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?


- 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 $\equiv$ 45662-4 Seizure disorder
  - LP7216-7:Extremities $\equiv$ LP7395-9:Limbs

- Infer a hierarchy
  - Glucose | Urine $\rightarrow$ Carbohydrates | Urine

- Find inconsistencies
  - 44084-2 Fatty acids in Serum or Plasma $\rightarrow$
  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


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:

<table>
<thead>
<tr>
<th>Part Type</th>
<th>Part No.</th>
<th>Part Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>LP14355-9</td>
<td>Creatinine</td>
</tr>
<tr>
<td>Property</td>
<td>LP6827-2</td>
<td>MCnc  [Mass Concentration]</td>
</tr>
<tr>
<td>Time</td>
<td>LP6960-1</td>
<td>Pt   [Point in time (spot)]</td>
</tr>
<tr>
<td>System</td>
<td>LP7576-4</td>
<td>Ser/Plas  [Serum or Plasma]</td>
</tr>
<tr>
<td>Scale</td>
<td>LP7753-9</td>
<td>Qn</td>
</tr>
</tbody>
</table>
METHODS
We used part links to create logical definitions for codes

<table>
<thead>
<tr>
<th>Part Type</th>
<th>Part Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Creatinine</td>
</tr>
<tr>
<td>Property</td>
<td>MCnc</td>
</tr>
<tr>
<td>Time</td>
<td>Pt</td>
</tr>
<tr>
<td>System</td>
<td>Ser/Plas</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn</td>
</tr>
</tbody>
</table>

Code:

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

DL definition:

\[(\text{has\_component\ some Creatinine})\ \text{and}\ \ (\text{has\_property\ some MCnc})\ \text{and}\ \ (\text{has\_time\_aspect\ some Pt})\ \text{and}\ \ (\text{has\_system\ some Ser/Plas})\ \text{and}\ \ (\text{has\_scale\ some Qn})\]
Component 2\textsuperscript{nd} subpart: challenge

Code:

<table>
<thead>
<tr>
<th>Part Type</th>
<th>Part No.</th>
<th>Part Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>LP14635-4</td>
<td>Glucose</td>
</tr>
<tr>
<td>Challenge</td>
<td>LP20355-1</td>
<td>post CFst</td>
</tr>
<tr>
<td>Property</td>
<td>LP6827-2</td>
<td>MCnc  ([\text{Mass Concentration}])</td>
</tr>
<tr>
<td>Time</td>
<td>LP6960-1</td>
<td>Pt  ([\text{Point in time (spot)}])</td>
</tr>
<tr>
<td>System</td>
<td>LP7576-4</td>
<td>Ser/Plas  ([\text{Serum or Plasma}])</td>
</tr>
<tr>
<td>Scale</td>
<td>LP7753-9</td>
<td>Qn</td>
</tr>
</tbody>
</table>
# Component 3rd subpart: adjustment

## Code:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>23811-3</td>
<td>Alpha-1-Fetoprotein [Multiple of the median] adjusted in Serum or Plasma</td>
</tr>
</tbody>
</table>

## Parts:

<table>
<thead>
<tr>
<th>Part Type</th>
<th>Part No.</th>
<th>Part Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>LP14331-0</td>
<td>Alpha-1-Fetoprotein</td>
</tr>
<tr>
<td>Adjustment</td>
<td>LP20174-6</td>
<td>adjusted</td>
</tr>
<tr>
<td>Property</td>
<td>LP71590-1</td>
<td>MoM [Multiple of the median]</td>
</tr>
<tr>
<td>Time</td>
<td>LP6960-1</td>
<td>Pt [Point in time (spot)]</td>
</tr>
<tr>
<td>System</td>
<td>LP7576-4</td>
<td>Ser/Plas [Serum or Plasma]</td>
</tr>
<tr>
<td>Scale</td>
<td>LP7753-9</td>
<td>Qn</td>
</tr>
</tbody>
</table>
LOINC parts are not available in the public release (2.36)
Multiaxial hierarchy in LOINC could be vastly improved with DL

<table>
<thead>
<tr>
<th>Type</th>
<th>Category or ShortName</th>
<th>Component</th>
<th>System</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chemistry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sugars/Sugar metabolism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose [Mass/time] in 10...</td>
<td>Glucose</td>
<td>Urine</td>
<td>21307-4</td>
</tr>
<tr>
<td></td>
<td>Glucose [Mass/time] in 18...</td>
<td>Glucose</td>
<td>Urine</td>
<td>58997-8</td>
</tr>
<tr>
<td></td>
<td>Glucose [Presence] in 24...</td>
<td>Glucose</td>
<td>Urine</td>
<td>32174-5</td>
</tr>
<tr>
<td></td>
<td>Glucose [Mass/volume] in...</td>
<td>Glucose</td>
<td>Urine</td>
<td>21305-8</td>
</tr>
<tr>
<td></td>
<td>Glucose [Mass/time] in 24...</td>
<td>Glucose</td>
<td>Urine</td>
<td>2351-5</td>
</tr>
<tr>
<td></td>
<td>Glucose [Moles/volume] in...</td>
<td>Glucose</td>
<td>Urine</td>
<td>25916-8</td>
</tr>
<tr>
<td></td>
<td>Glucose tetrasaccharide</td>
<td>Urine</td>
<td>Glucose tetrasaccharide/Creatinine</td>
<td>Urine</td>
</tr>
<tr>
<td></td>
<td>Chemistry, challenge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urinalysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analytes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose [Presence] in Urine...</td>
<td>Glucose</td>
<td>Urine</td>
<td>25428-4</td>
</tr>
<tr>
<td></td>
<td>Glucose [Presence] in Urine...</td>
<td>Glucose</td>
<td>Urine</td>
<td>50555-2</td>
</tr>
<tr>
<td></td>
<td>Glucose [Mass/volume] in...</td>
<td>Glucose</td>
<td>Urine</td>
<td>5792-7</td>
</tr>
<tr>
<td></td>
<td>Glucose [Mass/volume] in...</td>
<td>Glucose</td>
<td>Urine</td>
<td>53328-1</td>
</tr>
<tr>
<td></td>
<td>Glucose [Moles/volume] in...</td>
<td>Glucose</td>
<td>Urine</td>
<td>22705-8</td>
</tr>
<tr>
<td></td>
<td>Glucose [Moles/volume] in...</td>
<td>Glucose</td>
<td>Urine</td>
<td>59156-0</td>
</tr>
</tbody>
</table>

Screenshot from the Regenstrief LOINC Mapping Assistant (RELMA)
Multiaxial hierarchy in LOINC could be vastly improved with DL.
Separated codes and parts and defined corresponding observations

- Multiaxial
  - Glucose
  - Glucose | Urine
  - Glucose in 10 hour Urine
  - Glucose in Urine by Test strip

- Urine

- Inferred
  - Protein & Glucose panel in Urine by Test strip
SNOMED CT compensates for missing parts relations in LOINC

Body fluid LP30504-2 owl:EquivalentTo Body fluid 32457005

Body Fluids C0005889 ISA

Urine LP7681-2 owl:EquivalentTo Urine 78014005

Urine C0042036
We can identify semantically equivalent LOINC parts via UMLS

Erythrocyte LP16699-8

RBC LP7536-8

Erythrocytes LP14304-7

Erythrocytes C0014792
Reasoner infers logical consequences from a set of asserted facts or axioms

OBS Glucose | Urine

Glucose in 10 hour Urine

has_component some Glucose and
has_system some Urine

has_component some Glucose and
has_property some Arbitrary Concentration and
has_time_aspect some Point in time (spot) and
has_system some Urine and
has_scale some Ord and
has_method some Test strip

DL definition
Huge Knowledge Base classified with ConDOR reasoner

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

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+

c) DL definition

... and (has_component some Cells.CD3+CD56+)
and (has_component some Cells.CD3-CD56+)
LOINC must have realised the problem

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>56897-2</td>
<td>CD3-CD56+ cells/100 cells in Cerebral spinal fluid</td>
</tr>
</tbody>
</table>

**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

Monocytes+Macrophages
LP14312-0

Monocytes+Macrophages
/100 leukocytes in Peritoneal fluid by Manual count (32029-1)

Macrophages
/100 leukocytes in Peritoneal fluid by Manual count (40517-5)

Macrophages
LP14314-6
Pop quiz: removing which `has_component` relation changes equivalence to subsumption?

Monocytes+Macrophages
LP14312-0

Monocytes+Macrophages
/100 leukocytes in Peritoneal fluid by Manual count
(32029-1)

Macrophages
/100 leukocytes in Peritoneal fluid by Manual count
(40517-5)

Macrophages
LP14314-6
Issues with referential integrity

Type of Enema device (8950-8)

* LP28805-7

Type of Enema device (8932-6)

Enema device
LP7209-2
SNOMED CT enrichment gives 102 sets of equivalent LOINC codes

- 46062-6: **Treatments**: 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
Schistocytes [Presence] in Blood by Light microscopy (800-3)

Helmet cells [Presence] in Blood by Light microscopy (10374-7)

LP14570-3:Helmet cells
LP14738-6:Cells
LP29945-0:Schistocytes

(has_component some 'Helmet cells') and (has_component some Cells)

SCT_70310009: Helmet cell
is_a SCT_362837007:Entire cell
Infered hierarchy has more connected nodes and is better connected

<table>
<thead>
<tr>
<th></th>
<th>LOINC</th>
<th>Inferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of connected nodes</td>
<td>73,387</td>
<td>82,350</td>
</tr>
<tr>
<td>Network diameter</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Connected components</td>
<td>8</td>
<td>513</td>
</tr>
<tr>
<td>Shortest paths</td>
<td>425,976</td>
<td>1,119,232</td>
</tr>
<tr>
<td>Characteristic path length</td>
<td>3.39</td>
<td>3.81</td>
</tr>
<tr>
<td>Average number of neighbours</td>
<td>2.01</td>
<td>3.40</td>
</tr>
</tbody>
</table>
Inferred nodes are better connected locally

![Graph showing logarithm of average connectivity vs logarithm of number of neighbours for LOINC and inferred nodes.](image)
Find all carbohydrate observations

Regenstrief LOINC Mapping Assistant (RELMA)
Find all carbohydrate observations?!
It is not easy

Gene tests

HLA tests

Evaluation and management

Skin tests

Patient information

HPA tests

Everything else

Here Be Dragons
COMPONENT LP14635-4: Glucose is the most connected node
COMPONENT LP14635-4: Glucose is the most connected node
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
239 LOINC codes were found to be inconsistently 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

<table>
<thead>
<tr>
<th>Category or ShortName</th>
<th>Component</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturated fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monounsaturated fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyunsaturated fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty Acids.esterified</td>
<td>Bld-Ser-Plas</td>
<td></td>
</tr>
<tr>
<td>Fatty acids.nonesterified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty acids.ethyl esters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty acids.long chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-Hydroxy fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-hydroxyoctanoate</td>
<td>Urine</td>
<td></td>
</tr>
<tr>
<td>Fatty acids [interpretation] in Serum or Plasma</td>
<td>Fatty acids</td>
<td>Ser/Plas</td>
</tr>
<tr>
<td>Fatty acids [interpretation] in Serum or Plasma</td>
<td>Fatty acids</td>
<td>Ser/Plas</td>
</tr>
<tr>
<td>Fatty acids [Mass/volume] in Serum or Plasma</td>
<td>Fatty acids</td>
<td>Ser/Plas</td>
</tr>
<tr>
<td>Fatty acids [Moles/volume] in Serum or Plasma</td>
<td>Fatty acids</td>
<td>Ser/Plas</td>
</tr>
</tbody>
</table>
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 ;)
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