SDTM-ETL 4.6 User Manual and Tutorial

Author: Jozef Aerts, XML4Pharma

Last update: 2024-08-05

Caching LOINC to SDTM Mapping Information

Table of Contents

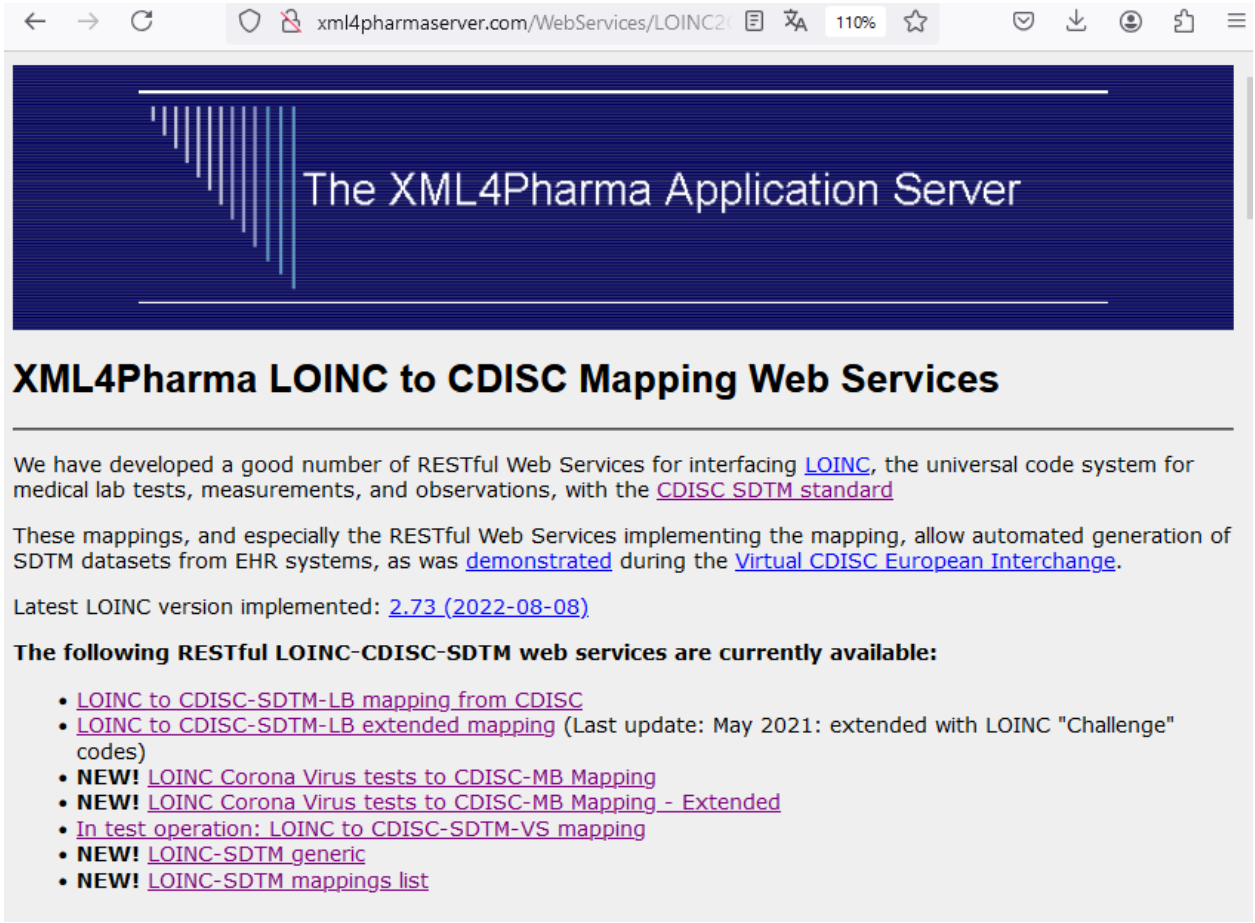
Introduction

CDISC has for a long time tried to ignored LOINC as the really unique identifier for lab and microbiology tests, as well as vital signs and ECG tests, and even questionnaires.
This only changed a bit when the FDA started requiring the LOINC code for LB datasets.

Instead, CDISC developed controlled terminology based on the "postcoordinated" principle, where one uses different variables to describe different aspects of the tests. So, SDTM uses a combination of --TESTCD (and --TEST), --METHOD, --SPEC, etc. to describe tests. LOINC however is "precoordinated", meaning that a single LOINC code defines a single, unique test. As such, LOINC is a superior system for describing lab and other tests, and is used for this in medicine and medical informatics all over the world[1].
When the FDA started mandating LOINC for lab tests, CDISC reacted by developing mappings between LOINC codes and SDTM variables for the 1,400 most used LOINC codes (almost 2,300 mappings). These mappings were later extended to over 9,500 LOINC codes (over 18,000 mappings) and a RESTful web service (RWS) developed which is described at:

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

[1] In many countries, LOINC is even mandated by law to be used in health records for labs, microbiology, vitals signs etc..



SDTM-ETL has a number of functions that use these RWS, allowing to automatically generate values for variables such as --TESTCD, --TEST, --SPEC, --METHOD etc.. These functions are described in the document "Using the LOINC-SDTM-LB Mapping and similar functions" at http://www.xml4pharma.com/SDTM-ETL/tutorials/SDTM-ETL_LOINC-SDTM-Mapping.pdf

Using these functions, generating SDTM datasets, especially for LB (Laboratory) and MB (Microbiology), becomes very easy when the LOINC codes are provided by the labs, and avoids making incorrect mapping decisions. Some customers even reported that it saved them many days of work, with a superior quality outcome.

Improvements in SDTM-ETL v.4.6: Caching

The LOINC-SDTM mapping functions use the RESTful web service as described above. Although this RWS is very fast, when many thousand of mappings between LOINC codes and SDTM variables need to be executed, this may take considerable time. Even in the case of a few LOINC codes only, the process needs to be repeated for each row in the SDTM dataset, which is not very efficient for large SDTM-LB and SDTM-MB datasets.
Also, some customers wanted to be able to avoid using the RWS, as it requires an internet connection.
Therefore, it was decided to provide a "caching" method, by using a local XML file that contains the user's mappings between LOINC and SDTM, which resides in the "CDISC-CT" folder and is named "LOINC2SDTM_cached.xml". The distribution already comes with such a file, which can be replaced, recreated or extended by the user.

The contents of the "LOINC2SDTM_cached.xml" file are:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<LOINC2SDTM>
    <LOINC2SDTMMapping MappingSource="CDISC" TargetSDTMDomain="LB">
        <LBTESTCD NCICode="C64431">ALB</LBTESTCD>
        <LBTEST NCICode="C64431">Albumin</LBTEST>
        <LBORRESU_Example NCICode="C42576">g/L</LBORRESU_Example>
        <LBPOS/>
        <LBLOINC>1751-7</LBLOINC>
        <LBSPEC NCICode="C105706">SERUM OR PLASMA</LBSPEC>
        <LBLOC/>
        <LBMETHOD/>
        <LBANMETH/>
        <LBFAST/>
        <LBTPT/>
        <SUPPLB.LBPTFL>Y</SUPPLB.LBPTFL>
        <SUPPLB.LBPDUR/>
        <SUPPLB.LBRESTYP>MASS CONCENTRATION</SUPPLB.LBRESTYP>
        <SUPPLB.LBRSLSCL>QUANTITATIVE</SUPPLB.LBRSLSCL>
        <SUPPLB.LBTSTOPO/>
        <SUPPLB.LBLLOD/>
        <SUPPLB.LBTSTCND/>
        <SUPPLB.LBMTHSEN/>
        <Example_UCUM_Units>g/dL</Example_UCUM_Units>
    </LOINC2SDTMMapping>
    <LOINC2SDTMMapping MappingSource="CDISC" TargetSDTMDomain="LB">
        <LBTESTCD NCICode="C51946">RBC</LBTESTCD>
        <LBTEST NCICode="C51946">Erythrocytes</LBTEST>
        <LBORRESU_Example NCICode="C67308">10^12/L</LBORRESU_Example>
        <LBPOS/>
        <LBLOINC>789-8</LBLOINC>
        <LBSPEC NCICode="C12434">BLOOD</LBSPEC>
        <LBLOC/>
        <LBMETHOD NCICode="C154794">AUTOMATED COUNT</LBMETHOD>
        <LBANMETH/>
        <LBFAST/>
        <LBTPT/>
        <SUPPLB.LBPTFL>Y</SUPPLB.LBPTFL>
        <SUPPLB.LBPDUR/>
        <SUPPLB.LBRESTYP>NUMBER CONCENTRATION</SUPPLB.LBRESTYP>
        <SUPPLB.LBRSLSCL>QUANTITATIVE</SUPPLB.LBRSLSCL>
        <SUPPLB.LBTSTOPO/>
        <SUPPLB.LBLLOD/>
        <SUPPLB.LBTSTCND/>
        <SUPPLB.LBMTHSEN/>
        <Example_UCUM_Units>10*6/uL</Example_UCUM_Units>
    </LOINC2SDTMMapping>
</LOINC2SDTM>
```

It contains the LOINC-SDTM mappings for two popular LOINC codes 1751-7 and 789-8.

When this file is present in the "CDISC-CT" folder, the software will always first use to retrieve the mapping from this file instead of using the RESTful Web Service when using one of the LOINC-SDTM mapping functions. If the LOINC code is not found in the file, then the (slower) RESTful Web Service is called.
This is also visible in the messages that are generated during the mapping execution. For example:

```
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
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
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
Trying to find value for variable LBMETHOD for LOINC code 2093-3
Using RESTful web service to find LBMETHOD value for LOINC code 2093-3
Trying to find value for variable LBTESTCD for LOINC code 1975-2
Using RESTful web service to find LBTESTCD value for LOINC code 1975-2
Value for LBTESTCD for LOINC code 1975-2 = BILI
Trying to find value for variable LBTEST for LOINC code 1975-2
Using RESTful web service to find LBTEST value for LOINC code 1975-2
Value for LBTEST for LOINC code 1975-2 = Bilirubin
```

As LOINC code 1751-7 is in the "cache" file "LOINC2SDTM_cached.xml", the values for LBTESTCD, LBTEST and LBMETHOD are directly retrieved from it, whereas for LOINC code 2093-3, there is no entry in the "cache" file, so the RESTful Web Service is called to obtain the values for LBTESTCD, LBTEST, LBSPEC, and LBMETHOD.

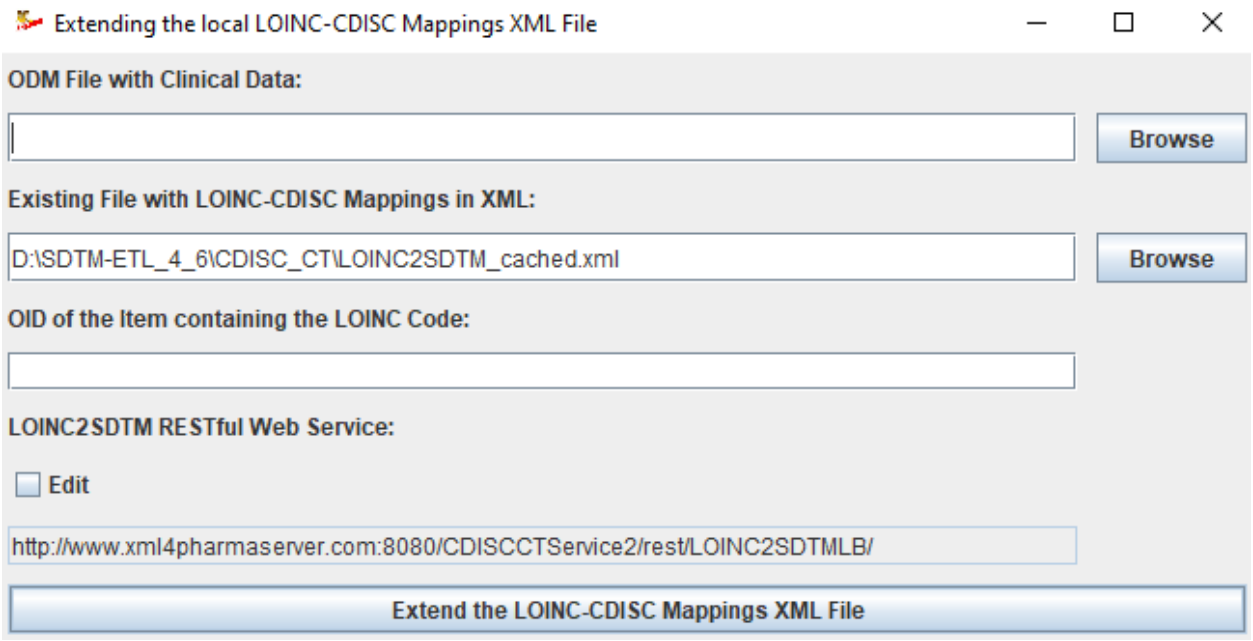
Obtaining mappings for storage in the "cache" file

In order to allow the users of the software to generate their own "LOINC2SDTM_cached.xml"file, a utility program "LOINC2SDTMCacheFileExtender" is provided. It can be started using the line command (CLI)

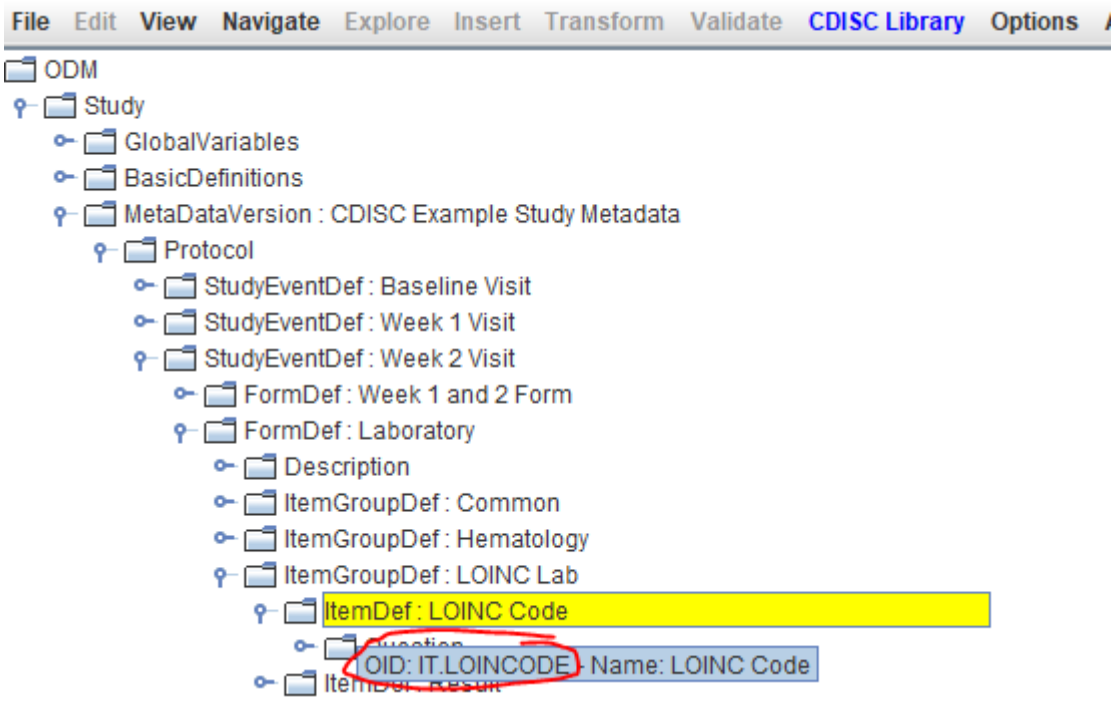
java -cp SDTM-ETL.jar com.xml4pharma.sdtmetl.loinc2sdtm.LOINC2SDTMCacheFileExtender

or by a double-click on the file "LOINC2SDTMCacheFileExtender.bat".

When started, a graphical user interface is provided:

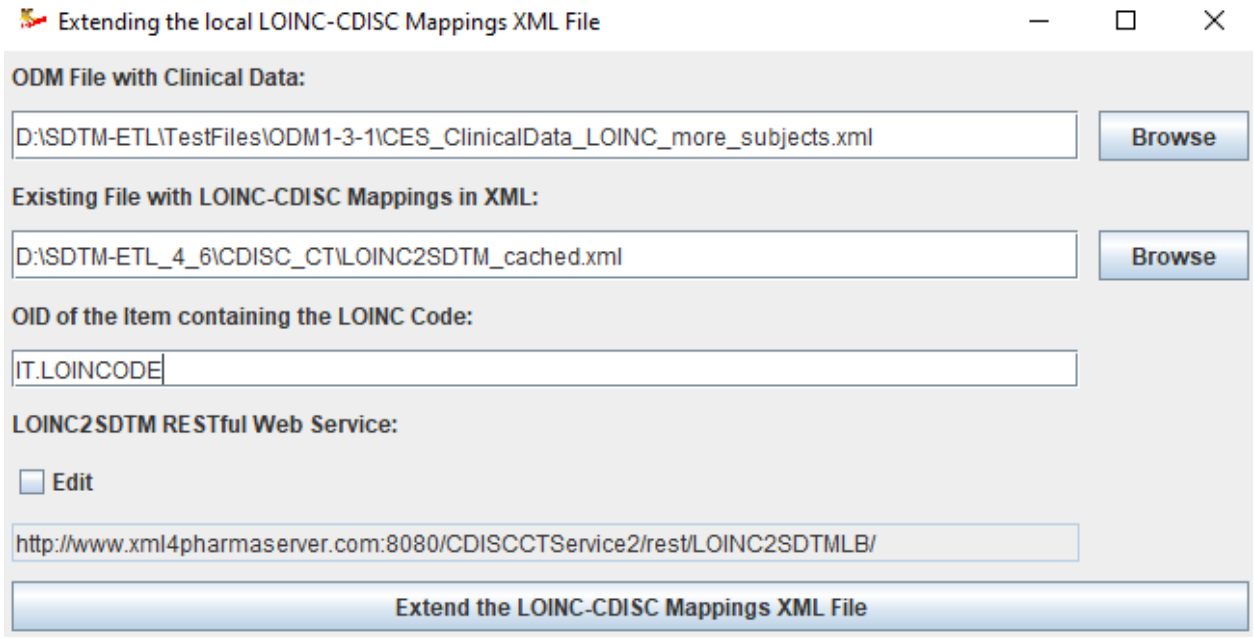


It asks for an ODM file with clinical data that also contains LOINC codes, and the OID (Object Identifier) of the data point containing the LOINC code. The latter can e.g. be found by inspecting the ODM tree in the SDTM-ETL application itself. For example:

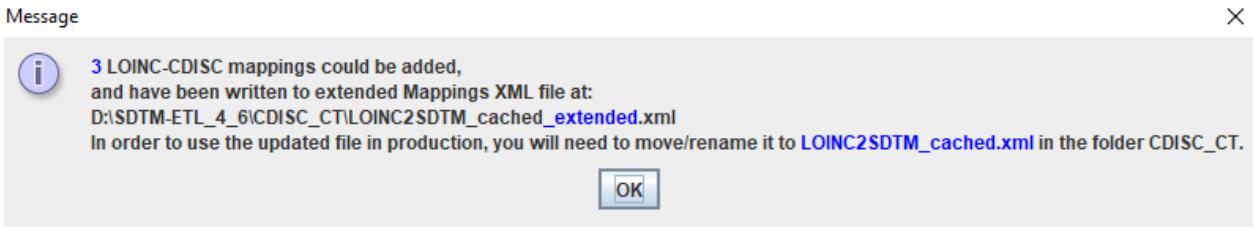


where one finds that the OID for the data point containing the LOINC code is "IT.LOINCCODE".

We can then provide this information to the GUI of the LOINC2SDTMCacheFileExtender application:



and clicking the button "Extend the LOINC-CDISC Mappings XML file" starts the generation, with a final message like:



and when inspecting the generated file "LOINC2SDTM_cached_extended.xml", finding:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<LOINC2SDTM>
   <LOINC2SDTMMapping MappingSource="CDISC" TargetSDTMDomain="LB"> [21 lines]
   <LOINC2SDTMMapping MappingSource="CDISC" TargetSDTMDomain="LB"> [21 lines]
   <LOINC2SDTMMapping MappingSource="CDISC" TargetSDTMDomain="LB">
      <LBTESTCD NCICode="C105586">CHOL</LBTESTCD>
      <LBTEST NCICode="C105586">Cholesterol</LBTEST>
      <LBORRESU_Example NCICode="C67015">mg/dL</LBORRESU_Example>
      <LBPOS/>
      <LBLOINC>2093-3</LBLOINC>
      <LBSPEC NCICode="C105706">SERUM OR PLASMA</LBSPEC>
      <LBLOC/>
      <LBMETHOD/>
      <LBANMETH/>
      <LBFAST/>
      <LBTPT/>
      <SUPPLB.LBPTFL>Y</SUPPLB.LBPTFL>
      <SUPPLB.LBPDUR/>
      <SUPPLB.LBRESTYP>MASS CONCENTRATION</SUPPLB.LBRESTYP>
      <SUPPLB.LBRSLSCL>QUANTITATIVE</SUPPLB.LBRSLSCL>
      <SUPPLB.LBTSTOPO/>
      <SUPPLB.LBLLOD/>
      <SUPPLB.LBTSTCND/>
      <SUPPLB.LBMTHSEN/>
      <Example_UCUM_Units>mg/dL</Example_UCUM_Units>
   </LOINC2SDTMMapping>
   <LOINC2SDTMMapping MappingSource="CDISC" TargetSDTMDomain="LB">
      <LBTESTCD NCICode="C38037">BILI</LBTESTCD>
      <LBTEST NCICode="C38037">Bilirubin</LBTEST>
      <LBORRESU_Example NCICode="C67015">mg/dL</LBORRESU_Example>
      <LBPOS/>
      <LBLOINC>1975-2</LBLOINC>
      <LBSPEC NCICode="C105706">SERUM OR PLASMA</LBSPEC>
      <LBLOC/>
      <LBMETHOD/>
      <LBANMETH/>
      <LBFAST/>
      <LBTPT/>
      <SUPPLB.LBPTFL>Y</SUPPLB.LBPTFL>
      <SUPPLB.LBPDUR/>
      <SUPPLB.LBRESTYP>MASS CONCENTRATION</SUPPLB.LBRESTYP>
      <SUPPLB.LBRSLSCL>QUANTITATIVE</SUPPLB.LBRSLSCL>
```

containing 3 additional LOINC-SDTM mappings.
When one then renames the file "LOINC2SDTM_cached_extended.xml" to "LOINC2SDTM_cached.xml", these new mappings will be used from the file instead of using the RESTful web service.

Conclusions

The automated generation of values from the LOINC code for variables such as LBTESTCD, LBTEST, LBSPEC and LBMETHOD not allows to save enormous amounts of time when developing mappings for lab tests, but also provides high quality mappings, much higher than when done by human mappers.

This automated generation uses a RESTful web service. Although it is very fast, users can now generate and maintain a "cache" file with mappings. During execution of the mappings, when the

LOINC code is present in the "cache" file, the mapping will be taken from that. If it isn't, the (somewhat) slower RESTful Web Service is called. This can save considerable time in the case a very large amount of lab tests need to be treaded.

A separate program, the "LOINC2SDTMCacheFileExtender" allows to retrieve all distinct LOINC codes from an ODM file with clinical data, and generate or extend the "cache" file with the mappings between the LOINC codes and the SDTM variable values.

Limitations and future developments

This method does not work yet for new SDTM variables for SDTMIG-3.4 that correspond to other "LOINC parts": LBRESSCL (result scale, example: QUANTITATVE), LBRESTYP (result type, examples: NUMBER CONCENTRATION, RATIO), LBCOLSRT (Collected Summary Type, examples: MEAN, MINIMUM, MAXIMUM). These will be added in future. They are important as additional identifiers for the test when relying on "post-coordinated" variables[2].

[2] Essentially, when the LOINC code is provided, it should not be needed to also submit LBTESTCD, LBTEST, LBSPEC, LBMETHOD, LBRESSCL, etc.., as only the LOINC code provides the unique identifier for the test. This is however still not recognized by CDISC, LOINC still be seen as "not invented here".