Quality Assurance in LOINC® using Description Logic

Tomasz Adamusiak MD PhD

Postdoc at NIH NLM LHC CgSB
10/11 – 03/12

Objective

Asses whether areas for improvement can be identified in LOINC by changing its representation to OWL DL and comparing its classification to that of SNOMED CT

Why do it the hard way?

- Rector, A. L., & Brandt, S. **Why do it the hard way?** (2008) The case for an expressive description logic for SNOMED.

- More flexibility in a more expressive language
- A uniform, clear, and understandable schema
- Modularisation
- Access to standard tooling developed by the wider Semantic Web and OWL communities
  - Protégé, OWL API

Description Logic immediate benefits for LOINC

- Identify duplicates (codes, parts)
  - *45424-9 Epilepsy ≡ 45662-4 Seizure disorder*
  - *LP7216-7:Extremities ≡ LP7395-9:Limbs*
- Infer a hierarchy
  - Glucose | Urine → Carbohydrates | Urine
- Find inconsistencies
  - *44084-2 Fatty acids in Serum or Plasma → 7-hydroxyoctanoate | Urine*

BACKGROUND

Web Ontology Language (OWL)

OWL Manchester Syntax

has_component some Glucose

A number of papers explored LOINC SNOMED CT integration and DL

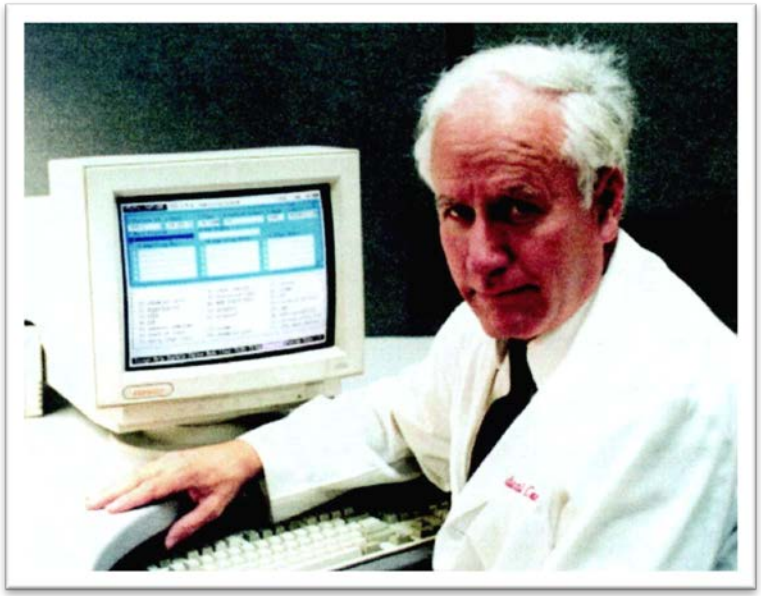
- Dolin, R. H., Huff, S. M., Rocha, R. A., Spackman, K. A., & Campbell, K. E. (**1998**). Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies.
- Spackman, K. A. (**1998**). Integrating sources for a clinical reference terminology: experience linking SNOMED to LOINC and drug vocabularies.
- Srinivasan A. et al. (2006). Semantic web representation of LOINC: an ontological perspective.
- Bodenreider, O. (2008). Issues in mapping LOINC laboratory tests to SNOMED CT.

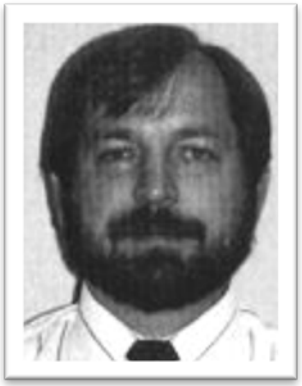
Quality Assurance in literature

- Geller et al. (2009). **Special issue on auditing of terminologies**. *Journal of biomedical informatics*

- Bodenreider, O., & Peters, L. B. (2009). A graph-based approach to **auditing RxNorm**.
- Wei, D., & Bodenreider, O. (2010). Using the abstraction network in complement to description logics for **quality assurance** in biomedical terminologies - a case study in **SNOMED CT**.
- Rector, A., & Iannone, L. (2011). Lexically suggest, logically define: **Quality assurance** of the use of qualifiers and expected results of post-coordination in **SNOMED CT**.
- Lin, M. C., Vreeman, D. J., McDonald, C. J., & Huff, S. M. (2012). **Auditing** consistency and usefulness of **LOINC use** among three large institutions - Using version spaces for grouping LOINC codes.

A universal code system for identifying laboratory and clinical observations





LOINC codes consist of parts

Code:

2160-0	**Creatinine** [Mass/volume] in **Serum or Plasma**

Parts:

Part Type	Part No.	Part Name
Component	LP14355-9	**Creatinine**
Property	LP6827-2	MCnc [*Mass Concentration*]
Time	LP6960-1	Pt [*Point in time (spot)*]
System	LP7576-4	**Ser/Plas** [*Serum or Plasma*]
Scale	LP7753-9	Qn

METHODS

We used part links to create logical definitions for codes

Code:

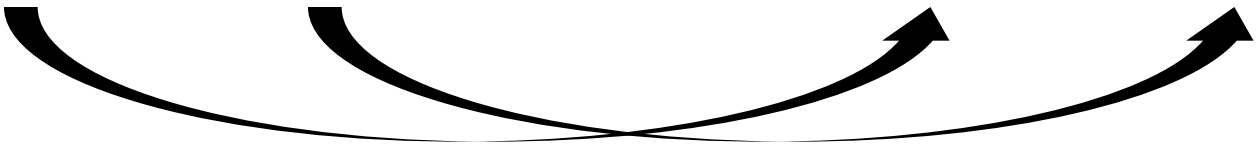
2160-0	Creatinine [Mass/volume] in Serum or Plasma

Parts:

Part Type	Part Name
Component	**Creatinine**
Property	**MCnc**
Time	**Pt**
System	**Ser/Plas**
Scale	**Qn**

DL definition:

(*has_component* some **Creatinine**) and
(*has_property* some **MCnc**) and
(*has_time_aspect* some **Pt**) and
(*has_system* some **Ser/Plas**) and
(*has_scale* some **Qn**)



Component 2nd subpart: challenge

Code:

1558-6	Fasting glucose [Mass/volume] in Serum or Plasma

Parts:

Part Type	Part No.	Part Name
Component	LP14635-4	Glucose
Challenge	**LP20355-1**	**post CFst**
Property	LP6827-2	MCnc *[Mass Concentration]*
Time	LP6960-1	Pt *[Point in time (spot)]*
System	LP7576-4	Ser/Plas *[Serum or Plasma]*
Scale	LP7753-9	Qn

Component 3rd subpart: adjustment

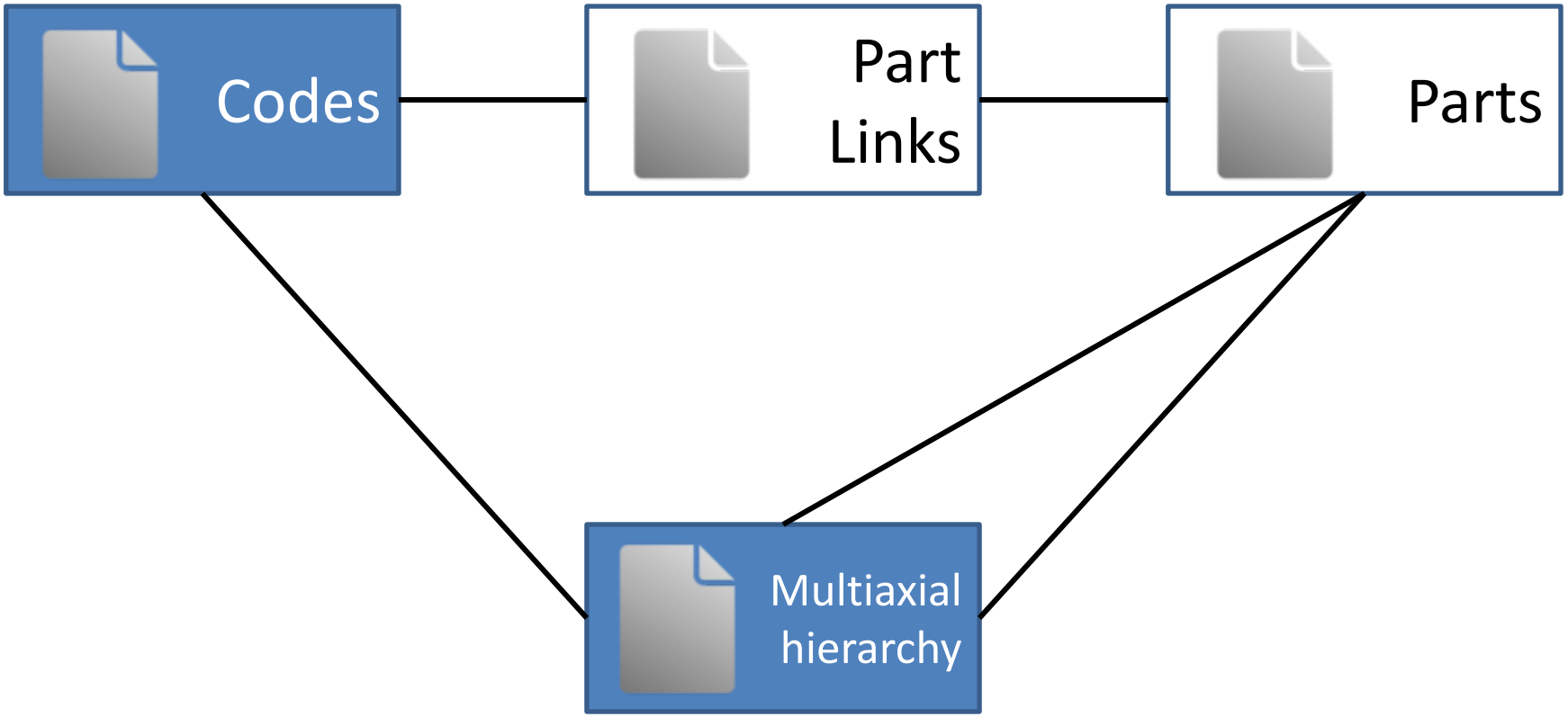
Code:

23811-3	Alpha-1-Fetoprotein [Multiple of the median] **adjusted** in Serum or Plasma

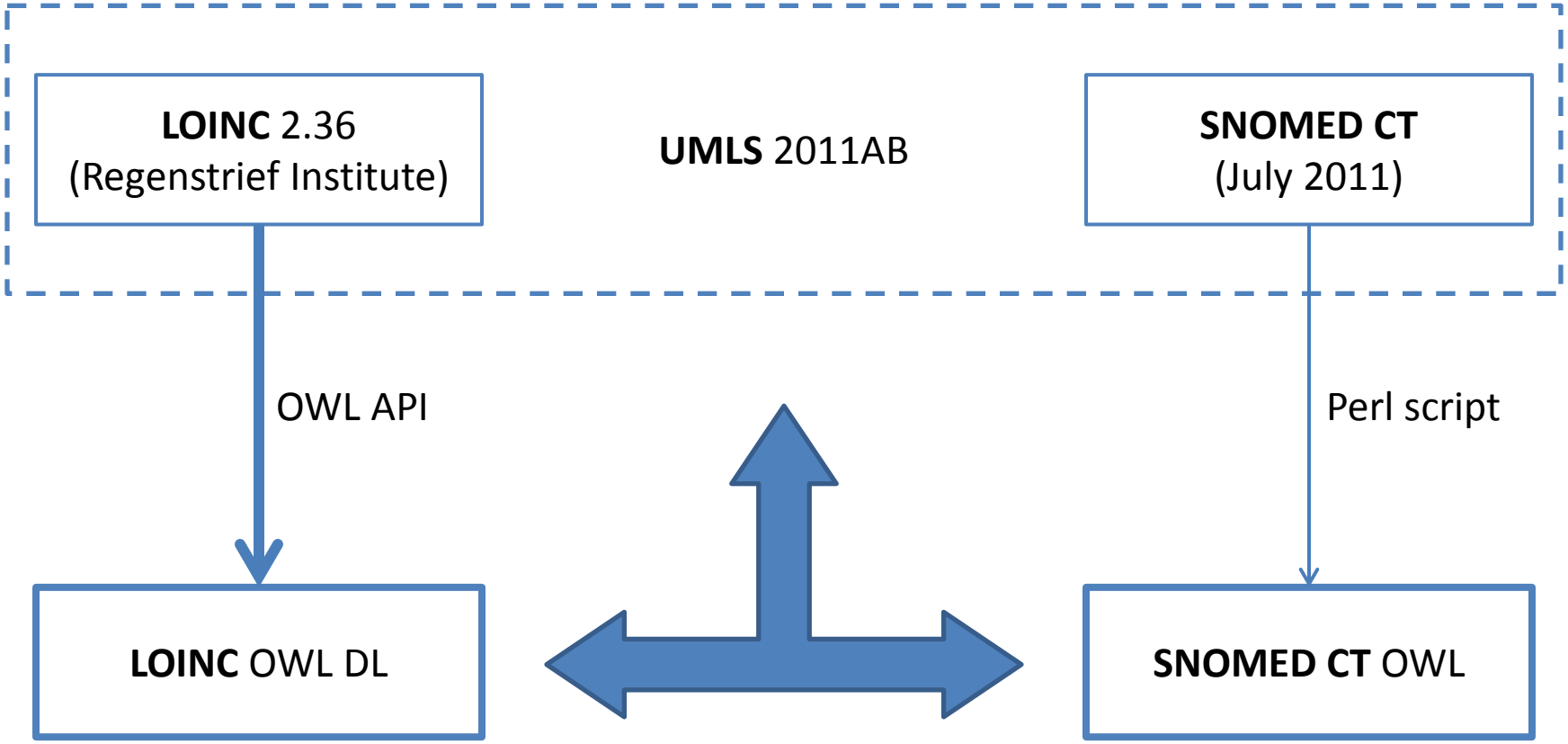
Parts:

Part Type	Part No.	Part Name
Component	LP14331-0	Alpha-1-Fetoprotein
Adjustment	**LP20174-6**	**adjusted**
Property	LP71590-1	MoM *[Multiple of the median]*
Time	LP6960-1	Pt *[Point in time (spot)]*
System	LP7576-4	Ser/Plas *[Serum or Plasma]*
Scale	LP7753-9	Qn

LOINC parts are not available in the public release (2.36)



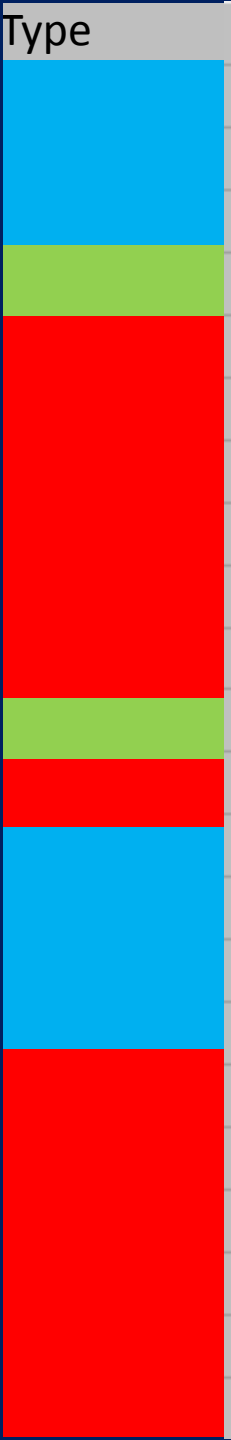
Materials


LOINC 2.36
(Regenstrief Institute)
UMLS 2011AB
SNOMED CT
(July 2011)
OWL API
Perl script
LOINC OWL DL
SNOMED CT OWL

Multiaxial hierarchy in LOINC could be vastly improved with DL

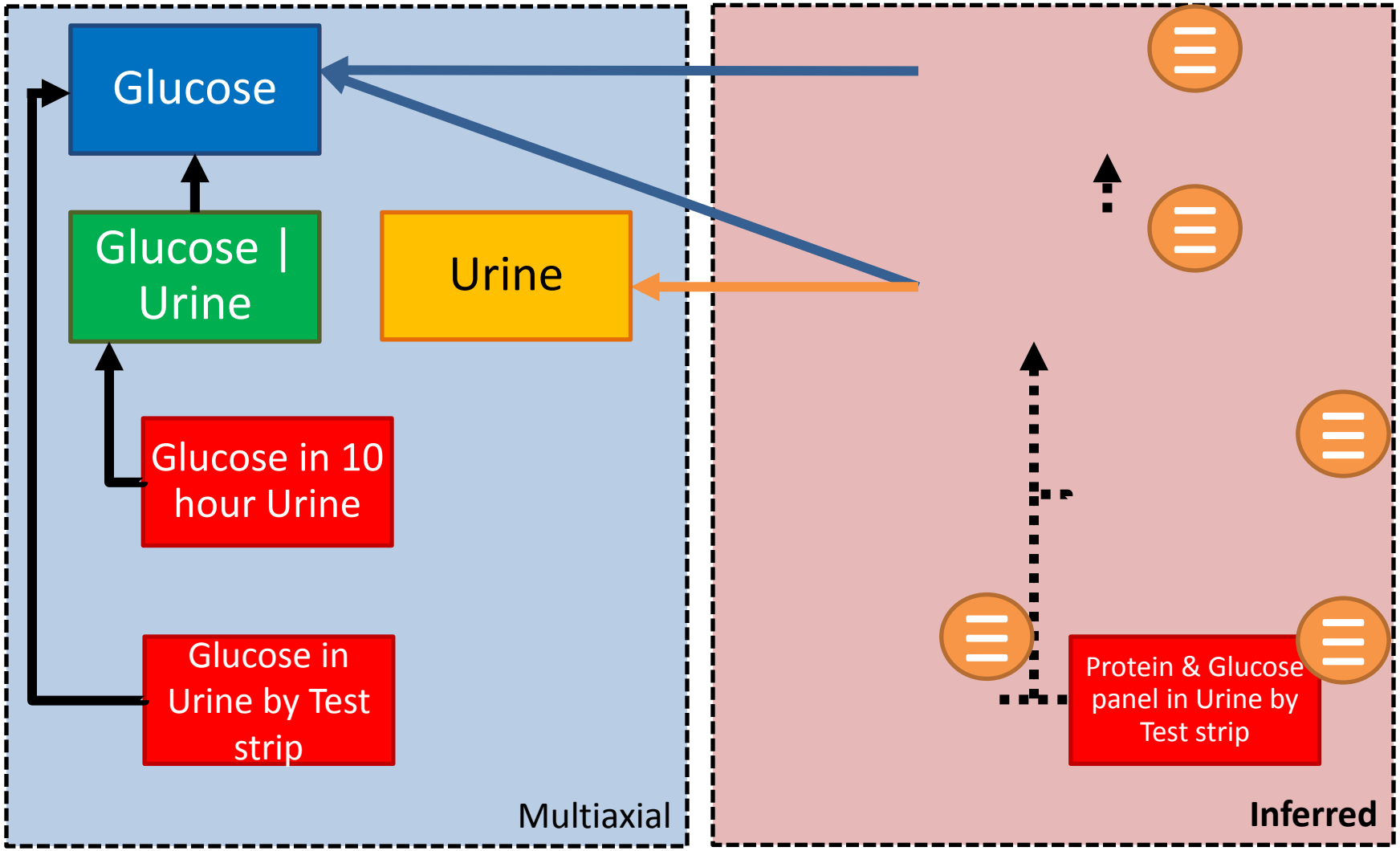
Type	Category or ShortName	Component	System	Code
	Chemistry			**LP31388-9**
	Sugars/Sugar metabolism			**LP31399-6**
	Glucose			**LP14635-4**
	Glucose | Urine			**LP43854-6**
	Glucose [Mass/time] in 10...	Glucose	Urine	21307-4
	Glucose [Mass/time] in 18...	Glucose	Urine	58997-8
	Glucose [Presence] in 24...	Glucose	Urine	32174-5
	Glucose [Mass/volume] in...	Glucose	Urine	21305-8
	Glucose [Mass/time] in 24...	Glucose	Urine	2351-5
	Glucose [Moles/volume] in...	Glucose	Urine	25916-8
	Glucose tetrasaccharide | Urine			**LP71460-7**
	Glucose...	Glucose tetrasaccharide/Creatinine	Urine	53868-6
	Chemistry, challenge			**LP40271-6**
	Urinalysis			**LP32744-2**
	Analytes			**LP40317-7**
	Glucose			**LP14635-4**
	Glucose [Presence] in Urine...	Glucose	Urine	25428-4
	Glucose [Presence] in Urine...	Glucose	Urine	50555-2
	Glucose [Mass/volume] in...	Glucose	Urine	5792-7
	Glucose [Mass/volume] in...	Glucose	Urine	53328-1
	Glucose [Moles/volume] in...	Glucose	Urine	22705-8
	Glucose [Moles/volume] in...	Glucose	Urine	59156-0

Screenshot from the Regenstrief LOINC Mapping Assistant (RELMA)

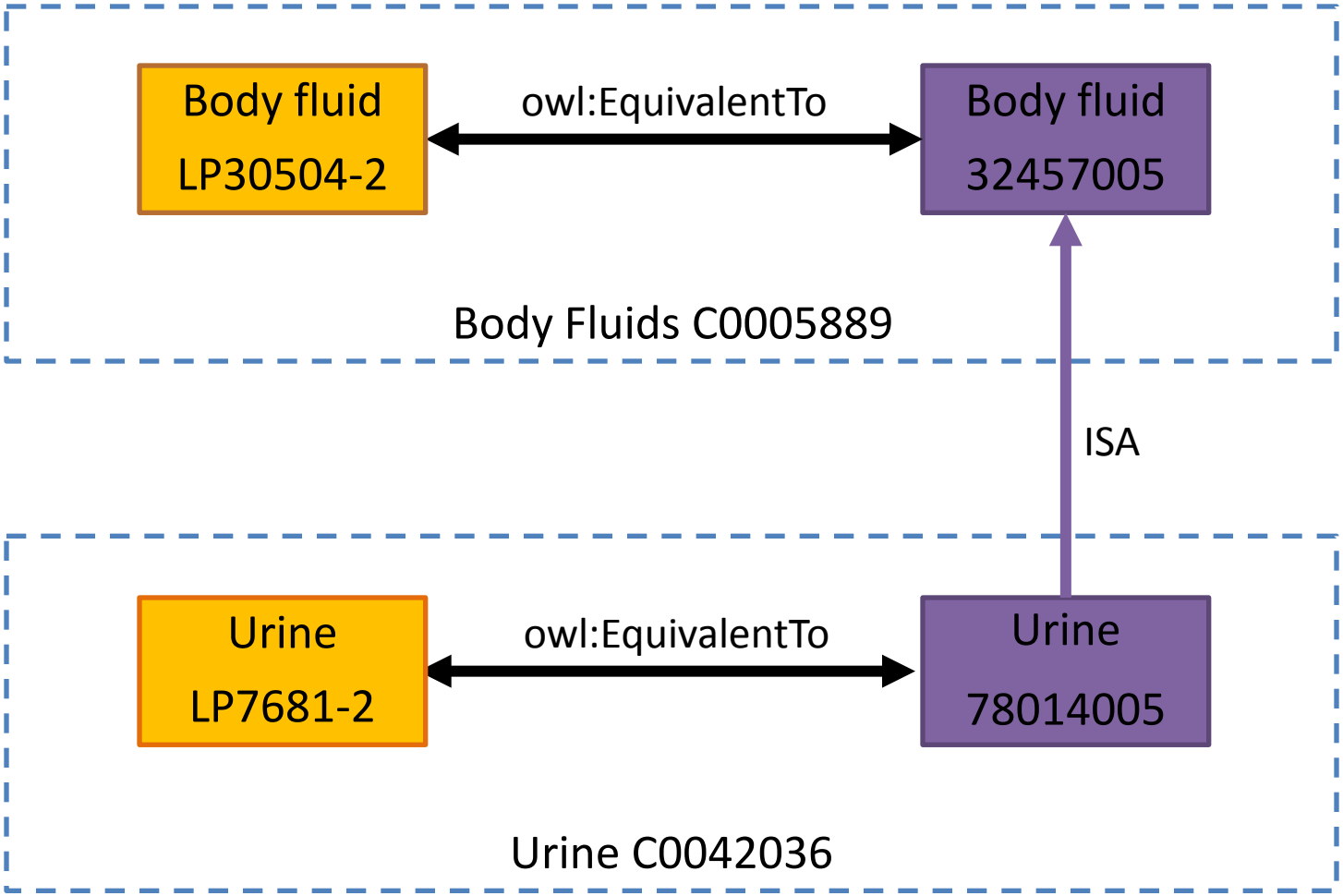


Type	Category or ShortName	Component	System
	Chemistry		
	Sugars/Sugar metabolism		
	Glucose		
	Glucose | Urine		
	Glucose [Mass/time] in 10...	Glucose	Urine
	Glucose [Mass/time] in 18...	Glucose	Urine
	Glucose [Presence] in 24...	Glucose	Urine
	Glucose [Mass/volume] in...	Glucose	Urine
	Glucose [Mass/time] in 24...	Glucose	Urine
	Glucose [Moles/volume] in...	Glucose	Urine
	Glucose tetrasaccharide | Urine		
	Glucose...	Glucose tetrasaccharide/Creatinine	Urine
	Chemistry, challenge		
	Urinalysis		
	Analytes		
	Glucose		
	Glucose [Presence] in Urine...	Glucose	Urine
	Glucose [Presence] in Urine...	Glucose	Urine
	Glucose [Mass/volume] in...	Glucose	Urine
	Glucose [Mass/volume] in...	Glucose	Urine
	Glucose [Moles/volume] in...	Glucose	Urine
	Glucose [Moles/volume] in...	Glucose	Urine

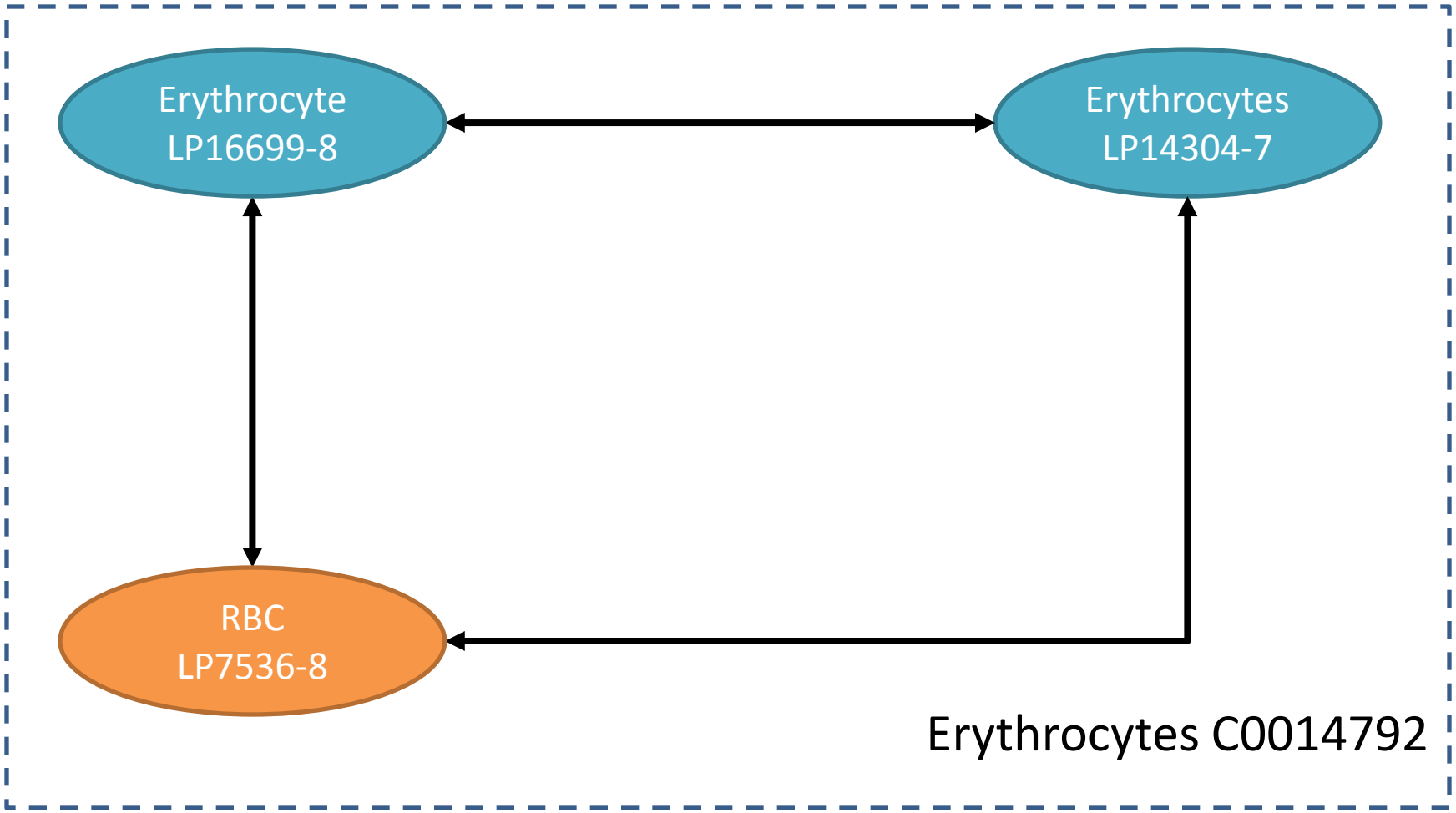
Separated codes and parts and defined corresponding observations



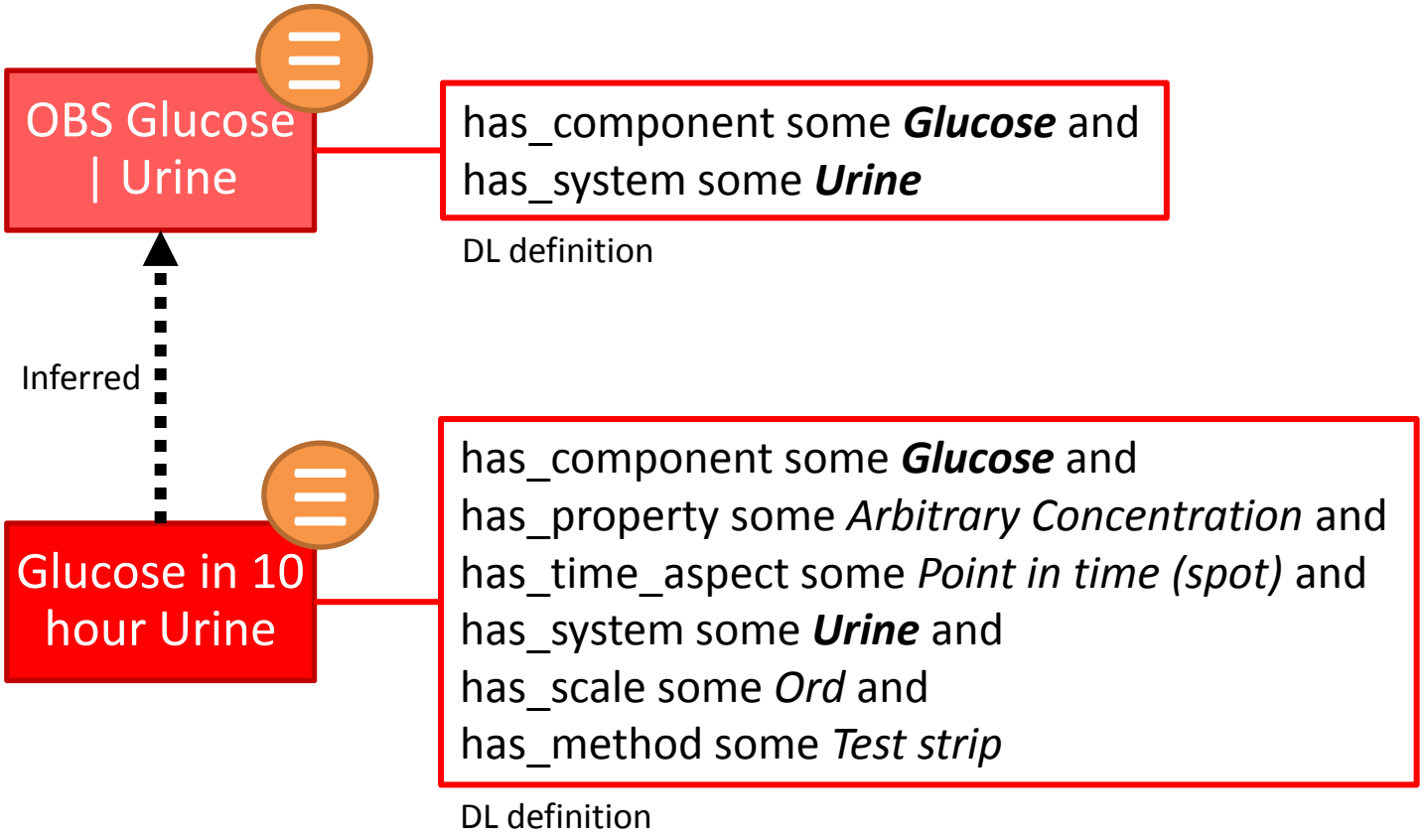
SNOMED CT compensates for missing parts relations in LOINC



We can identify semantically equivalent LOINC parts via UMLS



Reasoner infers logical consequences from a set of asserted facts or axioms



Huge Knowledge Base classified with ConDOR reasoner

	LOINC	LOINC+SNOMED CT
Number of classes	173,091	468,572
Number of asserted axioms:	677,023	1,577,861
Number of inferred axioms	126,020	413,050
LOINC codes	65,003	
LOINC parts	82,102	
	LOINC multiaxial hierarchy	
LOINC codes	47,405	
LOINC parts	25,982	

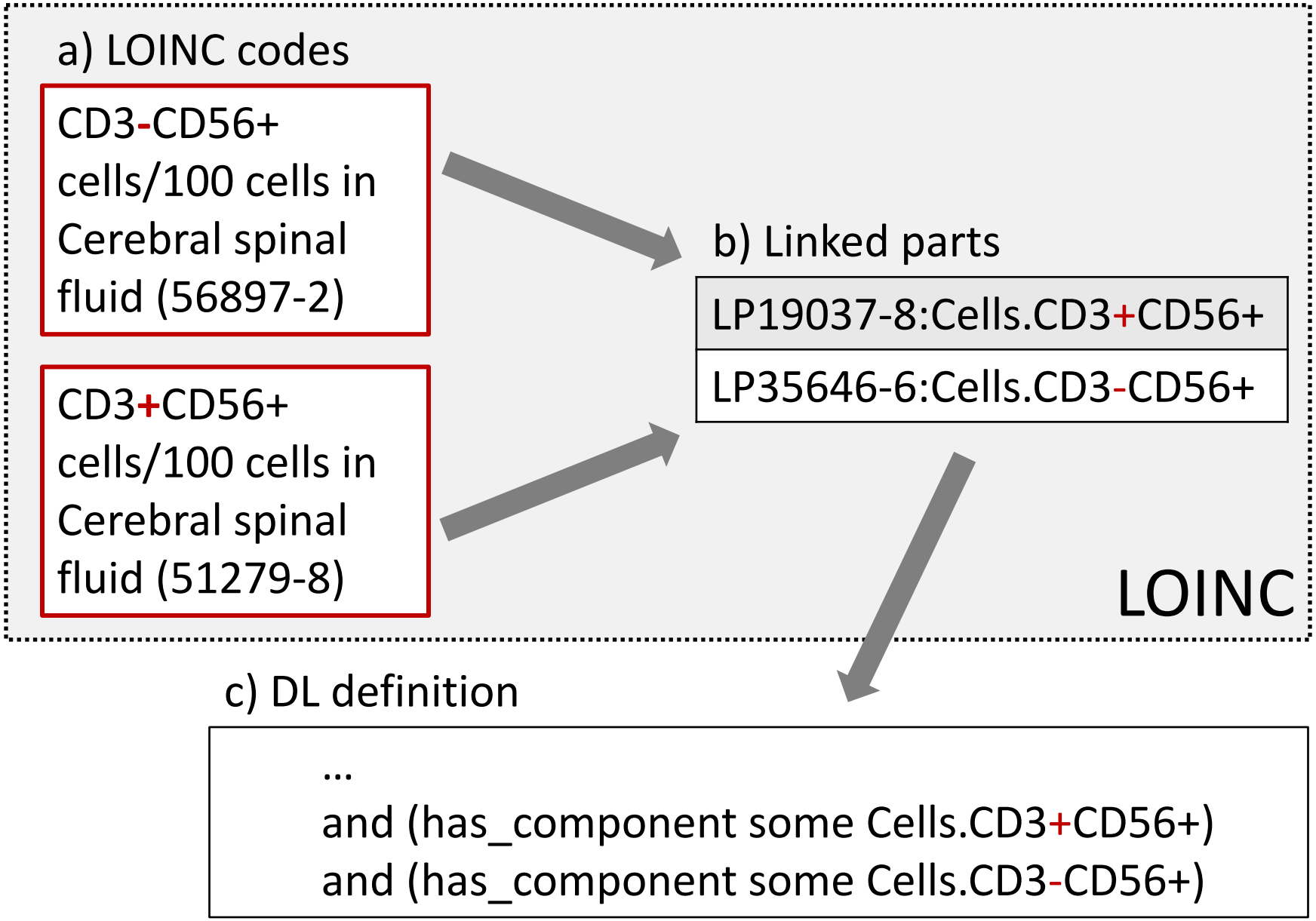
RESULTS

Without SNOMED CT: inferred 325 sets of equivalent LOINC codes

- 56897-2:Cells.CD3-CD56+/100 cells:NFr:Pt:CSF:Qn
- 51279-8:Cells.CD3+CD56+/100 cells:NFr:Pt:CSF:Qn

- 10132-9:T' wave amplitude.lead AVR:Elpot:Pt:Heart:Qn:EKG
- 10144-4:T wave amplitude.lead AVR:Elpot:Pt:Heart:Qn:EKG

- 36748-2:Views oblique:Find:Pt:Spine.cervical:Nar:XR
- 42164-4:Views & oblique:Find:Pt:Spine.cervical:Nar:XR


a) LOINC codes
CD3-CD56+ cells/100 cells in Cerebral spinal fluid (56897-2)
CD3+CD56+ cells/100 cells in Cerebral spinal fluid (51279-8)
b) Linked parts
LP19037-8:Cells.CD3+CD56+
LP35646-6:Cells.CD3-CD56+
LOINC
c) DL definition
…
and (has_component some Cells.CD3+CD56+)
and (has_component some Cells.CD3-CD56+)

LOINC must have realised the problem

56897-2 CD3-CD56+ cells/100 cells in Cerebral spinal fluid

NAME

Fully-Specified Name:	**Component**
	Cells.CD3-CD56+/100 cells

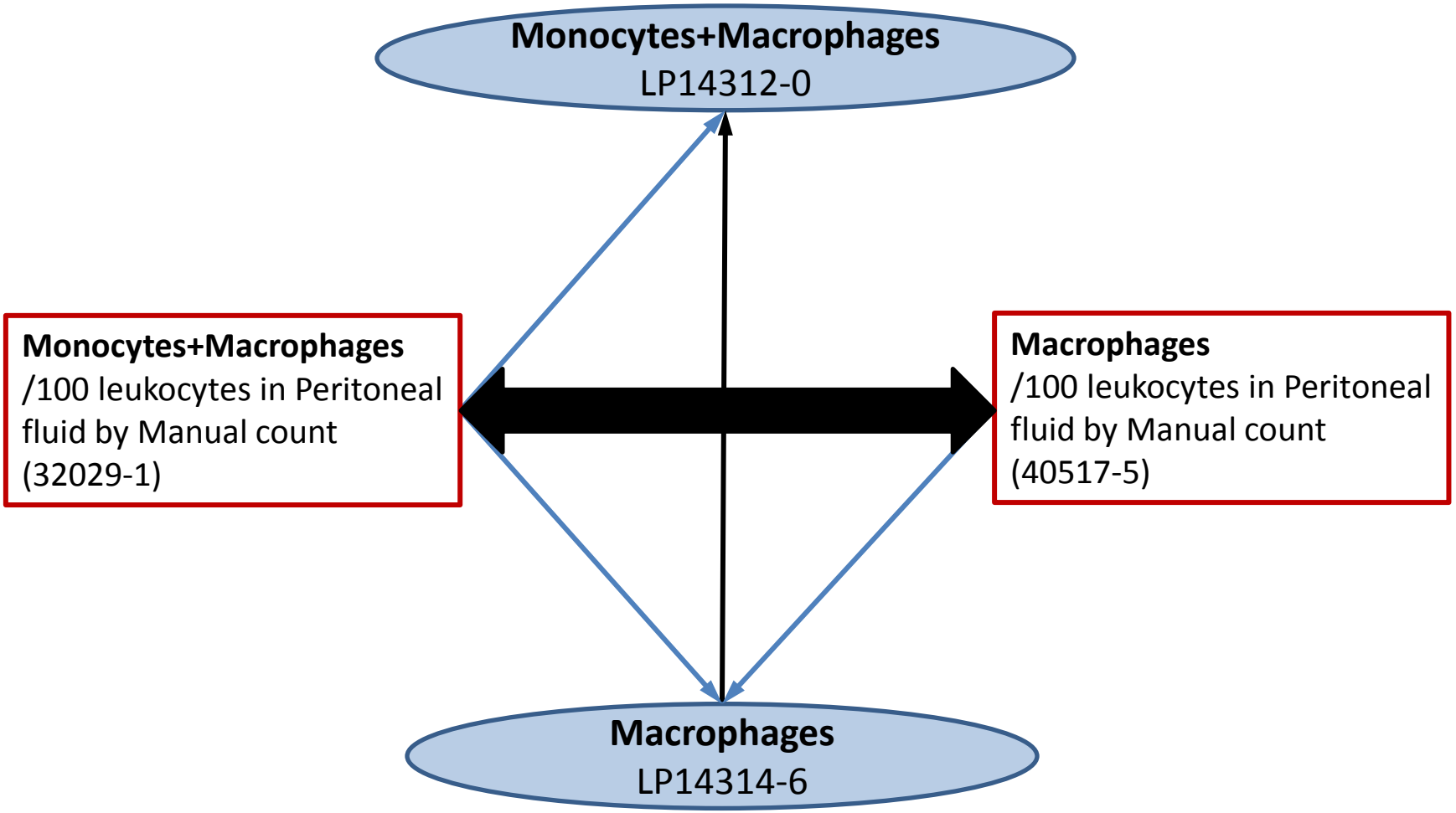
RELATED NAMES

CD3 Cells

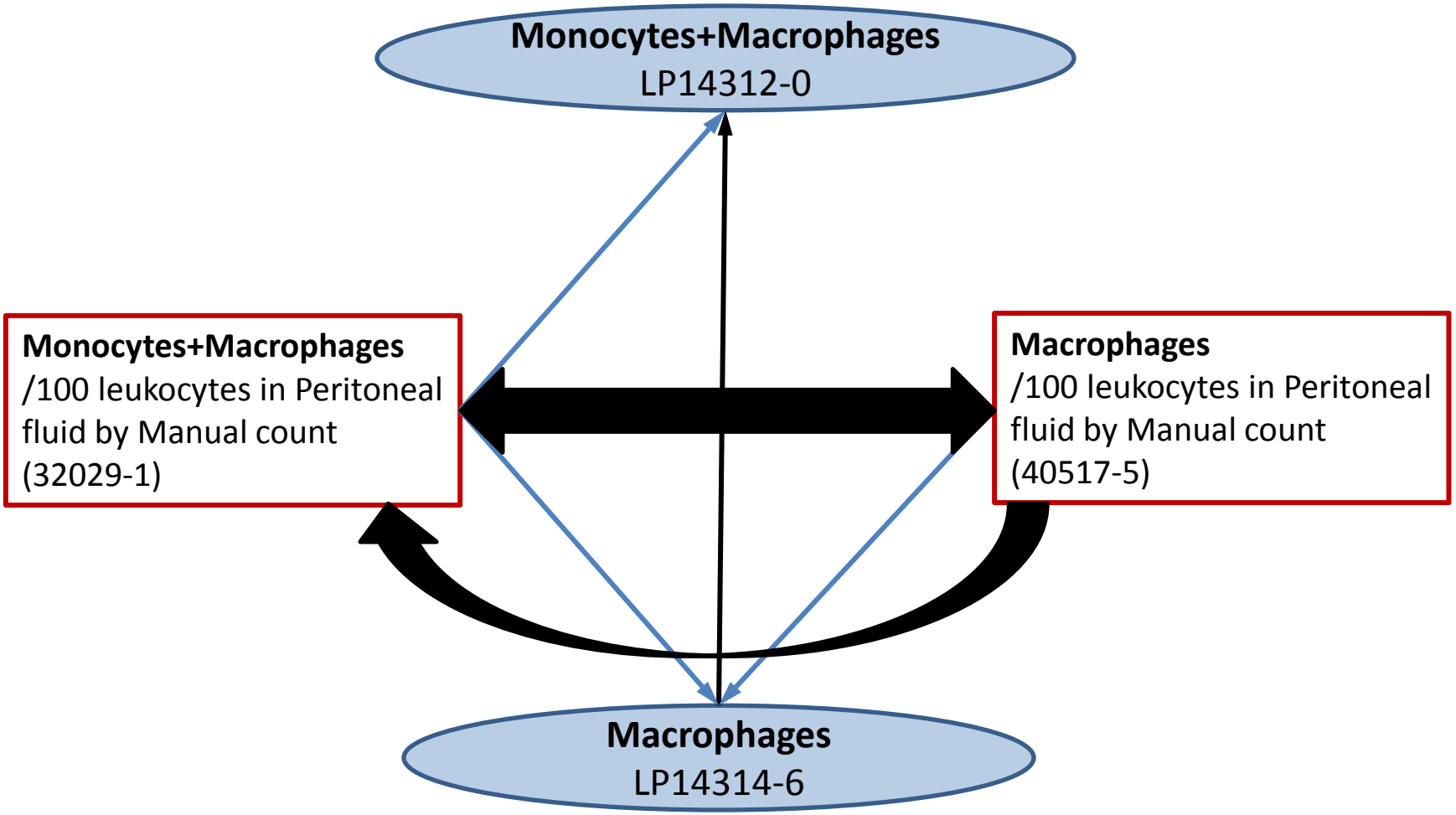
CD3+CD56+ Cells

CD3-CD56+ Cells

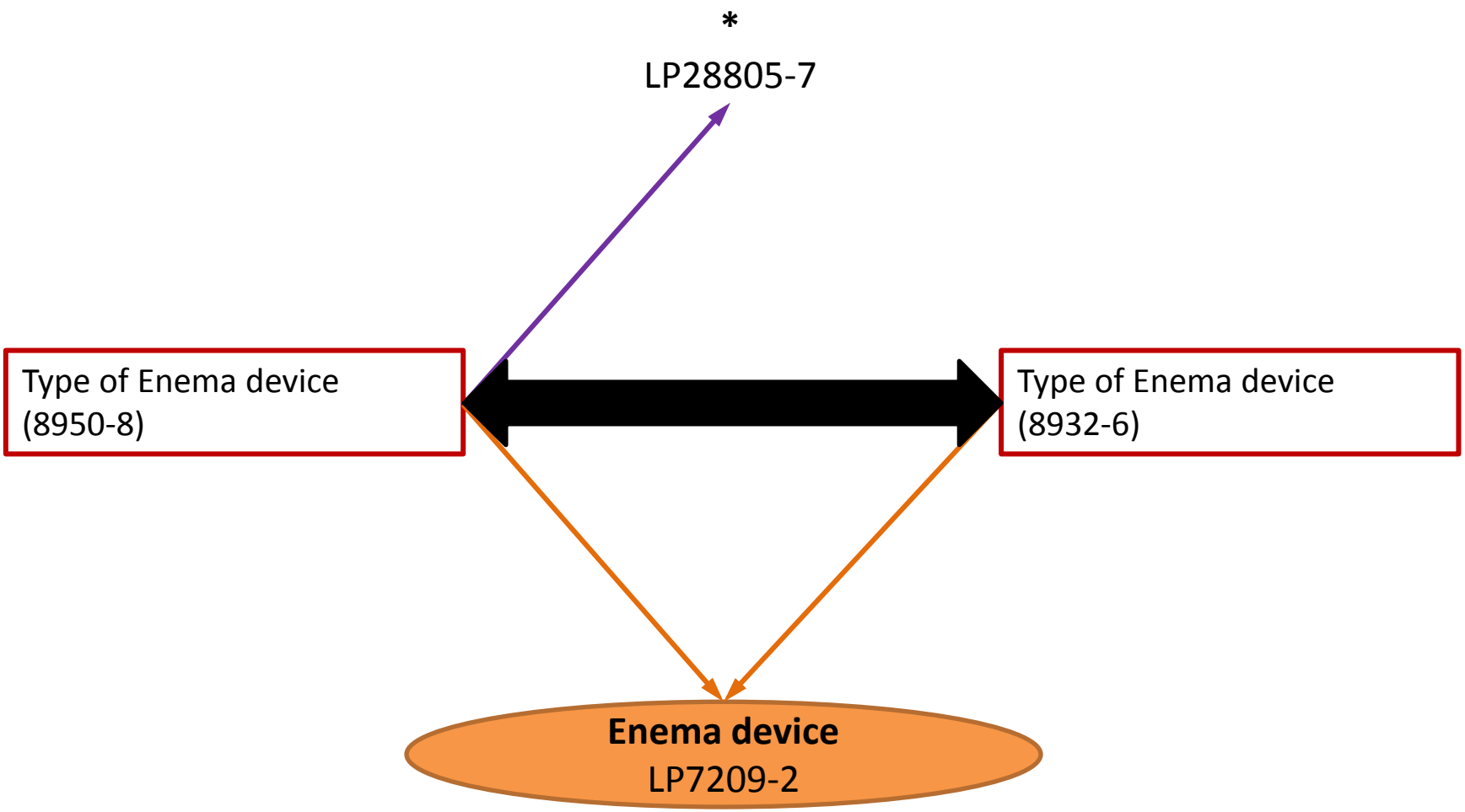
Inconsistencies in part hierarchy result in incorrect inference



Pop quiz: removing which has_component relation changes equivalence to subsumption?



Issues with referential integrity



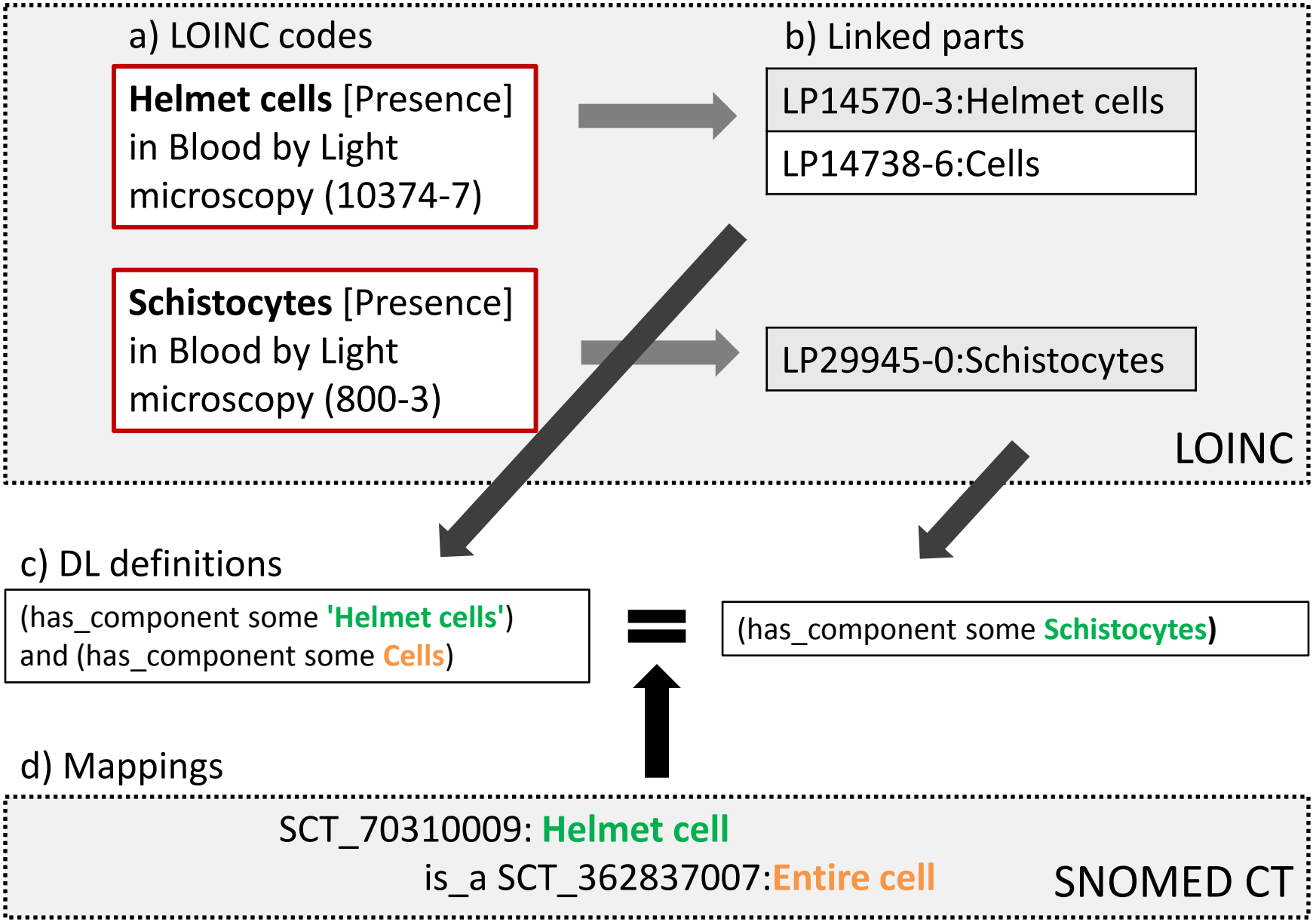
SNOMED CT enrichment gives 102 sets of equivalent LOINC codes

- 46062-6:**Treatment**s:-:Pt:^Patient:Set:
- 46064-2:**Therapies**:-:Pt:^Patient:Set:

- 45424-9:**Epilepsy**:Find:Pt:^Patient:Ord:MDS
- 45662-4:**Seizure disorder**:Find:Pt:^Patient:
  Ord:MDS

- 8703-1:Physical findings:Find:Pt:**Extremities**:Nom:Observed
- 32430-1:Physical findings:Find:Pt:**Extremity**:Nom:Observed

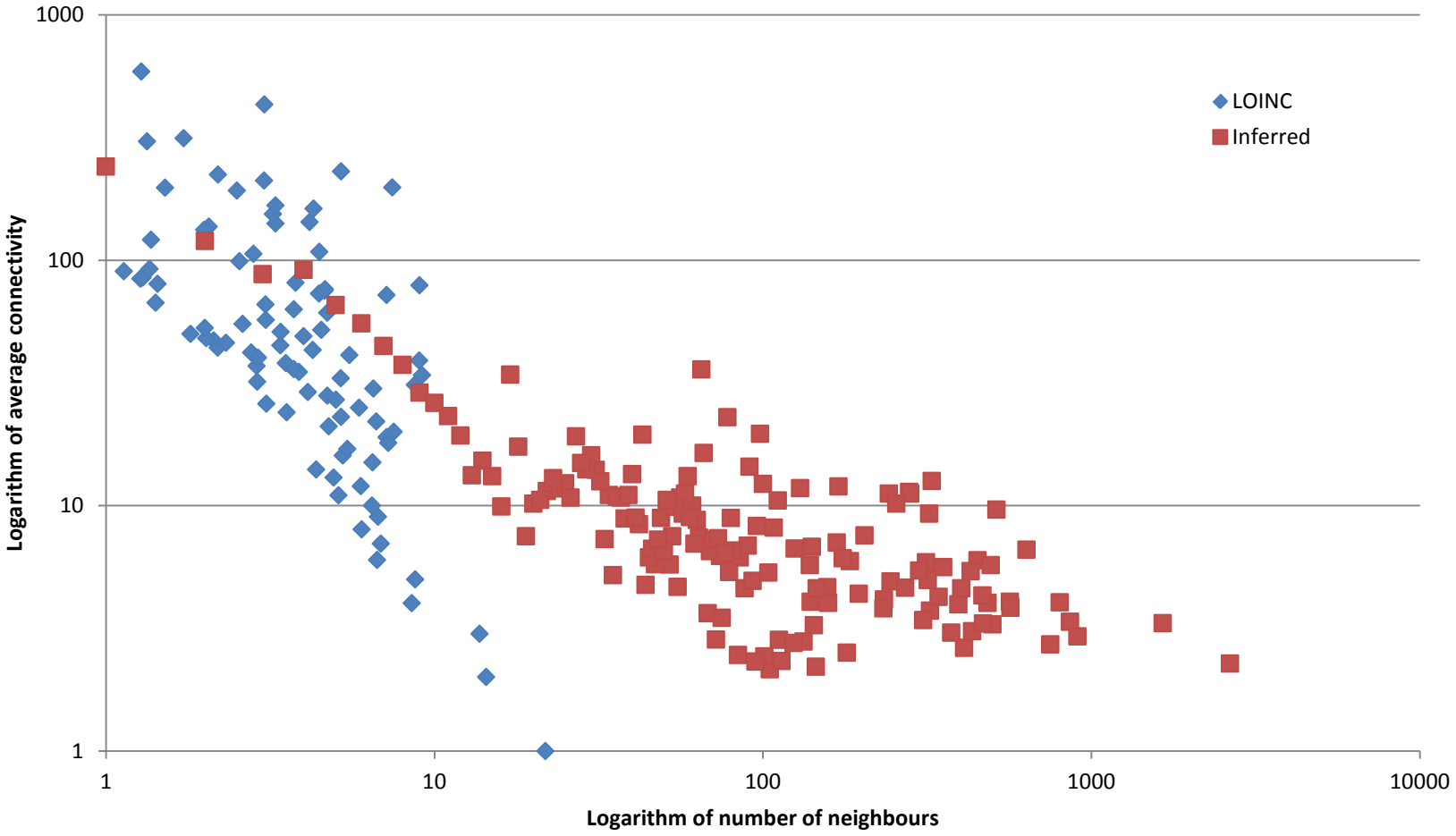
- 39037-7:Multisection^W contrast **IV:Find:Pt:Upper extremity**:Nar:MRI
- 36208-7:Multisection^W contrast IV:Find:Pt:**Upper arm**:Nar:MRI


a) LOINC codes
Helmet cells [Presence] in Blood by Light microscopy (10374-7)
Schistocytes [Presence] in Blood by Light microscopy (800-3)
b) Linked parts
LP14570-3:Helmet cells
LP14738-6:Cells
LP29945-0:Schistocytes
LOINC
c) DL definitions
(has_component some 'Helmet cells') and (has_component some Cells)
=
(has_component some Schistocytes)
d) Mappings
SCT_70310009: Helmet cell
is_a SCT_362837007:Entire cell
SNOMED CT

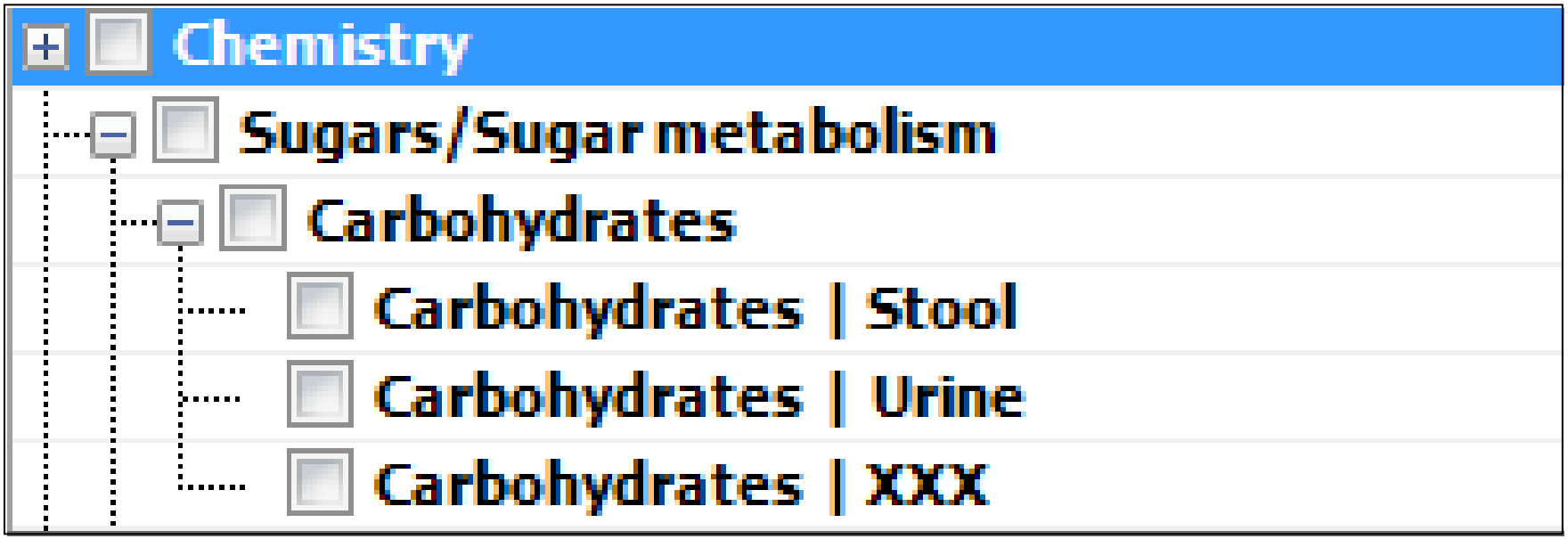
Inferred hierarchy has more connected nodes and is better connected

	LOINC	Inferred
Number of connected nodes	73,387	82,350
Network diameter	15	13
Connected components	8	513
Shortest paths	425,976	1,119,232
Characteristic path length	3.39	3.81
Average number of neighbours	2.01	3.40

Inferred nodes are better connected locally

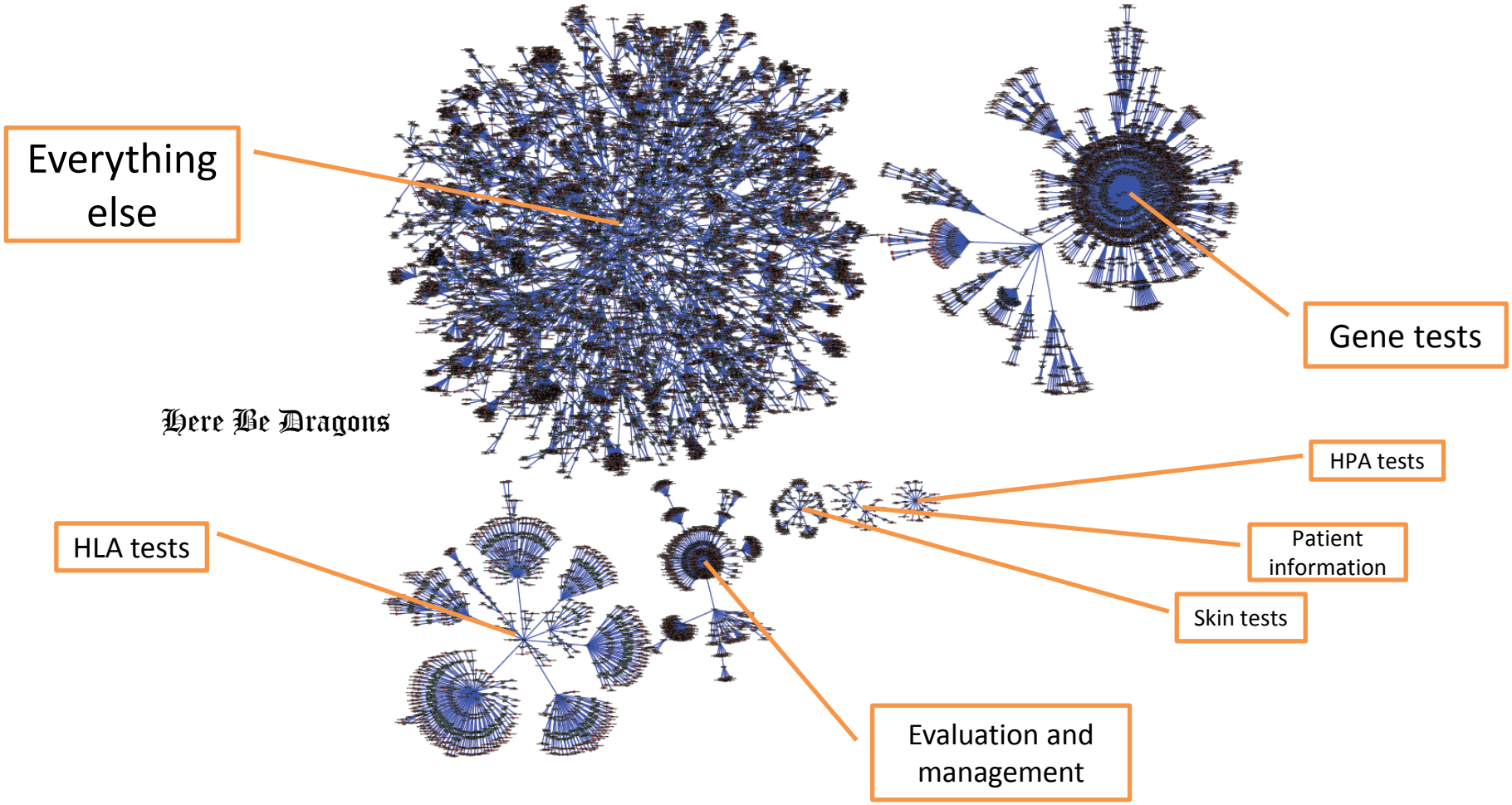


Find all carbohydrate observations

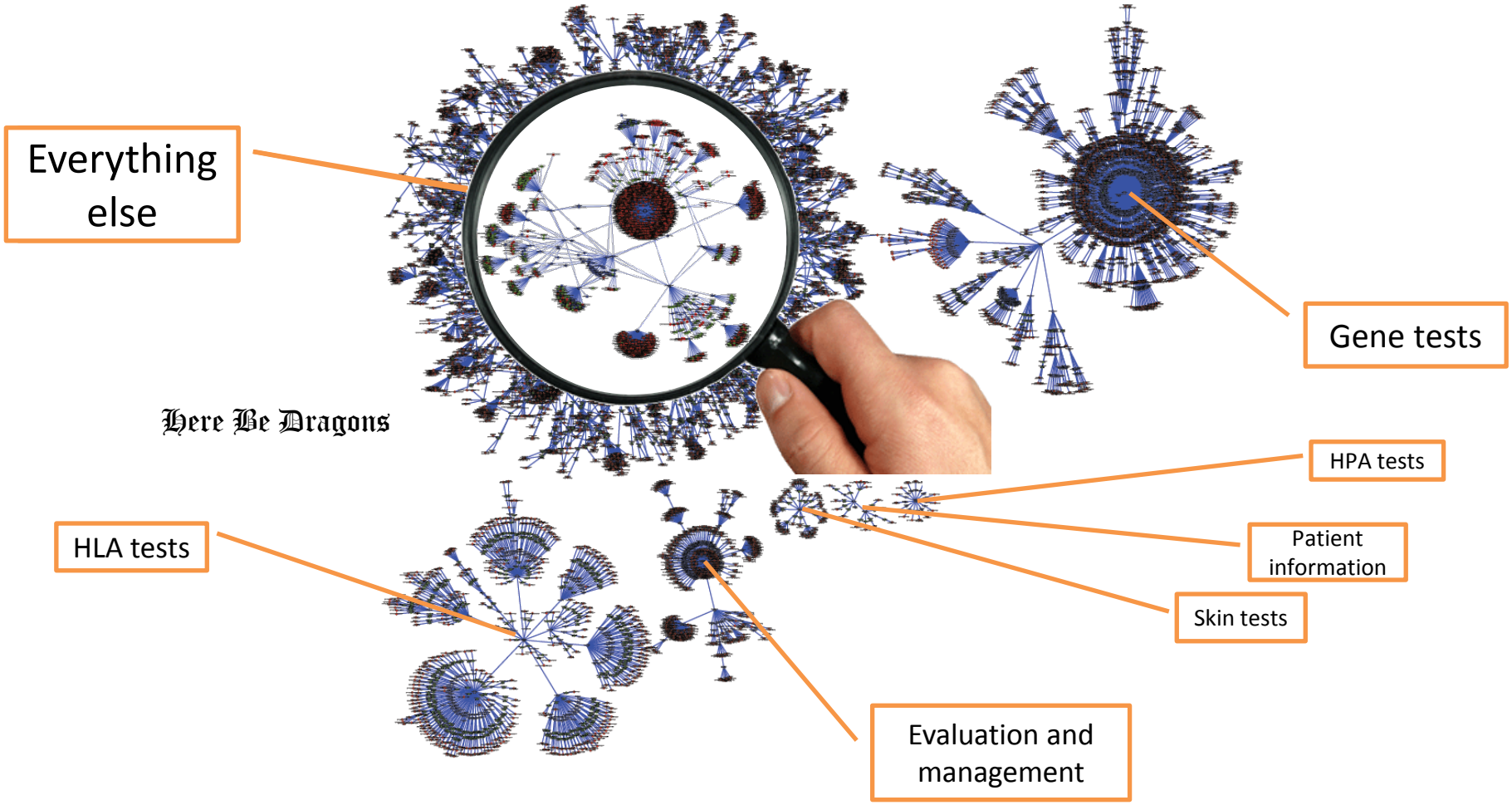


Regenstrief LOINC Mapping Assistant (RELMA)

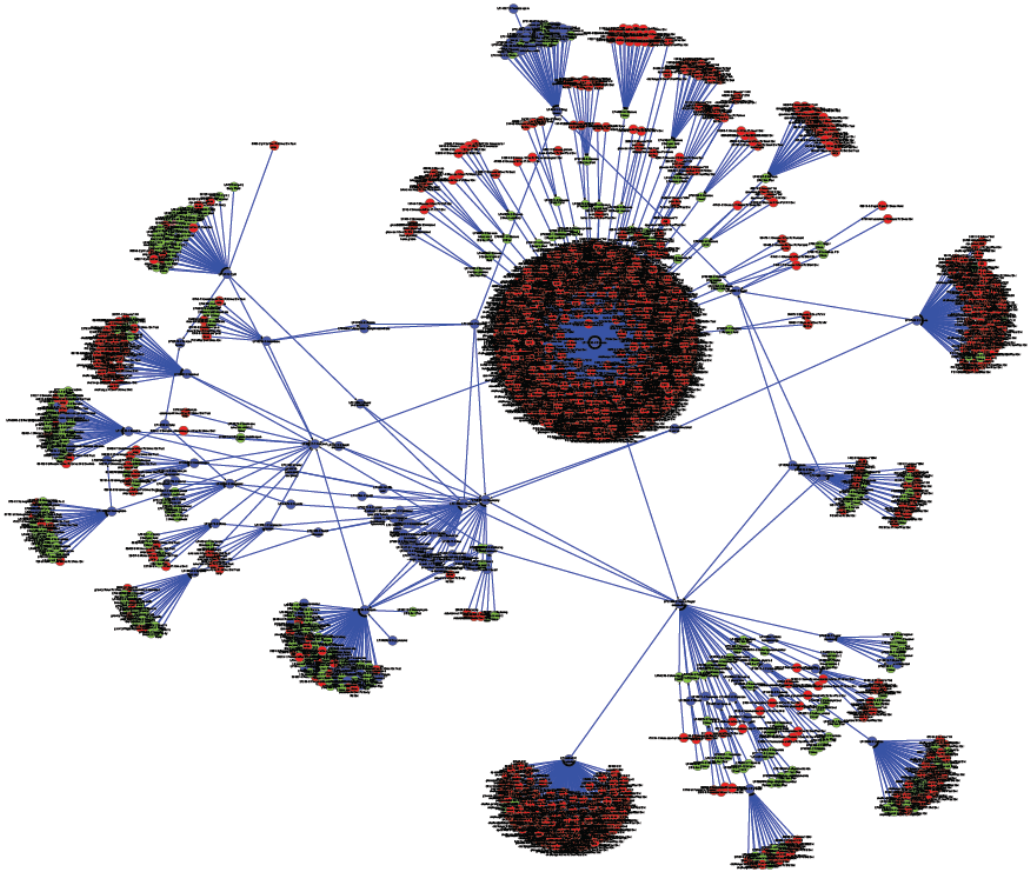
¡¿Find all carbohydrate observations?!



It is not easy



COMPONENT LP14635-4:Glucose is the most connected node

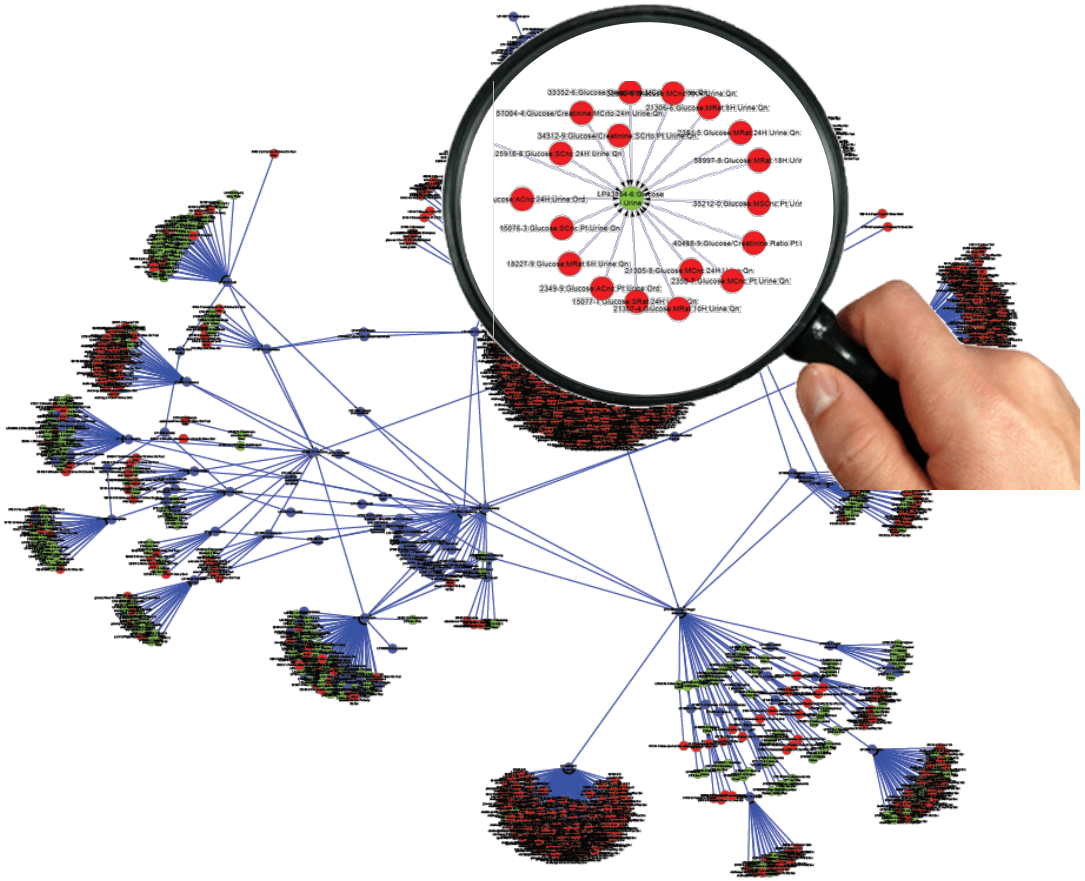


Legend:

Test

Mx

COMPONENT LP14635-4:Glucose is the most connected node

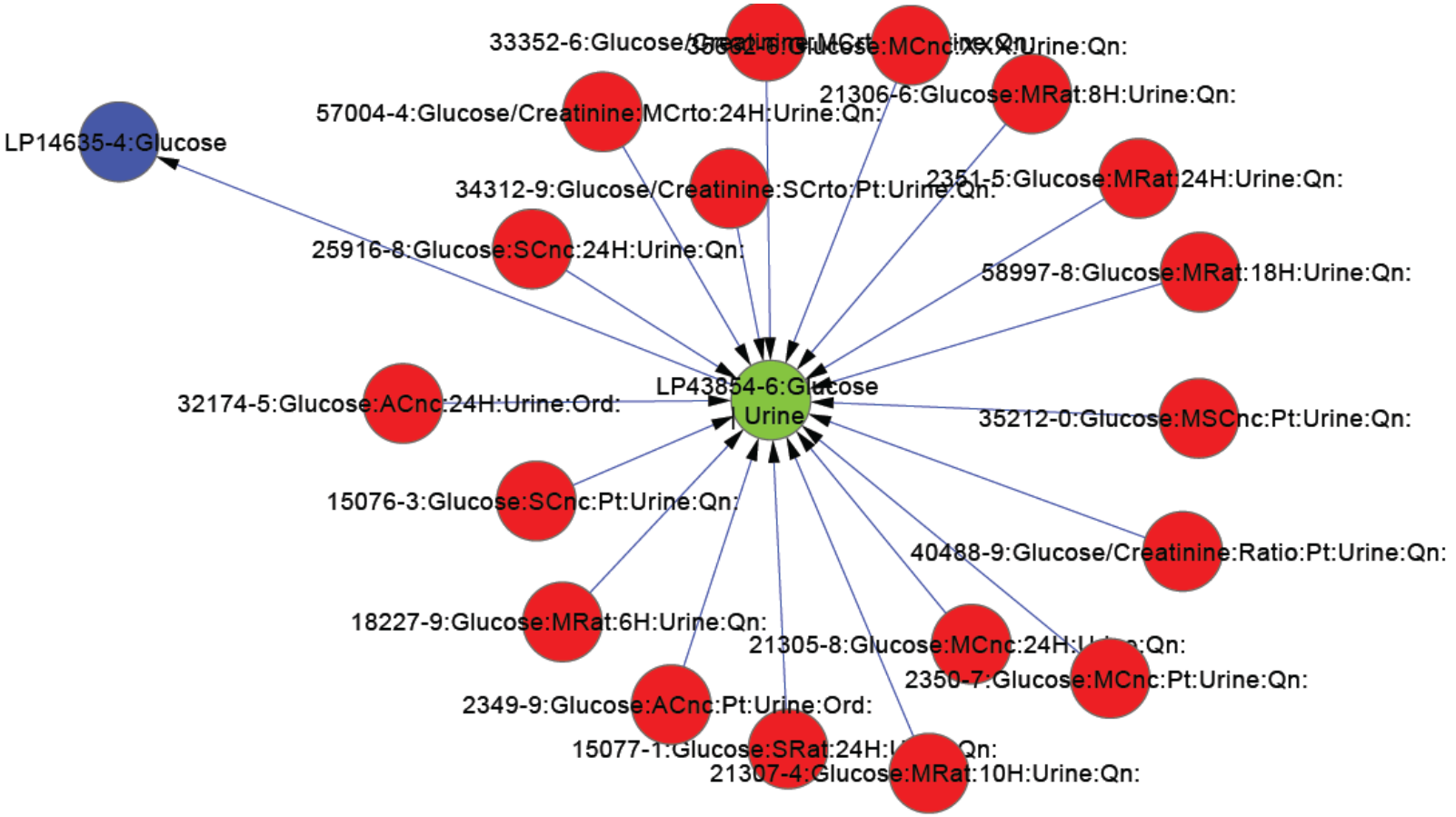


Legend:

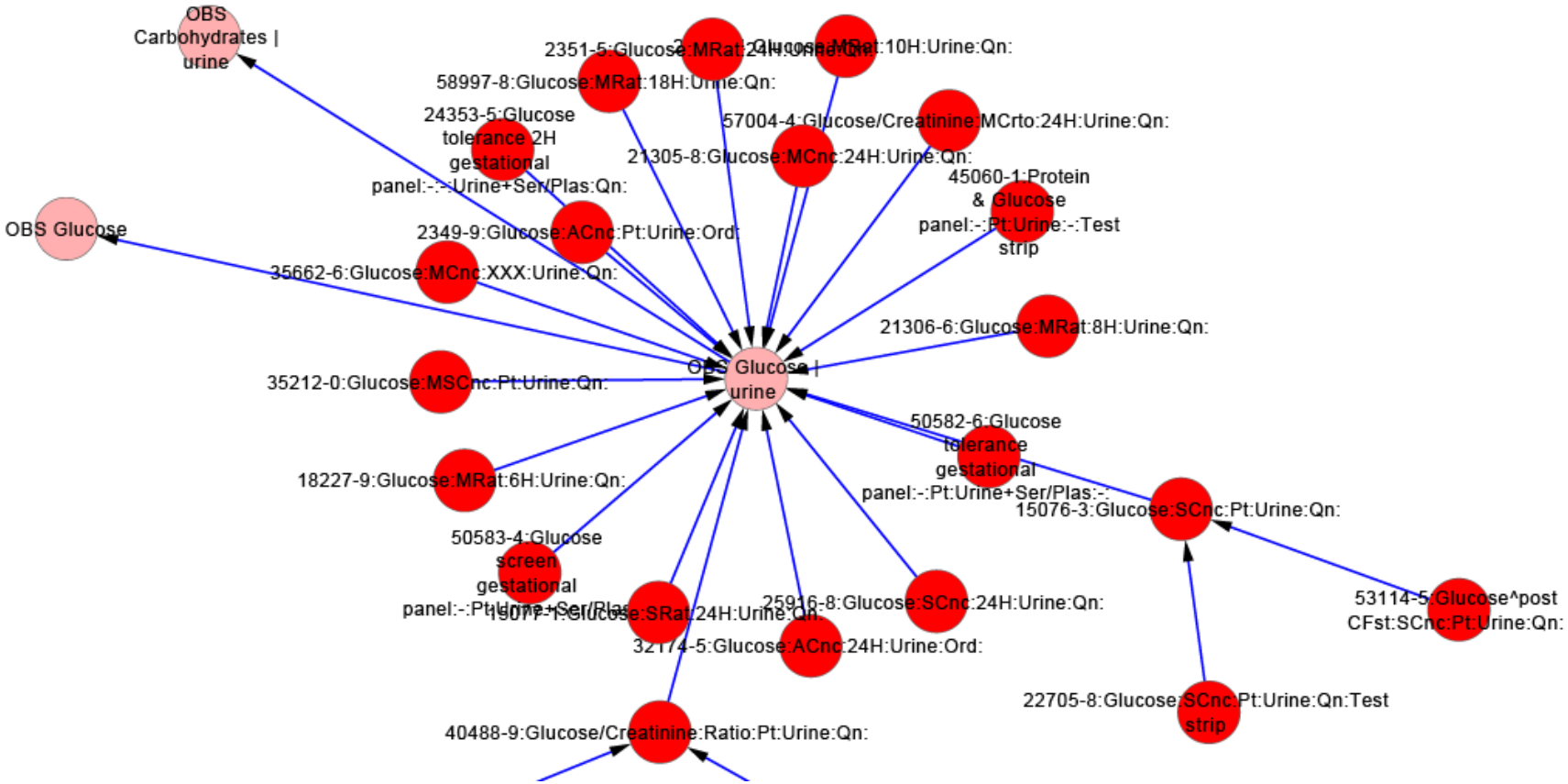
Test

Mx

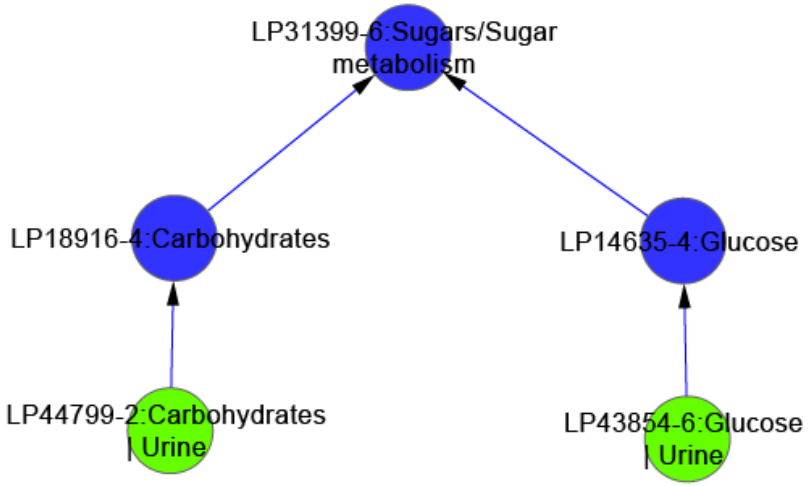
MULTIAXIAL LP43854-6:Glucose|Urine is an example of a grouping LOINC observation



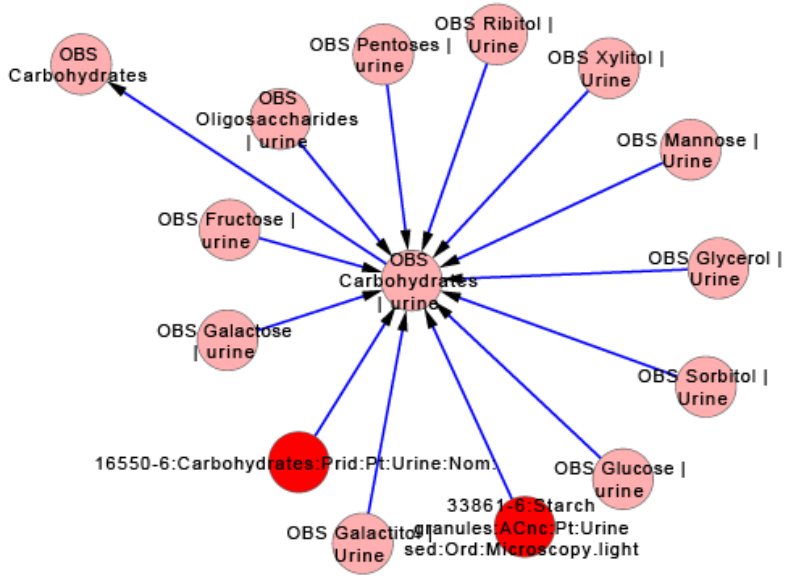
Inferred hierarchy provides new access points and codes subsumption



No direct path between *Carbohydrates | Urine* and *Glucose | Urine* originally



(a) LOINC



(b) Inferred

239 LOINC codes were found to be inconsistenly asserted in the hierarchy

- 183 concepts of scale type Document

- 28626-0:**History and physical note**:Find:Pt:Setting:Doc:Physician
  - Asserted *History and physical note*
  - Inferred *Note*

- Mostly insufficient modelling

Reasoner correctly infers them under Lipids | Bld-ser-plas

Category or ShortName	Component	System
Chemistry		
Lipids		
Fatty acids		
Saturated fatty acids		
Monounsaturated fatty acids		
Polyunsaturated fatty acids		
Fatty Acids.esterified | Bld-Ser-Plas		
Fatty acids.nonesterified		
Fatty acids.ethyl esters		
Fatty acids.long chain		
3-Hydroxy fatty acids		
7-hydroxyoctanoate | Urine		
Fatty acids [interpretation] in Serum or Plasma...	Fatty acids	Ser/Plas
Fatty acids [interpretation] in Serum or Plasma	Fatty acids	Ser/Plas
Fatty acids [Mass/volume] in Serum or Plasma	Fatty acids	Ser/Plas
Fatty acids [Moles/volume] in Serum or Plasma	Fatty acids	Ser/Plas

LOINC curators are doing a splendid job and the terminology is consistent

Significance of DL

1. Error detection
   a) Duplicates
   b) Missing hierarchical relations
   c) Inconsistencies in hierarchy
2. Enhanced navigation
3. Enhanced subsumption
4. Maintenance

Recommendations

1. Create logical definitions for codes
2. Have an inferred hierarchy
3. Parts vs. codes
4. Alignment with SNOMED CT

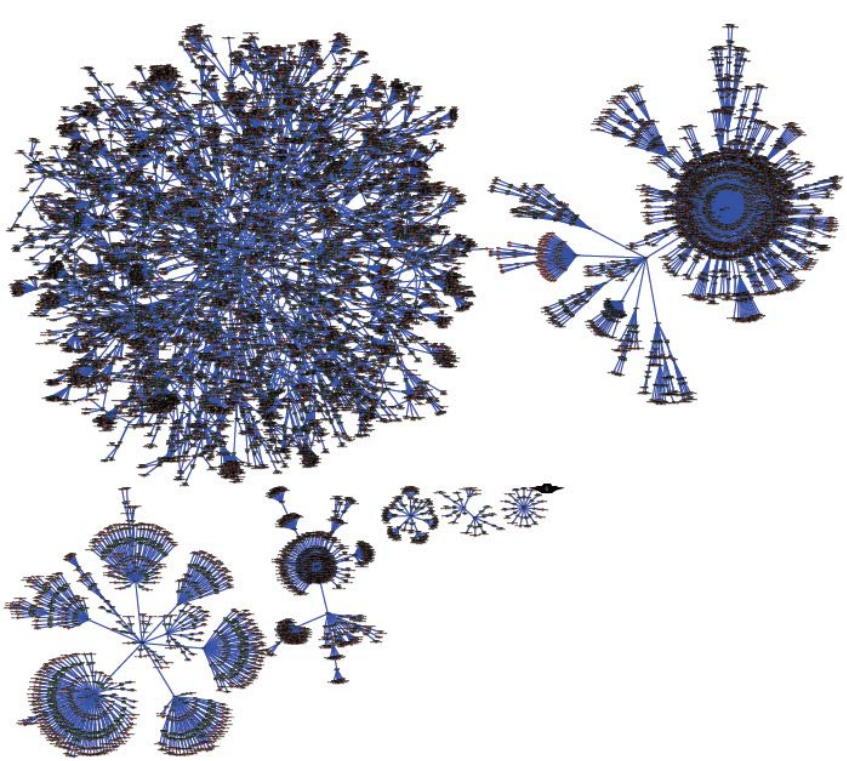
What does it mean to have several parts in LOINC map to SNOMED CT?

- SCT_3711007:Structure of great blood vessel (organ)
  - SYSTEM LP7303-3:Heart.great vessels
  - SYSTEM LP33690-6:Great vessel
  - SYSTEM LP30622-2:Great vessels
- SCT_66019005:Limb structure
  - COMPONENT LP121777-9:Extremity
  - SYSTEM LP7216-7:Extremities
  - SYSTEM LP7395-9:Limbs
  - SYSTEM LP29945-0:Extremity

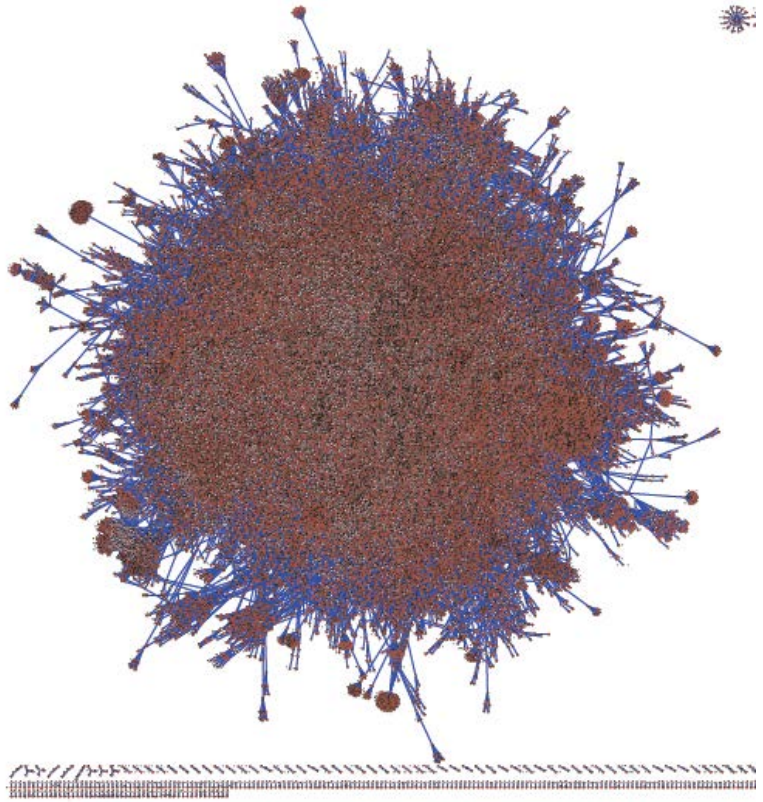
Limitations

- Relying on UMLS to provide mappings
- Imposing a specific ontological commitment
- Modelling with conjunctions likely suboptimal for more complex observations

Inferred is bigger and better ;)



MULTIAXIAL



INFERRED

Acknowledgments

- **Olivier Bodenreider MD PhD** (mentor)
- Bastien Rance PhD
- Rainer Winnenburg PhD
- Clement McDonald MD
- Daniel J. Vreeman PT DPT MSc (Regenstrief Institute)

This work was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM) and the Oak Ridge Institute for Science and Education (ORISE) Training Program in Clinical Informatics managed for the U.S. Department of Energy (DOE) by Oak Ridge Associated Universities (ORAU).

Thank you