SDTM-ETL 4.0: Performing Unit Conversions in SDTM-ETL

Author: Jozef Aerts, XML4Pharma

Last update: **2020-09-19**

Introduction

For SDTM/SEND Findings domains, unit conversions are often necessary, especially for the population of -STRESN (numeric standardized result) variable. Classically, this requires a lot of programming. Unfortunately, CDISC controlled terminology (CDISC-CT) does not provide us a way to do such conversions. For example, "inches" and "cm" are both listed as units for "length" in CDISC-CT, but the conversion factor is not provided in a machine-readable way. For laboratory results, the situation is very often even more complicated. For example, if I need to convert/standardize an albumin concentration from "US conventional" units like mg/dL to "SI" units like mmol/L, how can I automate this? I probably need to do a lookup in conversion tables or conversion factor lists, and then program that into my mapping, and that for each different analyte.

With LOINC and UCUM however, everything becomes much more simple.

Using LOINC and UCUM

For laboratory results, the FDA requires to submit the LOINC code for new studies. For lab tests that measure a quantity, there are often two LOINC codes, one for measurements using "conventional units" (e.g. mass/volume), and one for "SI units" (e.g. moles/volume). We can easily use this fact for making unit conversions easy.

Each LOINC code essentially represents 5 or 6 parts, one of them being the chemical entity (analyte) to be measured or tested for, when the LOINC code is about a laboratory test[1]. The analyte part (which has a code itself) has an associated molecular weight (e.g. 180.156 for glucose), which is then used for "conventional" to "SI" units conversion (first chemistry class in secondary school…).

UCUM is the worldwide standard for notation of units. It is used in aerospace, engineering, in healthcare (especially in electronic health records – EHRs). Unfortunately, CDISC decided against UCUM many years ago, and still develops its own controlled terminology for units, which is not a notation, not a system, but a list. Whereas CDISC-CT does not allow unit conversions, UCUM has been developed in such a way (as a system) that unit conversions become very easy. This is done by the use of the famous "ucum-essence-xml" file, which can easily be used in software programs.

The **good news** however is that for concentrations, the UCUM-notation is very often identical to the CDISC unit from the list. "mg/dL" and "mmol/L" are both valid in UCUM and in CDISC-CT, and also mean exactly the same thing. So very often, we will still be able to use a "CDISC-CT unit" in UCUM conversions, as they are identical.
For other properties, we will be mostly out of luck. For example, the CDISC-CT pressure unit

[1] LOINC is much more than only about laboratory tests! It is about any kind of tests in healthcare, like vital signs test, ECG tests, standardized questionnaires, and much much more.

"mmHg" is not valid in UCUM, there "mm[Hg]" is used. And that for a very good reason, as it allows conversions to other pressure units, such as "mm[H2O]" (millimeter water column) or "bar" or "Pascal". Such conversions cannot be done when using CDISC-CT units.

But for concentrations, we are mostly safe.

We also expect that the FDA will soon make the use of UCUM notation in electronic submissions mandatory. It does it already mandate UCUM for SPL and has also included UCUM in the "standards catalog". So, we expect that CDISC-CT for units will soon become obsolete, and be replaced by UCUM.

RESTful web services

Although clinical research is, as usual, tremendously behind in implementation of modern IT, it is now more and more following healthcare, where RESTful web services are already used for a longer time.
Essentially, a RESTful web service is nothing else than a service offered by a server to any other application in the world over the internet. Usually, the service request is simply an HTTP call, and the response is usually either coming in the form of XML or of JSON.

For example, when I use the HTTP call

https://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/2.5/from/mg/dL/to/mmol/L/LOINC/2339-0

to the NLM server, requesting the service "ucumtransform" to convert 2.5 mg/dL to mmol/L for the analyte of the LOINC test with code 2339-0 (glucose mass concentration in blood), the response comes in the form of XML:

```
▼<UCUMWebServiceResponse ServerDateTime="2020-09-18T17:54:31">
   <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/2.5/from/mg/dL/to/mmol/L/LOINC/2339-0</WebServiceRequest>
  ▼<Response>
     <SourceQuantity>2.5</SourceQuantity>
     <SourceUnit>mg/dL</SourceUnit>
     <TargetUnit>mmol/L</TargetUnit>
     <MolecularWeightUsed>180.156</MolecularWeightUsed>
     <ResultQuantity>0.13876862</ResultQuantity>
   </Response>
 </UCUMWebServiceResponse>
```

This XML can easily be parsed by any program in any modern computer language (each modern computer language also has full support for RESTful web services). The value in "ResultQuantity" then tells our software that 2.5 mg/dL of glucose in blood (LOINC code 2339-0) corresponds to 0.139 mmol/L.
As the amount of information exchanged is also really small, the speed is very high. The NLM service for example, easily handles 50-100 requests per second. Smart client programs using the service can even come to much higher conversion speeds by caching results.

SDTM-ETL and unit conversions

The NLM server offers three services which are of importance for us:

- Simple conversions where no LOINC code nor molecular weight is needed. Examples are inch to centimeter ([in_i] to cm), or mg/dL to g/L. For these, we developed the SDTM-ETL function:

  rws:conversion*(string quantity, string source-unit, string target-unit*). For example:

$lbstresn = ***rws:conversion****('2.5', 'mg/dL', 'g/L'*).
Of course, you will usually not hardcode the strings in your SDTM-ETL mapping scripts but use variables, that are the passed to the function. For example:

$quantity = ...
$source = ...
$target =
$lbstresn = rws:conversion($quantity, $source, $target);

- "Conventional" to "SI" unit conversions and vice versa using the LOINC code of the test. In the case of laboratory tests, you will have the LOINC code already (in LBLOINC), as this is an FDA requirement anyway. The function is:

  rws:unitconversionLoinc*(string quantity, string source-unit, string target-unit, string loinc-code)*

  For example:

  $lbstresn = rws:unitconversionLoinc('2.5', 'mg/dL', 'mmol/L', '2339-0')

  which will return a value of "0.1388".

- In the case that you don't have a LOINC code for the test, but you have the molecular weight of the analyte, you can use the function:

  rws:unitconversionMW*(string quantity, string source-unit, string target-unit, string molecular-weight)*

  For example:

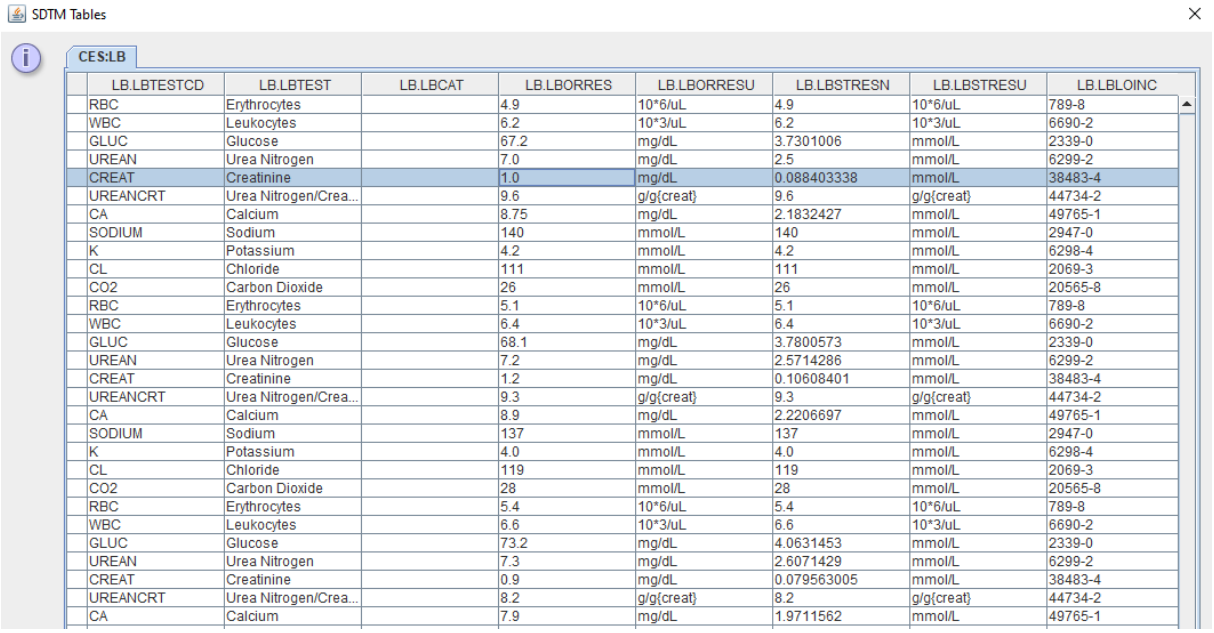
  rws:unitconversionMW('2.5', 'mg/dL', 'mmol/L', '180.156')

  Of course also here, you will not hardcode the strings, but use variables.

The SDTM-ETL distribution comes with an example file where these functions are demonstrated. You can find it in the distribution under
/TestFiles/define_2_0_mappings/CES_LB_define_loinc2cdisc_testing.xml

When this mapping define.xml is executed on the example file with clinical data located at:
/TestFiles/ODM-1-3-1/CES_ClinicalData_simple_subject_LOINC_Coded.xml

The result looks like the following:



	LB.LBTESTCD	LB.LBTEST	LB.LBCAT	LB.LBORRES	LB.LBORRESU	LB.LBSTRESN	LB.LBSTRESU	LB.LBLOINC
	RBC	Erythrocytes		4.9	10*6/uL	4.9	10*6/uL	789-8
	WBC	Leukocytes		6.2	10*3/uL	6.2	10*3/uL	6690-2
	GLUC	Glucose		67.2	mg/dL	3.7301006	mmol/L	2339-0
	UREAN	Urea Nitrogen		7.0	mg/dL	2.5	mmol/L	6299-2
	CREAT	Creatinine		1.0	mg/dL	0.088403338	mmol/L	38483-4
	UREANCRT	Urea Nitrogen/Crea...		9.6	g/g{creat}	9.6	g/g{creat}	44734-2
	CA	Calcium		8.75	mg/dL	2.1832427	mmol/L	49765-1
	SODIUM	Sodium		140	mmol/L	140	mmol/L	2947-0
	K	Potassium		4.2	mmol/L	4.2	mmol/L	6298-4
	CL	Chloride		111	mmol/L	111	mmol/L	2069-3
	CO2	Carbon Dioxide		26	mmol/L	26	mmol/L	20565-8
	RBC	Erythrocytes		5.1	10*6/uL	5.1	10*6/uL	789-8
	WBC	Leukocytes		6.4	10*3/uL	6.4	10*3/uL	6690-2
	GLUC	Glucose		68.1	mg/dL	3.7800573	mmol/L	2339-0
	UREAN	Urea Nitrogen		7.2	mg/dL	2.5714286	mmol/L	6299-2
	CREAT	Creatinine		1.2	mg/dL	0.10608401	mmol/L	38483-4
	UREANCRT	Urea Nitrogen/Crea...		9.3	g/g{creat}	9.3	g/g{creat}	44734-2
	CA	Calcium		8.9	mg/dL	2.2206697	mmol/L	49765-1
	SODIUM	Sodium		137	mmol/L	137	mmol/L	2947-0
	K	Potassium		4.0	mmol/L	4.0	mmol/L	6298-4
	CL	Chloride		119	mmol/L	119	mmol/L	2069-3
	CO2	Carbon Dioxide		28	mmol/L	28	mmol/L	20565-8
	RBC	Erythrocytes		5.4	10*6/uL	5.4	10*6/uL	789-8
	WBC	Leukocytes		6.6	10*3/uL	6.6	10*3/uL	6690-2
	GLUC	Glucose		73.2	mg/dL	4.0631453	mmol/L	2339-0
	UREAN	Urea Nitrogen		7.3	mg/dL	2.6071429	mmol/L	6299-2
	CREAT	Creatinine		0.9	mg/dL	0.079563005	mmol/L	38483-4
	UREANCRT	Urea Nitrogen/Crea...		8.2	g/g{creat}	8.2	g/g{creat}	44734-2
	CA	Calcium		7.9	mg/dL	1.9711562	mmol/L	49765-1

Where one can see that all tests that have "conventional" units in LBORRES/LBORRESU were converted (using the function and thus using the RESTful web service) to SI units, i.e. from mg/dL to mmol/L.

This demo mapping file also uses another set of functions, i.e. to populate LBTESTCD, LBTEST, LBSPEC, etc. starting from the LOINC code, using a RESTful web service that implements the by CDISC published LOINC-SDTM-LB mapping. These functions are further explained in the manual "Using the LOINC-SDTM-LB mapping and similar functions". Unnecessary to say that, when the LOINC code is available, these functions make the mapping process considerably easier, as the population of several variables can be automated.

Remark: NONE of the RESTful web services that are implemented ever send patient-identifying information to any server. The above-mentioned functions only send quantity, source and target units, and LOINC code or molecular weight to the server. No demographic data is ever send.