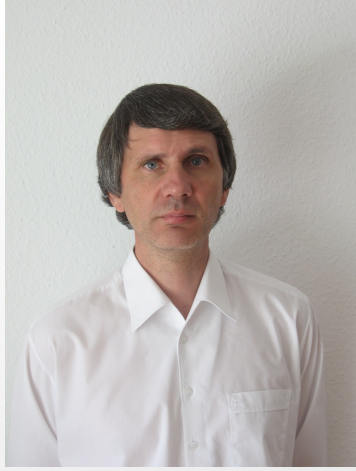

Generating FDA-ready submission datasets for new COVID-19 therapies directly from Electronic Health Records using the new COVID-19 LOINC codes

Jozef Aerts
XML4Pharma

LOINC® from Regenstrief

A bit about Jozef …

- CDISC (Clinical Data Interchange Standards Consortium) volunteer since 2001
- Co-developer of several CDISC Standards
- Software development for use in clinical research in combination with standards
- LOINC, UCUM and FHIR aficionado
- CEO of XML4Pharma
  - and former Professor in Medical Informatics in Graz (Austria)

Submission of Clininical Research Data to FDA and post-coordinated Terminolgy

- For e-submissions to FDA, pharma companies ("sponsors") need to transform their captured study data to a standard, named "Submission Dataset Tabulation Standard" (SDTM), which essentially is a set of electronic tables.
- Many of the variables in SDTM are under controlled terminology
- For "Findings" (observations), this controlled terminology is mostly post-coordinated

USUBJID	MBS...	...	MBTESTCD	MBTEST	MBTSTDTL	MBLOC	MBMETHOD
1171862	1	...	SAR2RNA	SARS-CoV-2 RNA	DETECTION	RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST
1171863	1	...	SAR2RNA	SARS-CoV-2 RNA	DETECTION	RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST
1171864	1	...	SAR2RNA	SARS-CoV-2 RNA	DETECTION	RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST
1171865	1	...	SAR2RNA	SARS-CoV-2 RNA	DETECTION	RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST

The Use of post-coordinated Controlled Terminology in SDTM - Why?

- The reason is probably historical …
- Study protocols do not exactly describe what tests need to be performed
- Measure "glucose in urine" is often encountered in study protocols
- The sponsor either further specifies this in the "instructions to the site", or leaves it to the the study site to interprete this
- Sponsors do not know LOINC
- This often results in a variety of tests being performed, even for the same analyte and original purpose

The Use of post-coordinated Controlled Terminology in SDTM - Why?

- With such a variety of tests for the same analyte, all one can do is a post-study categorization => post-coordinated terminology

- Idea is "I do not know what will come, I will categorize afterwards"

Then, in 2017, things changed

- In 2017, FDA announced it will start requiring to submit the LOINC code for lab tests in SDTM submissions

- This caused a lot of panic at CDISC
  (LOINC is still "not invented here" at CDISC

- Together with Regenstrief, a project was started to map the most common LOINC lab test codes to CDISC (post-coordinated) controlled terminology

- The results of this mapping has just (September 2020) been published by CDISC

The LOINC-CDISC mapping

- Restricted to 2300 mappings for 1300 LOINC codes

- Some LOINC codes have multiple mappings
  E.g. CDISC does not like "Ser/Plas" => 3 mappings "SERUM", "PLASMA", "SERUM OR PLASMA"

- CDISC statement: "… *the team feels very strongly that people should NOT be using this file for electronic or automated processing. The intention of this file is simply to show examples, it is not meant for purposes beyond that* …"

The LOINC-CDISC mapping

- Res
- Son
  E.g. OR
  PLA
- CDI *this file*
  for amples, it
  is n

What the hell should we use this mapping for, other than use it or automation?

Should we just use it for staring at it?

When using it in automation, we soon found out that …

- Mappings for 1,300 LOINC codes is far from sufficient
- Starting from LOINC codes is a huge opportunity to automate the generation of SDTM-LB datasets
  - Usually, the generation of the LB dataset takes weeks, due to categorization steps and programming these
  - When automating using the (extended) LOINC-SDTM-LB mappings, LB datasets can be generated in minutes instead of taking weeks
  - This of course requires the LOINC code is available
- The same applies if we had mappings for vital signs, ECG, questionnaires …

And then came COVID-19

- LOINC published a lot of new codes for Corona Virus tests
- CDISC published a "COVID-19 Interim User Guide"
  (but did even not mention the new LOINC codes …)
- Most of the new LOINC codes map to the CDISC Microbiology (MB)

- But there is no mapping available between LOINC and MB!
  (and CDISC said it will not develop one: "*LOINC in MB is not an FDA requirement …*"

https://www.cdisc.org/standards/therapeutic-areas/covid-19

So we developed a LOINC-MB mapping for Corona-virus tests ourselves!

Developing the LOINC-MB Mapping for the new LOINC Corona Virus tests

- Uses the same principles as for the CDISC LOINC-LB mappings
- Required a lot of discussions with, and numerous "new term requests" to the CDISC "Controlled Terminology Team"
  - Lots of the CDISC-CT for virus tests was not existing yet
- We then made the mappings available through a RESTful web service, so that it can be used in any modern software

http://xml4pharmaserver.com/WebServices/LOINC2CDISC_webservices.html

Problems encountered during developing the LOINC-MB Mapping for the new LOINC Corona Virus tests

- In SDTM "TESTCD" (test code) means something different depending on the domain
  - LB: the analyte
  - VS: the physical property (systolic blood pressure, weight …)
  - MB: the species investigated
- Example: LOINC code 94509-7: "SARS-related coronavirus E gene …"
  - What is the species?
  - After discussion with CDISC-CT:
    - MBTESTCD = SAR2RNA
    - MBTEST = SARS-CoV-2 RNA
    - SUPPMB.MBSYM = N gene

What is "SUPPMB"?

- Subsequent versions of SDTM are "repairs" of the prior version
- Each time over and over again, CDISC finds out that the current SDTM model is insufficient to describe the medical world
- Things that "just don't fit" are put in so-called "Supplemental Qualifiers", also named "Non-standard Variables" (NSV)
- Often, in the next version of SDTM, these than become "standard"

- Such an "NSV" is "Genomic Symbol" - SUPPLB.MBSYM

The LOINC-MB Corona Virus Test Mappings Implementation

- As of date (October 2020): 82 mappings for 131 LOINC codes

Made available as a RESTful Web Service – see:

http://xml4pharmaserver.com/WebServices/LOINC2CDISC_webservices.html


```
<XML4PharmaServerWebServiceResponse ServerDateTime="2020-10-02T09:10:14">
  <WebServiceRequest>http://www.xml4pharmaserver.com:8080/CDISCCTService2/rest/LOINC2SDTMCorona/94500-6</WebServiceRequest>
  <Remark>This is a preliminary mapping! Results with a * in front are placeholders for future CDISC Controlled Terminology.</Remark>
 <Response>
  <LOINC2SDTMMapping MappingSource="XML4Pharma">
    <MBTESTCD>SAR2RNA</MBTESTCD>
    <MBTEST>SARS-CoV-2 RNA</MBTEST>
    <MBTSTDTL>DETECTION</MBTSTDTL>
    <MBCAT/>
    <MBPOS/>
    <MBLOINC>94500-6</MBLOINC>
    <MBSPEC/>
    <MBLOC>RESPIRATORY SYSTEM</MBLOC>
    <MBMETHOD>NUCLEIC ACID AMPLIFICATION TEST</MBMETHOD>
    <MBANMETH/>
    <MBFAST/>
    <MBTPT/>
    <MBEVLINT/>
    <MBEVINTX>PT</MBEVINTX>
    <SUPPMB.MBRESTYP>PRESENCE OR THRESHOLD</SUPPMB.MBRESTYP>
    <SUPPMB.MBRSLSCL>ORDINAL</SUPPMB.MBRSLSCL>
    <SUPPMB.MBSYM/>
   </LOINC2SDTMMapping>
  </Response>
</XML4PharmaServerWebServiceResponse>
```

Why a RESTful Web Service?

- New mappings (as new LOINC codes are published) can easily be added to the underlying database
- RESTful-WS can easily be consumed by any modern computer program. Current speed is about 20 requests treated, per second
- Consumer (program) can use the information returned in any way it wants
- No "hustling" with Excel files … 

What can one do with the RESTful Web Service?

- Electronic Health Records (EHRs) usually describe "observations" by the LOINC code
- EHRs that implement the HL7-FHIR interface are ideally suited as a source for study data in clinical research (avoids double data entry)
- Experience with transforming FHIR resource instances to CDISC-SDTM is growing
- Largest problem is the difference in coding systems used by CDISC
- Our mappings solve most of this problem
- The implementation as a RESTful web service enables the automated generation of CDISC SDTM datasets directly from FHIR resources

Automated Generation of SDTM-MB datasets directly from FHIR resources using the LOINC-MB Mapping for Corona Virus Tests

- Open Source project FHIRLOINC2SDTM
- Source code available at: https://sourceforge.net/projects/fhirloinc2sdtm
- Meant as pilot, proof-on-concept
- Almost daily new code added
- Output in modern CDISC Dataset-XML format, enabling FHIR source resources to be embedded into the SDTM record (traceability)

Demo time!

Visualizing the results

For visualization, we use the open source
"Smart Submission Dataset Viewer"
https://sourceforge.net/projects/smart-submission-dataset-viewe/

- The viewer itself uses a lot of RESTful web services as optional features, e.g. to show the meaning of a LOINC code to the reviewer

- The viewer also allows to visualize the embedded source FHIR resource (currently as a tooltip)

Extending the CDISC LOINC-LB Mapping

- The initial FHIRLOINC2SDTM project gave us appetite for more …

- If we can automatically generate CDISC SDTM-MB datasets for the Corona Virus tests, why wouldn't this work for other CDISC domains.
  Obvious candidates are:
  - Laboratory
  - Vital Signs
  - ECG
  - Questionnaires

Extending the CDISC LOINC-LB Mapping

- Initial tests indicated that mappings for 1,400 LOINC (laboartory codes) is insufficient
- So the question arised: is it possible to extend the LOINC-LB mapping considerably?
- A fully manual approach as used by the CDISC team is out of the question: it took the CDISC team 2 years to develop the mappings for the 1,400 LOINC codes
- Is an automated or semi-automated possible?

- Post-generation curation will however be necessary anyway

Extending the CDISC LOINC-LB Mapping: Principles

- LOINC "Component" maps to LBTESTCD/LBTEST
  - But need to take care of "subcomponents" and of "Ser/Plas", "Ser/Plas/Bld" etc.
- LOINC "Property" maps to SUPPLB.MBRESTYP
  - e.g. "ACnc" => "ARBITRARY CONCENTRATION"
- LOINC "Time Aspect" maps to either:
  - SUPPLB.LBPTFL=Y ("point in time flag") when "Pt"
  - SUPPLB.LBPDUR ("planned duration") when not "Pt"
- LOINC "System" maps to either:
  - LBLOC ("Location")
  - LBSPEC ("Specimen")
- LOINC "Method" maps to MBMETH (method), sometimes to LBANMETH ("Analysis Method")
- LOINC "Scale" maps to SUPPLB.LBRSLSCL ("Result Scale"), e.g. "QUANTITATIVE"

Extending the CDISC LOINC-LB Mapping: Difference between xxLOC and xxSPEC

- What is the difference between SDTM xxLOC (Location) and xxSPEC (Specimen)
- For example, when I take a hair, when it is still on my head, it is LBLOC, when I pull it out, it is LBSPEC
  - Other example: "Interstitial Fluid"
- Differentiating between is not alway easy => curation
- Personally, I consider this as an SDTM design error

Extending the CDISC LOINC-LB Mapping: Strategy and Algorithm

- For the 2300 existing LOINC-to-LB mappings,
  take the properties "Component", "Property", "System", "Scale", "TimeAspect" and "Method"
- Fix values of 5 of them, and allow the 6th to vary => query the LOINC database
  - With exception of "Component", would lead to too many for which no CDISC-CT exists
  - In case of fixing "method", also allow "methodless"
  - => 6 "scenarios"
- This results in a list of LOINC codes
- For each of them:
  - Check whether already in the mapping, if so, skip
  - Take the value of the variable that was allowed to vary, look whether a mapping for that variable exists. If so, use it, if not, copy the value (as an educated guess)

Extending the CDISC LOINC-LB Mapping: Strategy and Algorithm

- This resulted in >7,500 new mappings, BUT:
- Still need to multiplicate mappings for case "Ser/Plas" and "Ser/Plas/Bld" and a few other cases (where "or" exists)
- Remove duplicate mappings
- Where necessary, make decision whether value of "System" goes into LBLOC (Location) or into LBSPEC (specimen)
- Take care of composite "Component", e.g. "Glucose^post CFst"
- Merge mappings of the different scenarios
- Remove duplicates again

- The whole "curation" process took about 1 man-week

Extending the CDISC LOINC-LB Mapping: Implementation in DB and RESTful-WS

- This resulted in 5,400 additional mappings

- We decided to merge these into the existing database, with an additional column "source" (can be either "CDISC" or "XML4Pharma)
  - Also added a column "UCUM example units", as that was omitted by CDISC (CDISC does not like UCUM)

- To allow the receiving application to filter on whether the mapping comes from CDISC or from us, we added an attribute "MappingSource"

Extending the CDISC LOINC-LB Mapping: Example result

```
▼<XML4PharmaServerWebServiceResponse ServerDateTime="2020-10-09T10:38:59">
  <WebServiceRequest>http://www.xml4pharmaserver.com:8080/CDISCCTService2/rest/LOINC2SDTMLB/83103-2</WebServiceRequest>
 ▼<Response>
  ▼<LOINC2SDTMMapping MappingSource="XML4Pharma" TargetSDTMDomain="LB">
     <LBTESTCD NCICode="C74790">LH</LBTESTCD>
     <LBTEST NCICode="C74790">Luteinizing Hormone</LBTEST>
     <LBORRESU_Example NCICode="C67377">IU/mL</LBORRESU_Example>
     <LBPOS/>
     <LBLOINC>83103-2</LBLOINC>
     <LBSPEC NCICode="C105706">SERUM OR PLASMA</LBSPEC>
     <LBLOC/>
     <LBMETHOD NCICode="C16714">IMMUNOASSAY</LBMETHOD>
     <LBANMETH/>
     <LBFAST/>
     <LBTPT/>
     <SUPPLB.LBPTFL>Y</SUPPLB.LBPTFL>
     <SUPPLB.LBPDUR/>
     <SUPPLB.LBRESTYP>ARBITRARY CONCENTRATION</SUPPLB.LBRESTYP>
     <SUPPLB.LBRSLSCL>QUANTITATIVE</SUPPLB.LBRSLSCL>
     <SUPPLB.LBTSTOPO/>
     <SUPPLB.LBLLOD/>
     <SUPPLB.LBTSTCND/>
     <SUPPLB.LBMTHSEN/>
     <Example_UCUM_Units>[IU]/mL</Example_UCUM_Units>
    </LOINC2SDTMMapping>
  </Response>
</XML4PharmaServerWebServiceResponse>
```

Demo:
HAPI-FHIR
Diabetes Type 2:
Auto-generated
SDTM-LB dataset

With conversion from US-Conventional to SI units

Generation of a LOINC-VS Mapping for Vital Signs

- Started from set of LOINC panels that have "*Vital*" in the "Class"
- Drilled down to individual tests
- Special care needs to be taken when "Component" is composite, e.g.:
  "Intravascular systolic^standing"
  - Intravascular systolic => VSTESCD=SYSBP, VSTEST='SYSTOLIC BLOOD PRESSURE'
  - Standing => VSPOS=STANDING
  - Each such cases must be taken care of one by one …

LOINC-VS Mapping for Vital Signs
Current status

- 617 mappings for 578 LOINC codes
- Occasionally, one or more codes are added
- RESTful Web Service is already available and running

Are all these mappings really necessary?

- Unfortunately, currently they are …
- In SDTM, xxTESTCD, xxTEST, xxSPEC, xxLOC, … are required attributes
- Only delivering the LOINC and omit values for these variables would lead to rejection of the submission (FDA, PMDA, NMPA)
- This is essentially ridiculous, as all the information is already in the LOINC code, and can easily be visualized to the reviewer
- The CDISC SDTM team will however probably never give up the requirement rules for these attributes
  - LOINC is still "not invented here" at CDISC

Are all these mappings really necessary? Everything is already in the LOINC code

VSLOINC	VSLOC	VSLAT	VSMETHOD	VSBLFL	V
8302-2					
29463-7					
59576-9					
39156-5					
59576-9					
39156-5					
8302-2					
29463-7					
8302-2					
29463-7					
59576-9					
39156-5					
29463-7					

8302-2 (VSLOINC)
LOINC Name: Body height:Len:Pt:^Patient:Qn
LOINC Common Name: Body height
LOINC Class: BDYHGT.ATOM (General body height (length))
Component/Compound: Body height
Property: Len (Length)
Time Aspect: Pt (Point in time (spot))
System: ^Patient
Scale: Qn (Quantitative)
Example UCUM Units: [in_i]

Summary

- By developing mappings for Corona Virus to the CDISC Microbiology domain, generation of FDA-submission-ready SDTM datasets from EHR systems that have a FHIR interface, can be automated
- We made the by CDISC published LOINC-LB mappings available as a RESTful web service, allowing any modern software to use these
- This can also be used to auto-generate SDTM-LB datasets
- The by CDISC published LOINC-LB mappings were extended with over 5,400 mappings, and also be made available over the RESTful web service
- A set of >600 mappings between vital signs LOINC codes and the SDTM-VS domain was developed and implemented as a RESTful web service
- Mappings for ECG tests are in preparation

Thank you!

LOINC and UCUM:
also for clinical research and regulatory submissions!

Mail: Jozef.Aerts@XML4Pharma.com