >	P O ==> A;	< 1 >	R O L ==> W;	< 1 >	R S O E ==> D;
< 1 >	A R O ==> P;	< 1 >	O W L ==> R;	< 5 >	L E ==> R;
< 3 >	D O ==> S;	< 1 >	S F ==> D;	< 2 >	A R W ==> P E;
< 2 >	M T W ==> S L;	< 1 >	D F ==> S;	< 2 >	R S W ==> D E;
< 5 >	P W ==> A;	< 2 >	A R L ==> P E;	< 3 >	R D W ==> E;

Numbers indicate to how many of the 85 FCA-concepts this implication applies.

FCA applied to HARS and CIRS (3)

< 19 > M ==> T;
< 19 > T ==> M;

1	2		3		4		5	6	7
R	MT	OW	DMT	ALP	DMTS	OAMT	DALPS	DAEPRS	DAEPRSW
O	AP	LR	DRS	DER	DAPS	DMPT	DELRS	DLMTSW	
L	RS	AR	ORS	OAP	APRS	ALPW	DERSW	DALPSW	
S	DS	OL	APR	DOS	DORS	ELRW	DAMPT	AELPRW	
D	OR	OA	DLS	AMT	OAPR	OLRW	AELPR		
A	ER	RW	ERS	DAS	DLSW	LMTR	DALSW		
W	EW	DR	DAP	MTZ	AEPR	LMTZ	LMTSW		
E	PR	DW	ERW	APS	DLRS	DERW	DLMTS		
F	LW		LMT	MTR	DERS	EMTR	DOERS		
			ELR	DFS	AMPT	OMTR	ALPSW		
			LSW	OER			ALMPT		
							DAPRS		

1 + 2 + 3 + 2 + 6 + 4 + 1

Numbers indicate to how many of the 85 FCA-concepts this implication applies.

FCA applied to HARS and CIRS (4)

< 24 > P =[92%]=> < 22 > A;
< 20 > E =[90%]=> < 18 > R;
< 10 > P R =[90%]=> < 9 > A;
< 10 > A R =[90%]=> < 9 > P;
< 9 > A L =[89%]=> < 8 > P;
< 9 > P D =[89%]=> < 8 > A;
< 7 > W E =[86%]=> < 6 > R;
< 7 > R S E =[86%]=> < 6 > D;
< 6 > A W =[83%]=> < 5 > L;
< 6 > A W =[83%]=> < 5 > P;
< 11 > A S =[82%]=> < 9 > P;
< 5 > A W L =[80%]=> < 4 > P;

Approximate implications
('Luxenburger' base) that are valid for at least 80% of FCA concepts which have the antecedents as part of their attributes.

FCA applied to HARS and CIRS (5)

< 24 > P =[92%]=> < 22 > A;
< 20 > E =[90%]=> < 18 > R;
< 10 > P R =[90%]=> < 9 > A;
< 10 > A R =[90%]=> < 9 > P;
< 9 > A L =[89%]=> < 8 > P;
< 9 > P D =[89%]=> < 8 > A;
< 7 > W E =[86%]=> < 6 > R;
< 7 > R S E =[86%]=> < 6 > D;
< 6 > A W =[83%]=> < 5 > L;
< 6 > A W =[83%]=> < 5 > P;
< 11 > A S =[82%]=> < 9 > P;
< 5 > A W L =[80%]=> < 4 > P;

Example

7 out of 85 FCA-concepts have the attributes E(rroneous), R(eplaced-by), and S(ame as). 6 of these have also D(uplicate).

Focus 2: Changes in semantic tags

Saquinavir 200mg capsule
Saquinavir 200mg capsule (product)
Saquinavir 200mg capsule (substance)
Saquinavir (free base) 200mg capsule
Saquinavir (free base) 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (free base) 200mg capsule (substance)
Saquinavir mesylate 200mg capsule
Saquinavir 200mg capsule
Saquinavir (free base) 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir (as mesylate) 200mg capsule (substance)
Saquinavir (as mesylate) 200mg capsule
Saquinavir (as mesylate) 200mg capsule (product)
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir mesylate 200mg capsule (product)
Saquinavir mesylate 200mg capsule
Saquinavir (as mesylate) 200mg capsule

'Semantic tags' (per SNOMED CT documentation)

- Descriptions provide for each concept a Fully Specified Name (FSN) most of which '***end with a semantic tag in parentheses and which indicates the semantic category to which the concept belongs (e.g. clinical finding, disorder, procedure, organism, person, etc.)***' [4, p41].

IHTSDO. International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English). 2015. p. 757.

Semantic tags

Options

Search Mode: Full text search mode

Status: Active components only

Group by concept

Filter results by Language

english	618

Filter results by Semantic Tag

disorder	445
procedure	151
morphologic abnormality	12
situation	6
finding	2
specimen	2

Type at least 3 characters ✔ Example: *blistered finger*

hematoma

618 matches found in 0.146 seconds.

Hematoma	Hematoma (morphologic abnormality)
Hematoma	Hematoma (disorder)
Deep hematoma	Deep hematoma (disorder)
Liver hematoma	Liver hematoma (disorder)
Cecal hematoma	Cecal hematoma (disorder)
Wound hematoma	Wound hematoma (finding)
Hematoma block	Fracture infiltration with local anesthetic (procedure)
Rectal hematoma	Rectal hematoma (disorder)
Facial hematoma	Hematoma of face (disorder)
Hematoma sample	Hematoma sample (specimen)
Vulval hematoma	Hematoma of vulva (disorder)
Pelvic hematoma	Pelvic hematoma (disorder)

'Semantic tags' (per SNOMED CT documentation)

- Descriptions provide for each concept a Fully Specified Name (FSN) most of which '***end with a semantic tag in parentheses and which indicates the semantic category to which the concept belongs (e.g. clinical finding, disorder, procedure, organism, person, etc.)***' [4, p41].
- '***the semantic tag helps to disambiguate different concepts which may be referred to by the same commonly used word or phrase***' [4, p41].

IHTSDO. International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English). 2015. p. 757.

'Semantic tags' (per SNOMED CT documentation)

- Descriptions provide for each concept a Fully Specified Name (FSN) most of which '***end with a semantic tag in parentheses and which indicates the semantic category to which the concept belongs (e.g. clinical finding, disorder, procedure, organism, person, etc.)***' [4, p41].
- '***the semantic tag helps to disambiguate different concepts which may be referred to by the same commonly used word or phrase***' [4, p41].
- For example, it is the semantic tag 'morphologic abnormality' in the FSN 'Hematoma (morphologic abnormality)' that disambiguates the concept to which this FSN is assigned from a second concept with FSN 'Hematoma (disorder)'. The former is intended to be used for what '***a pathologist sees at the tissue level***', while the latter '***represents the clinical diagnosis that a clinician makes when they decide that a person has a "hematoma"***' [4, p41].

IHTSDO. International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English). 2015. p. 757.

Semantic tags

Options

Search Mode: Full text search mode

Status: Active components only

Group by concept

Filter results by Language

english	618

Filter results by Semantic Tag

disorder	445
procedure	151
morphologic abnormality	12
situation	6
finding	2
specimen	2

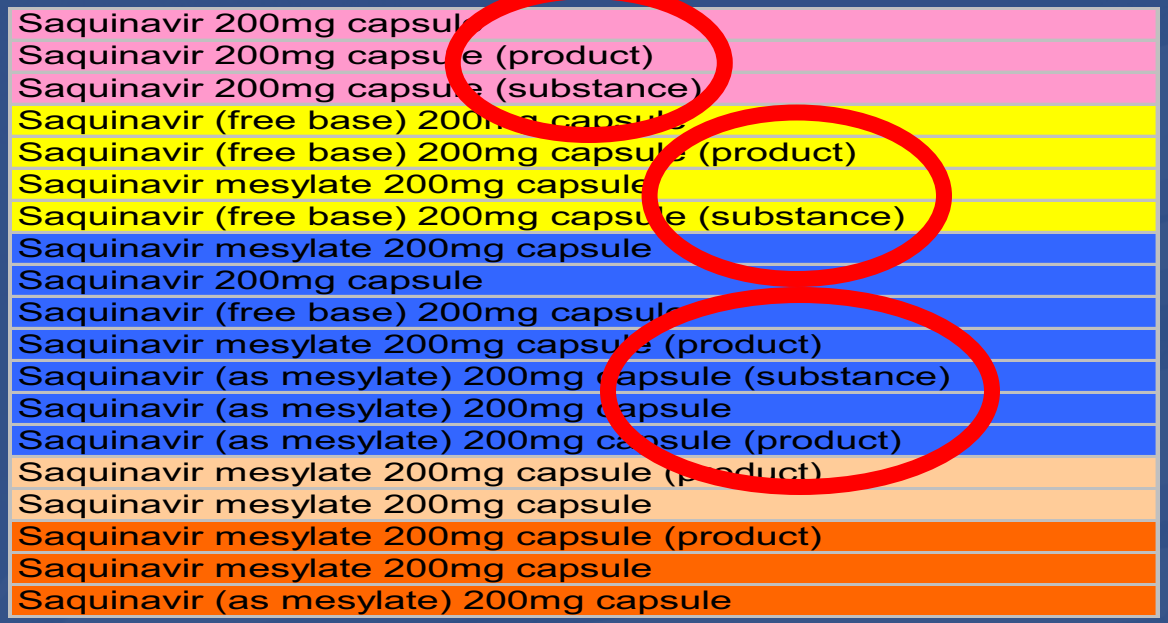
Type at least 3 characters ✔ Example: *blistered finger*

hematoma

618 matches found in 0.146 seconds.

Hematoma	Hematoma (morphologic abnormality)
Hematoma	Hematoma (disorder)
Deep hematoma	Deep hematoma (disorder)
Liver hematoma	Liver hematoma (disorder)
Cecal hematoma	Cecal hematoma (disorder)
Wound hematoma	Wound hematoma (finding)
Hematoma block	Fracture infiltration with local anesthetic (procedure)
Rectal hematoma	Rectal hematoma (disorder)
Facial hematoma	Hematoma of face (disorder)
Hematoma sample	Hematoma sample (specimen)
Vulval hematoma	Hematoma of vulva (disorder)
Pelvic hematoma	Pelvic hematoma (disorder)

Focus 2: Changes in semantic tags



Input:

- Fully Specified Names (FSN) of all versions

Methodology:

- Create chains of consecutive semantic changes for all concepts

For the 5 concepts above:

- (product) | (substance)
- (product) | (substance) | (product)

Examples of changes in semantic tag assignment

conceptID	Most recent FSN	Changes in semantic tags
66076007	Chewable tablet (qualifier value)	(substance)|(product)|(qualifier value)
66402002	Peritoneal dialysis education (procedure)	(procedure)|(regime/therapy)|(procedure)
68433009	Childhood (finding)	(function)|(observable entity)|(finding)
69736008	Vocational assessment (procedure)	(procedure)(regime/therapy)|(regime/therapy)|(procedure)
70409003	Mouthwash (qualifier value)	(substance)|(product)|(qualifier value)
70444001	Recessive gene (substance)	(function)|(observable entity)|(substance)
70790008	Absence of nausea and vomiting (situation)	(finding)|(context-dependent category)|(situation)
73669007	Kung fu (qualifier value)	(qualifier value)|(observable entity)|(qualifier value)
73905001	Sees flickering lights (finding)	(qualifier value)|(observable entity)|(finding)

Results for HARS and CIRS

- **Expected patterns do not always occur (1):**
  - C1 ‘possibly equivalent to’ C2/C3/… expected to go together with C1 being ambiguous
    →Not true for 4 concepts
  - 1,449 cases where a concept is stated to be duplicate without a corresponding same-as association,
  - 3,453 cases in which a same-as association was created without a duplicate assertion.

Results for HARS and CIRS

- **Expected patterns do not always occur (2):**
  - cases in which concepts are stated to be duplicates, yet denote clearly distinct entities.
  - '34759008: Urethral catheter, device (physical object)' is stated to be duplicated by 73 other concepts,
  - each duplicate denotes nevertheless a more precisely specified type of catheter, for example:
    - '349499005: Bard 10mL balloon 22Ch 1658 2-way all-silicone male length urethral Foley catheter', and
    - '349501002: Bard 10mL balloon 24Ch 1265LV 2-way Teflon coated male urethral Foley catheter'.
  - Several of these catheters are by means of other concepts, listed as descendants of '34759008'.

Was the brand name an issue rather than duplication?

Options

Search Mode: Partial matching search mode

Status: Active and inactive components

Group by concept

SAME AS associatio... ×

Filter results by Language

english 78

Filter results by Semantic Tag

78

Filter results by Module

SNOMED CT core module (core metadata concept) 78

Type at least 3 characters ✔ Example: *shou fra*

bard

78 matches found in 0.02 seconds.

Bard 10mL balloon 18Ch 1265LV 2-way Teflon coated male urethral Foley catheter	Bard 10mL balloon 18Ch 1265LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 22Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 22Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 18Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 18Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 16Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 16Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 16Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 16Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 26Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 26Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 26Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 26Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 24Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 24Ch 1266LV 2-way Teflon coated male urethral Foley catheter
Bard 30mL balloon 24Ch 1266LV 2-way Teflon coated male urethral Foley catheter	Bard 30mL balloon 24Ch 1266LV 2-way Teflon coated male urethral Foley catheter

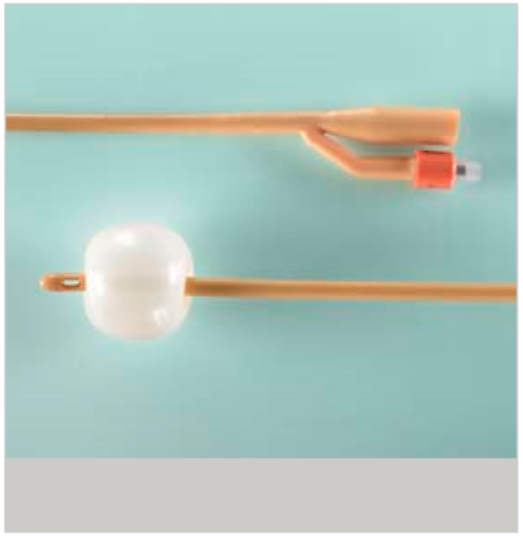
BARD | MEDICAL

ESSENTIAL. EFFECTIVE. EFFICIENT.

HOME ABOUT US PRODUCTS

Home › Products › Urological Drainage › Foley Cath

FOLEY CATHETERS



Results for HARS and CIRS

- Reasons for inactivations not always clear:
  - E.g.:
    - '391651001: Gluten-free/wheat-free baguette (product)'
    - '407775004 : Gluten-free/wheat-free baguette (product)'

ConceptID	Attribute	Value	History profile (one character per version)
391651001	AMB		AAAAYYYYYYYYYYYYYYYYYYYYY
	Is active		AAYYNNNNNNNNNNNNNNNNNNNNN
	Poss-equivalent-to	407775004	AAAAYYYYYYYYYYYYYYYYYYYYY
407775004	Semantic tag	product	AAAAYYYYYYYYYYYYYYYYYYYYY
	Is active		AAAAYYYYYYYYYYYYYYYYYYYYY

Some results for semantic tag changes

- There are 285 distinct patterns according to which SNOMED CT concepts underwent changes in the semantic tags assigned to them (including, 43 patterns, where there was a change from *no* tag to a tag).

Some results for semantic tag changes

- There are 285 distinct patterns according to which SNOMED CT concepts underwent changes in the semantic tags assigned to them (including, 43 patterns, where there was a change from *no* tag to a tag).
- A large amount of semantic tags were assigned to the FSN of concepts that were already inactive since many earlier versions. Why go through this trouble?

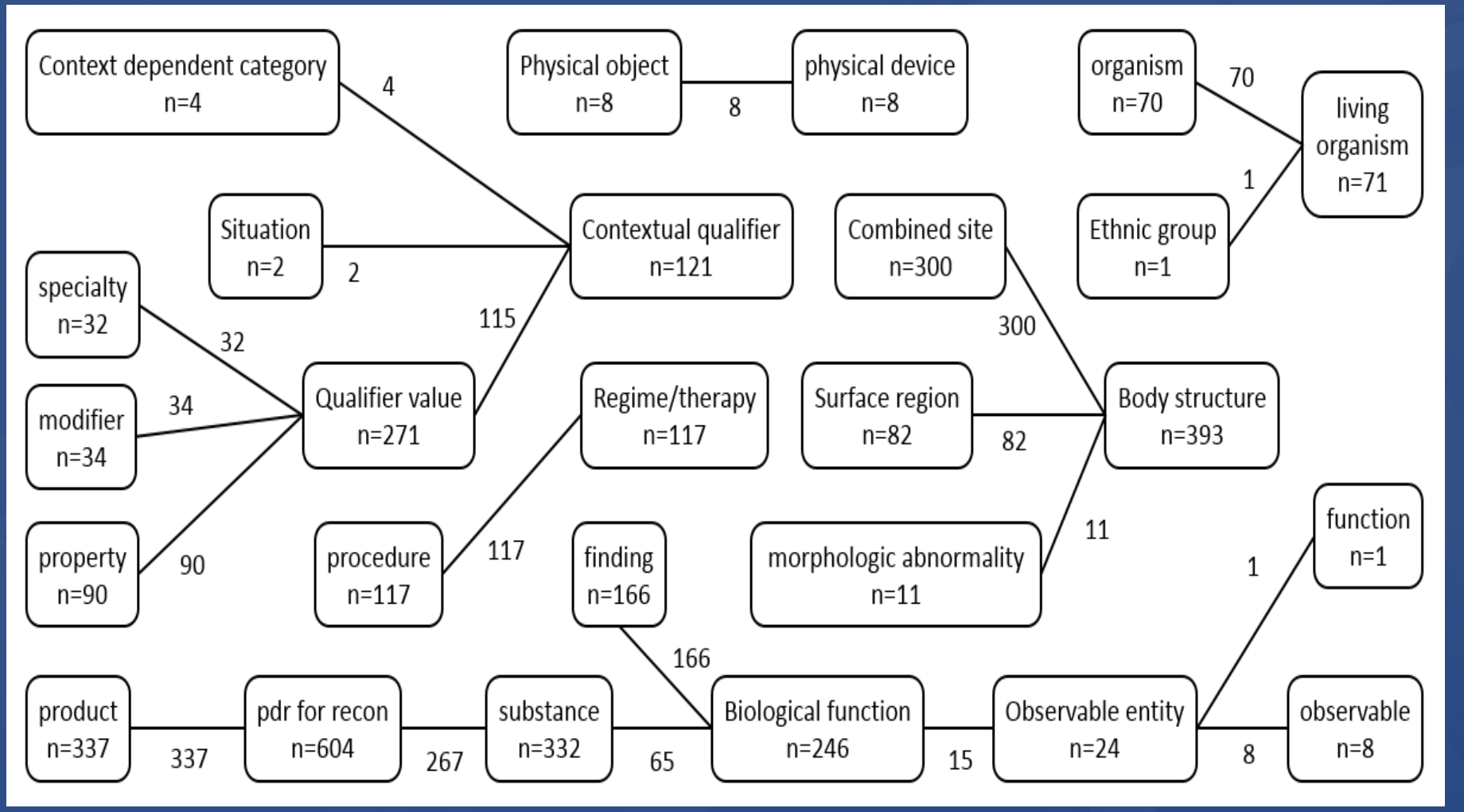
Some results for semantic tag changes

- There are 285 distinct patterns according to which SNOMED CT concepts underwent changes in the semantic tags assigned to them (including, 43 patterns, where there was a change from *no* tag to a tag).
- A large amount of semantic tags were assigned to the FSN of concepts that were already inactive since many earlier versions. Why go through this trouble?
- Certain change patterns occur frequently within a smaller subset of semantic tags, e.g. *disorder*, *finding*, *situation*, *morphologic abnormality*, *event* and *navigational concept*.
    - → Strong indication for issues with the ontology underlying SNOMED CT

Limitations

- Not clear from the distribution files what counts as semantic tags.
- Not everything at the end of an FSN – primarily of older FSNs – qualifies as FSN
- Some FSNs have more than one semantic tag (we assume).

Collocation of (assumed) semantic tags



Conclusions

- Release Format 2 (RF2) presents itself as a formidable resource to obtain a deeper insight in how SNOMED CT evolved.

Conclusions

- Release Format 2 (RF2) presents itself as a formidable resource to obtain a deeper insight in how SNOMED CT evolved.
- Yet, there are many pitfalls in attempting to derive from SNOMED CT's history mechanism what editorial and technical principles are followed, or whether they are applied consistently.

Conclusions

- Release Format 2 (RF2) presents itself as a formidable resource to obtain a deeper insight in how SNOMED CT evolved.
- Yet, there are many pitfalls in attempting to derive from SNOMED CT's history mechanism what editorial and technical principles are followed, or whether they are applied consistently.
- Whether it is the methodology proposed here itself, or a lack of, for instance, discriminatory power in the reasons for inactivation – one could even wonder why no reasons are given for the addition of new concepts –, is something that needs further to be researched.

Recommendations towards the IHTSDO

1. Formalize the relationships between semantic tags and SNOMED CT concept hierarchies,

Recommendations towards the IHTSDO

1. Formalize the relationships between semantic tags and SNOMED CT concept hierarchies,
2. Implement in the authoring environment mechanisms to prevent and detect incoherent and missing CIRS and HARS records, and

Recommendations towards the IHTSDO

1. Formalize the relationships between semantic tags and SNOMED CT concept hierarchies,
2. Implement in the authoring environment mechanisms to prevent and detect incoherent and missing CIRS and HARS records, and
3. Provide reasons for not only inactivations, but also activations, which reflect whether changes are
    1. purely internal in SNOMED CT (e.g. because of changes in the concept model) or
    2. external (changes in the covered domains).