SDTM-ETL 4.6 User Manual and Tutorial

Author: Jozef Aerts, XML4Pharma

Last update: 2024-08-05

Caching Unit Conversion Information

Table of Contents

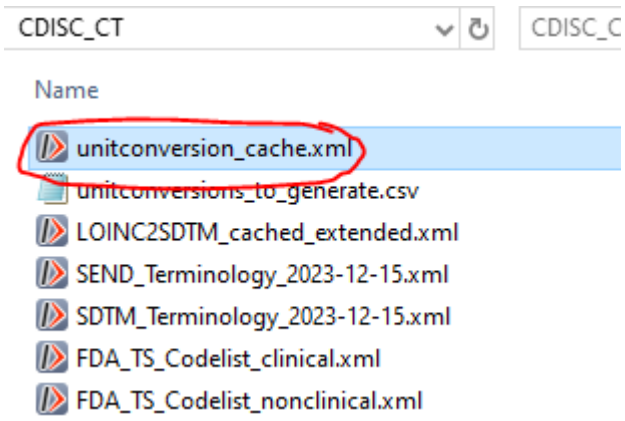
Introduction

One of the great features of the SDTM-ETL software is to automatically perform unit conversions using the RESTful Web Services (RWS) of the US National Library of Medicine NLM (https://ucum.nlm.nih.gov/ucum-service.html). These features are described in detail in the separate manual "Performing Unit Conversions in SDTM-ETL".
This automation of unit conversions is one of the most appreciated features of our mapping software.

Although the NLM RWS is extremely fast, when many thousands of conversions (e.g. between US conventional (mass concentration) units and "SI"[1] (molar concentration) units) need to be performed, this can take considerable time. Also, some of the users of the software prefer to be able to do this conversions without needing an internet connection that connects to the NLM RWS.

Therefore, as of version 4.6, it is possible to "cache" conversion factors. When using the unit conversion functions of SDTM-ETL, the software will first look whether the conversion factor between source and target unit is available in the "cache file", which is named "unitconversion_cache.xml" and resides in the "CDISC-CT" folder of the software distribution:

[1] Remark that essentially the designation "SI units" for "molar concentration units" is completely false, unfortunately introduced by the publications of the FDA.



The format of this file is:

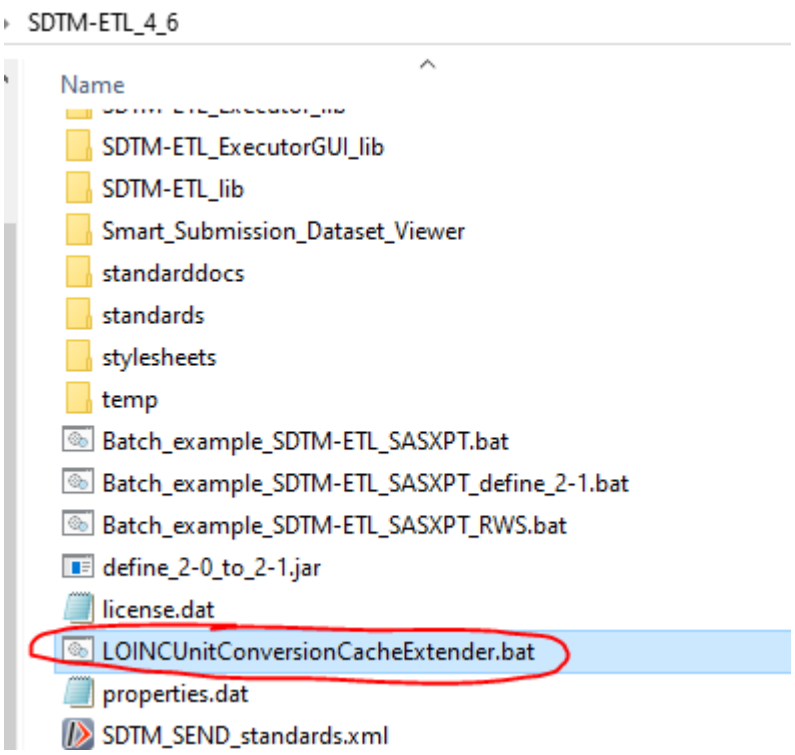
```
<UnitConversionCache>
    <UCUMWebServiceResponse ServerDateTime="2024-06-05T06:12:52">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1/from/mg/dL/to/umol/L/LOINC/1975-2</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>mg/dL</SourceUnit>
            <TargetUnit>umol/L</TargetUnit>
            <MolecularWeightUsed>584.6621</MolecularWeightUsed>
            <ResultQuantity>17.103896</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <!-- Conversion factors added 2024-06-10T07:44:25.985Z -->
    <UCUMWebServiceResponse ServerDateTime="2024-06-10T07:44:28">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/g/dL/to/umol/L/LOINC/1751-7</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>g/dL</SourceUnit>
            <TargetUnit>umol/L</TargetUnit>
            <MolecularWeightUsed>2754.105</MolecularWeightUsed>
            <ResultQuantity>3630.9436</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <UCUMWebServiceResponse ServerDateTime="2024-06-10T07:44:29">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/mg/dL/to/mmol/L/LOINC/2339-0</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>mg/dL</SourceUnit>
            <TargetUnit>mmol/L</TargetUnit>
            <MolecularWeightUsed>180.156</MolecularWeightUsed>
            <ResultQuantity>0.055507449</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <UCUMWebServiceResponse ServerDateTime="2024-06-10T07:44:31">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/mg/dL/to/umol/L/LOINC/14631-6</WebServiceRequest>
        <Response>
```

i.e. it is an XML document with the root element name being "UnitConversionCache", and the child elements being the XML responses as obtained from the NLM RWS server.

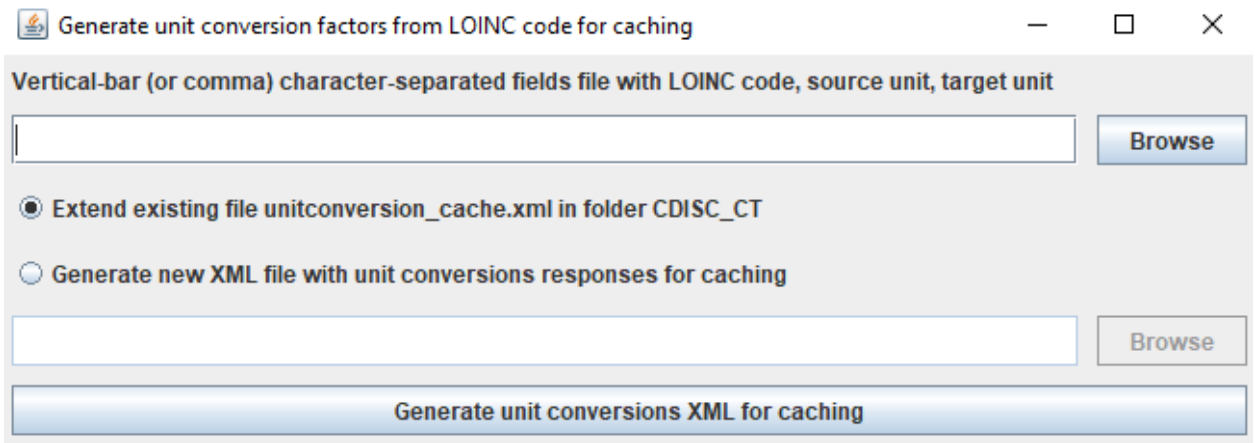
As we do not expect the users to fill this file manually, also a utility program is provided to generate or extend the "unitconversion_cache.xml" file. This program is named "LOINCUnitConversionCacheExtender" and can be started either by line command (CLI command):

java -cp SDTM-ETL.jar
com.xml4pharma.sdtmetl.loinc2sdtm.LOINCUnitConversionCacheExtender

from the main SDTM-ETL folder.
Or by a double click on the file "LOINCUnitConversionCacheExtender.bat" in the same folder:



In both cases, the utility application will start and show a graphical user interface:



In the first field, one should provide the path to a CSV file containing information of the unit conversions for which the conversion factor needs to be stored.
An example file "unitconversions_to_generate.csv" is provided in the folder "CDISC_CT":

*unitconversions_to_generate.csv - Editor

Datei Bearbeiten Format Ansicht Hilfe

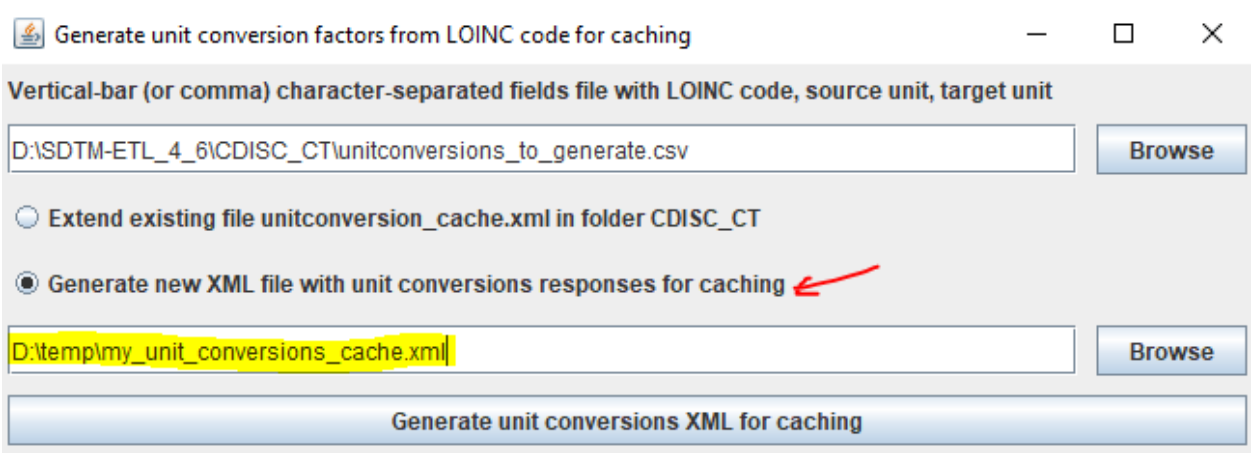
```
# List of LOINC codes, source unit, target unit
# for generating the conversion factors using the NLM RESTful web service
# See https://ucum.nlm.nih.gov/ucum-service.html for details
1751-7|g/dL|umol/L  # Albumin
2339-0|mg/dL|mmol/L # Glucose
12345 | g/l | mol/L # Invalid LOINC code
2339-0|g/dL|cm      # properties do not match
# test data from Princeton Pharmatech
14631-6|mg/dL|umol/L
14647-2|mg/dL|mmol/L
14682-9|mg/dL|umol/L
14749-6|mg/dL|mmol/L
14879-1|mg/dL|mmol/L
14927-8|mg/dL|mmol/L
17861-6|mg/dL|mmol/L
19123-9|mg/dL|mmol/L
20570-8|%|%   # for "%" testing
2157-6|[IU]/L|[IU]/L  # for [ and ] testing
2885-2|g/dL|mmol/L
718-7|g/dL|mmol/L
```

Lines starting with the "#" character are comment lines, and will not be treated by the software. One can also add a comment on lines with the information (as e.g. in lines 4-7) at the end of the line, explaining what the line is about.

Each line with non-comment information contains 3 pieces of information:
- the LOINC code of the test, defining the analyte. This is necessary as some unit conversions need the molecular weight of the analyte, which is obtained from the LOINC code.
Example: 1751-7 for "Albumin [Mass/volume] in Serum or Plasma"
- the source unit (UCUM-notation), e.g. "g/dL"
- the target unit (UCUM-notation), e.g. "umol/L"

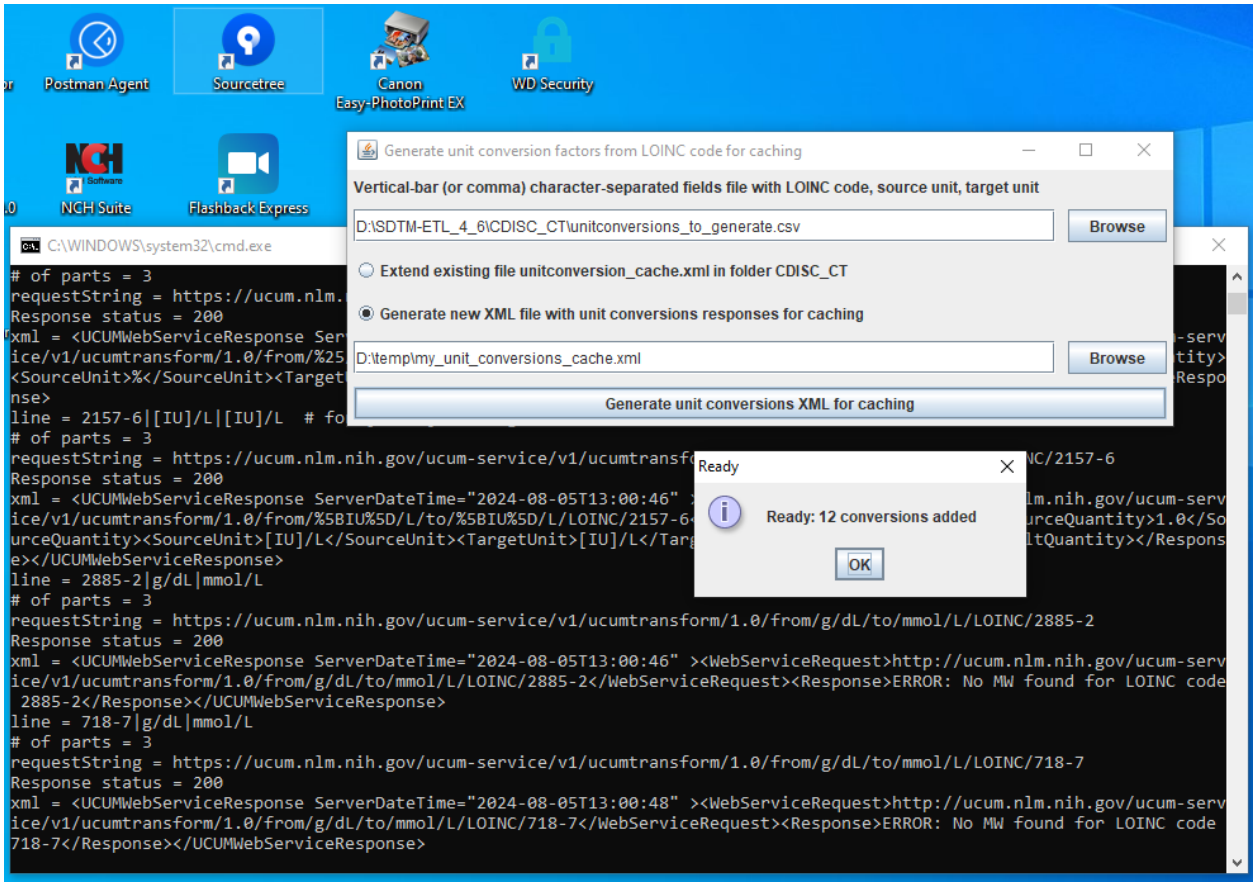
Remark that this file also contains some test entries, e.g. for non-matching properties (like line 7 with "g/dL" and "cm". Such lines will then not lead to a stored conversion factor.

In the Graphical User Interface:

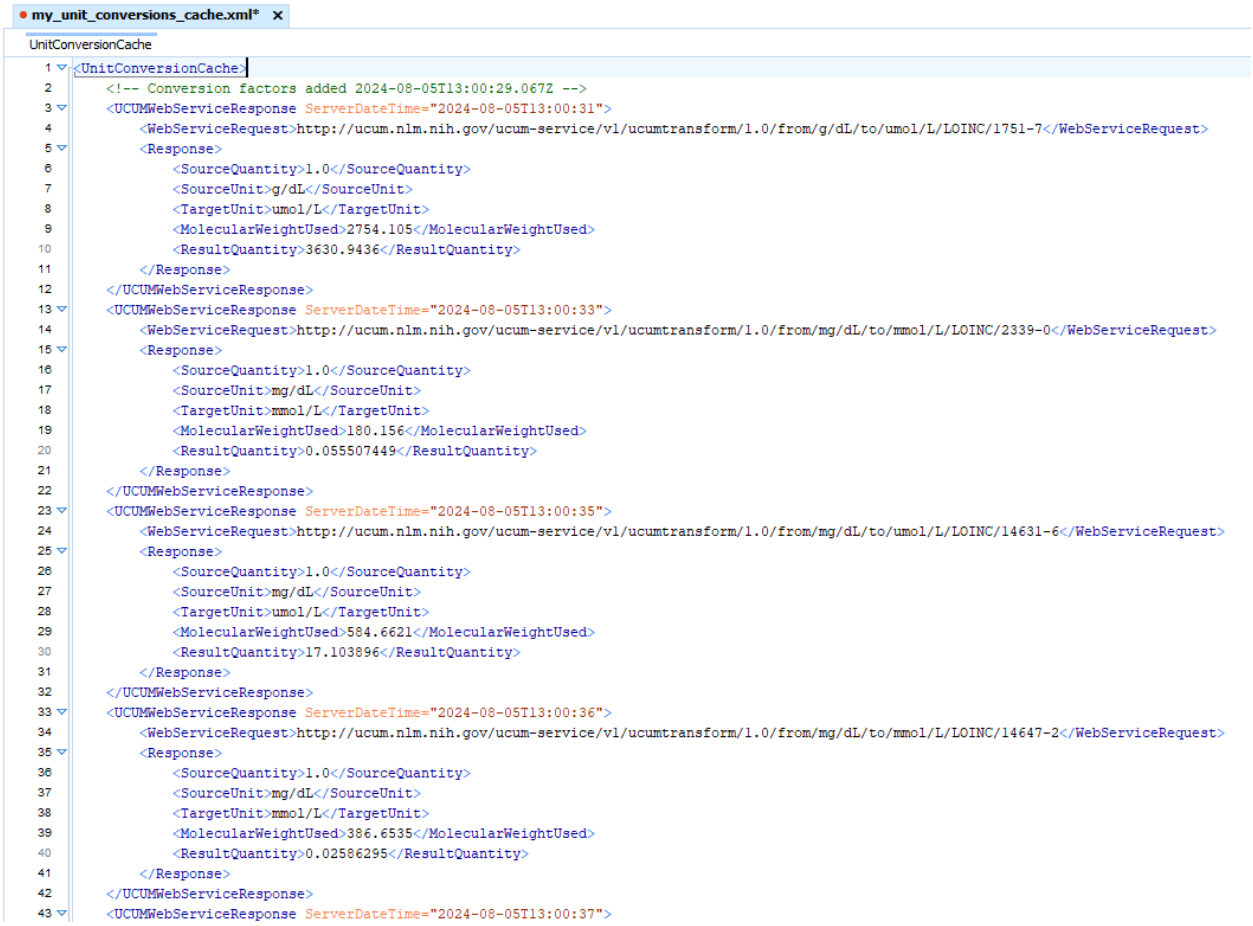


One can choose between either extending the existing "unitconversion_cache.xml" in the CDISC_CT folder, and generating a new one as "my_unit_conversions_cache.xml" in the folder "D:\temp".
Later, one can of course edit and/or merge such XML files manually, e.g. using NotePad++, or any simple XML editor.

When then clicking the button "Generate unit conversions XML for caching", the software starts and one can follow the progress in the console:



The generated XML in the file "my_unit_conversions_cache.xml" then has the following contents:



my_unit_conversions_cache.xml*

UnitConversionCache

```
<UnitConversionCache>
    <!-- Conversion factors added 2024-08-05T13:00:29.067Z -->
    <UCUMWebServiceResponse ServerDateTime="2024-08-05T13:00:31">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/g/dL/to/umol/L/LOINC/1751-7</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>g/dL</SourceUnit>
            <TargetUnit>umol/L</TargetUnit>
            <MolecularWeightUsed>2754.105</MolecularWeightUsed>
            <ResultQuantity>3630.9436</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <UCUMWebServiceResponse ServerDateTime="2024-08-05T13:00:33">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/mg/dL/to/mmol/L/LOINC/2339-0</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>mg/dL</SourceUnit>
            <TargetUnit>mmol/L</TargetUnit>
            <MolecularWeightUsed>180.156</MolecularWeightUsed>
            <ResultQuantity>0.055507449</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <UCUMWebServiceResponse ServerDateTime="2024-08-05T13:00:35">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/mg/dL/to/umol/L/LOINC/14631-6</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>mg/dL</SourceUnit>
            <TargetUnit>umol/L</TargetUnit>
            <MolecularWeightUsed>584.6621</MolecularWeightUsed>
            <ResultQuantity>17.103896</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <UCUMWebServiceResponse ServerDateTime="2024-08-05T13:00:36">
        <WebServiceRequest>http://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/1.0/from/mg/dL/to/mmol/L/LOINC/14647-2</WebServiceRequest>
        <Response>
            <SourceQuantity>1.0</SourceQuantity>
            <SourceUnit>mg/dL</SourceUnit>
            <TargetUnit>mmol/L</TargetUnit>
            <MolecularWeightUsed>386.6535</MolecularWeightUsed>
            <ResultQuantity>0.02586295</ResultQuantity>
        </Response>
    </UCUMWebServiceResponse>
    <UCUMWebServiceResponse ServerDateTime="2024-08-05T13:00:37">
```

If one needs more than one such files with unit conversions, it is recommended to create a small repository to manage them. Important is that the software will always use the current file "unitconversion_cache.xml" in the SDTM-ETL folder "CDISC_CT".

When then running SDTM-ETL, the NLM RESTful Web Service will not be called for the entries in this file, but the conversion factor will be taken from the file instead.
This can save considerable time when many thousands of conversions need to be executed.

It also allows users to change the conversion factor when disagreeing with it, but this should of course be the absolute exception.

International Units

In some cases. one has obtained concentrations in international units, and wants to convert them into "SI" units (i.e. molar concentration units). LOINC however does not always provide a conversion factor for these, also as in some cases, there is no international agreement what the conversion factor must be. IF one has a conversion factor for such, it is recommended to use that in the mapping code itself using an "if-elsif-else" structure. One should not try to incorporate such conversions in the "cache file".

Unit conversion execution

When the file "unitconversion_cache.xml" is present in the SDTM-ETL folder "CDISC_CT", and one uses the function rws:unitconversionLoinc(string quantity, string source-unit, string target-unit, String loinccode), then the system will always try to use the "cache" first. Only when the conversion requested is not in the "cache", then the RESTful Web Service will be called to make an attempt to do the conversion.
This is also visible from the messages generated in the "message" text area during the execution, e.g.:

```
Trying to find value for variable LBTESTCD for LOINC code 1751-7
Found LBTESTCD = ALB in XML file with cached mappings for LOINC code 1751-7
Trying to find value for variable LBTEST for LOINC code 1751-7
Found LBTEST = Albumin in XML file with cached mappings for LOINC code 1751-7
suggested SI unit for LOINC code 1751-7 = umol/L
For unit conversion: treating LOINC code = 1751-7
Using cache for unit conversion for LOINC code = 1751-7 with source unit = g/dL and target unit = umol/L
Conversion factor found from cache = 3630.9436
Result quantity = 14160.68004
Trying to find value for variable LBMETHOD for LOINC code 1751-7
Found LBMETHOD =  in XML file with cached mappings for LOINC code 1751-7
Trying to find value for variable LBTESTCD for LOINC code 2093-3
Using RESTful web service to find LBTESTCD value for LOINC code 2093-3
Value for LBTESTCD for LOINC code 2093-3 = CHOL
Trying to find value for variable LBTEST for LOINC code 2093-3
Using RESTful web service to find LBTEST value for LOINC code 2093-3
Value for LBTEST for LOINC code 2093-3 = Cholesterol
suggested SI unit for LOINC code 2093-3 = mmol/L
For unit conversion: treating LOINC code = 2093-3
Performing unit conversion using the NLM RESTful web service
NLM RESTful-WS = https://ucum.nlm.nih.gov/ucum-service/v1/ucumtransform/254/from/mg/dL/to/mmol/L/LOINC/2093-3
Trying to find value for variable LBSPEC for LOINC code 2093-3
Using RESTful web service to find LBSPEC value for LOINC code 2093-3
Value for LBSPEC for LOINC code 2093-3 = SERUM OR PLASMA
```

Conclusions

The automated unit conversion using the NLM RESTful Web Services allows to automate such conversions, also between "US conventional" and "SI" units and the other way around, and this without needing to "hardcode" the conversion factors. This surely reduces human errors when developing the mappings and saves an enormous amount of time. SDTM-ETL v.4.6, one can generate and manage a "cache file" with conversion. This "cache" file is generated by the new program "LOINCUnitConversionCacheExtender" that comes with the software distribution.
When executing unit conversions based on the LOINC code, the system will first try to use the conversion factors from the "cache file". Only when it does not find an entry for the requested conversion, it will use the (somewhat slower) NLM RESTful Web Service to execute the conversion.
This new feature can save a lot of time when many thousands of conversions need to be executed.