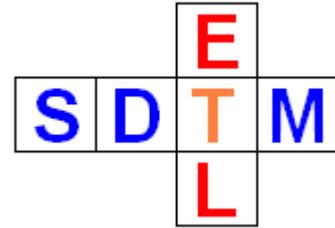


SDTM-ETL 4.0: Working with the CDISC Library

Author: Jozef Aerts, XML4Pharma

Last update: 2020-09-22



Introduction

The [CDISC Library](#) and its [API](#) provide electronic, machine-readable versions of most of the CDISC standards, that can be used directly from within applications. SDTM-ETL is one of the very first applications that uses the CDISC Library and its API.

In order to use the CDISC Library features in SDTM-ETL, you will need a CDISC Library account and key. These can be [obtained directly from CDISC](#), and is free of charge for CDISC members.

Installing the CDISC Library key

Once you have a CDISC Library user account and key, you will need to copy these into the file "properties.dat" that you can find in the main directory where you installed the software. You can edit that file using a simple text editor like "notepad" on Windows or "vi" on Linux.

You will need to set username and passkey as two parameters "cdisclibraryuser" and "cdisclibrarypass". For example:

```
*properties.dat - Editor
Datei Bearbeiten Format Ansicht Hilfe
language=en
languagefixed=true
#logfilepath=C:\temp
loglevel=INFO
sasviewerlocation=C:\Program Files\SAS Institute\SAS System Viewer\Sv.exe
adobereaderlocation=C:\Program Files (x86)\Adobe\Acrobat Reader DC\Reader\AcroRd32.exe
# CDISC Library username and key
cdisclibraryuser=xxxxxxx
cdisclibrarypass=yyyyyyy
```

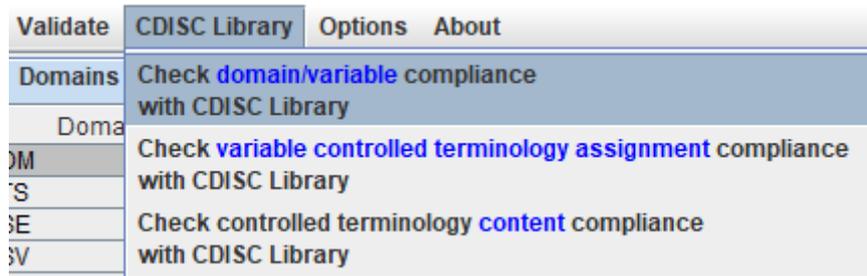
CDISC Library Features in SDTM-ETL

The CDISC Library features are constantly further developed and extended. At this moment, the CDISC Library is not yet used for execution of the mappings. Reason is that these are based on XSLT technology, and although XSLT has support for RESTful web services, it has no mechanism for "basic authentication", as is used by the CDISC Library RESTful web services, yet.

At this moment, the CDISC Library is mainly used for validation of defined dataset/domain definitions and the variables therein, and their associated codelists and codelist contents.

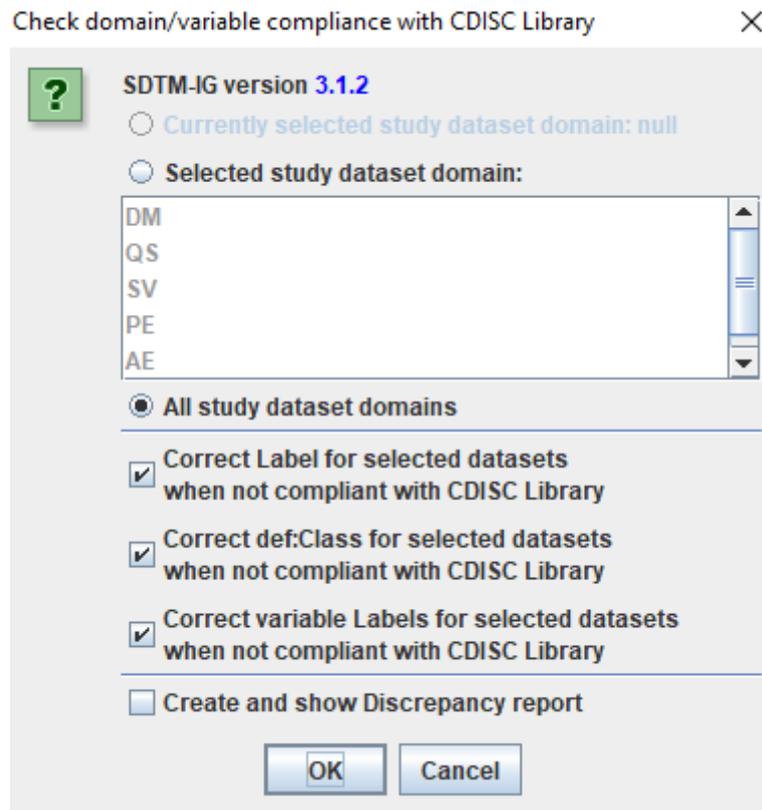
Validation of domain names and labels

In order to validate domain/dataset and variable labels, use the menu "CDISC Library – Check domain/variable compliance with CDISC Library":



In most cases, you will be asked whether you want to validate either:

- the currently selected domain/dataset
- one or more of the currently available "study" domains/datasets
- all the available study-specific domains/datasets



Furthermore, you can allow the software to "autocorrect" labels and class assignments when a discrepancy is found. A "discrepancy report" can be generated. The latter can be saved as an XML file for easy processing with other software and systems.

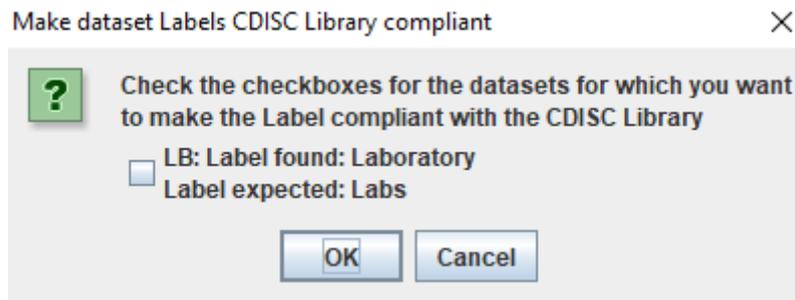
CDISC Library Discrepancies Report

File

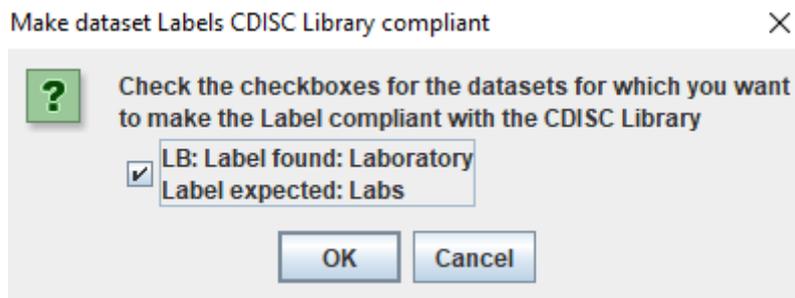
Discrepancies with CDISC Library

Dataset	Domain	Variable	Description	Expected Value
LB	LB		Incorrect Label for dataset = "CES:LB" - found = "Laboratory"	Labs
LB	LB	LBGRPID	Variable Label mismatch - found = Group Identifier	Group ID
LB	LB	LBREFID	Variable Label mismatch - found = Specimen Identifier	Specimen ID
LB	LB	LBORRES	Variable Label mismatch - found = Result or Findings as Collected	Result or Finding in Original Units

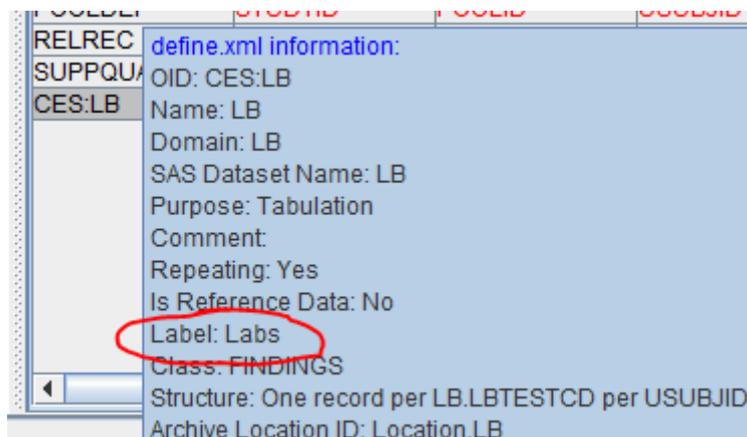
The discrepancy can then be automatically corrected for



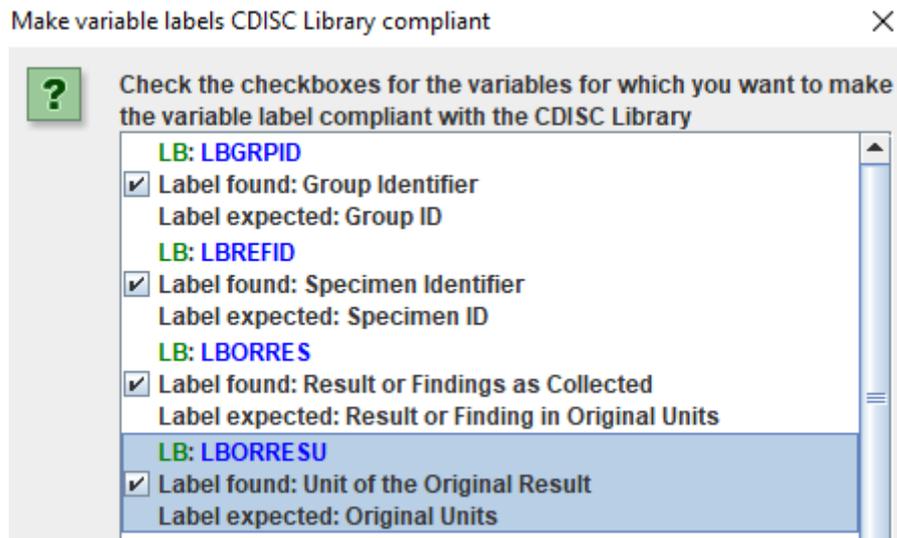
When the checkbox is selected:



As can be seen from the tooltip on the dataset definition cell:



Also labels for variables that were found not to be in accordance with the ones provided by the CDISC Library can be automatically adapted:



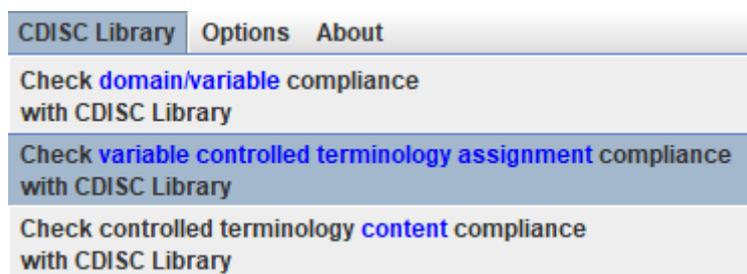
If you started from one of the by the SDTM-ETL software delivered templates, such discrepancies should of course not occur, but you might of course have started from another define.xml, such as from the sponsor, as a specification of what the sponsor expects to be delivered. In such cases, one often finds small deviations, sometimes even of 1 character only, or an uppercase-lowercase difference.

Also remark that in SDTM-IG 3.3, it is **allowed** to deviate from the labels provided in the SDTM-IG.

Validation of controlled terminology

For controlled terminology, there are two menu items, one for checking the correct assignment of codelists to the study dataset variables, and one for checking the content of the codelists assigned to the variables.

Using the menu "CDISC Library – Check variable controlled terminology assignment compliance with CDISC Library",



allows to validate the controlled terminology assigned and defined in the underlying define.xml, through the Codelists assigned and defined. For each variable, it is checked against the CDISC Library whether it is expected that a codelist is assigned, and if so, whether the NCI code of it matches the one expected from the CDISC Library.

Also here, a discrepancy report can be generated, looking like:

Dataset	Domain	Variable	Description	Expected Value
DM	DM	AGEU	Associated CodeList with OID = CL.AGEU was assigned to the variable, but it has a different NCI code = NCI-NONSENSE	C66781
DM	DM	SEX	A CodeList with OID = CL.SEX was assigned to the variable, but it does not have an NCI code	C66731
DM	DM	RACE	A CodeList with OID = CL.RACE was assigned to the variable, but it does not have an NCI code	C74457
DM	DM	ETHNIC	A CodeList with OID = CL.ETHNIC was assigned to the variable, but it does not have an NCI code	C66790
QS	QS	QSORRESU	No associated CodeList was assigned to the variable	C71620
QS	QS	QSSTRESU	No associated CodeList was assigned to the variable	C71620

Some of the items in the "discrepancy report" table are hyperlinked. For example, clicking the "Expected Value" item "C74457" looks up the "Race" codelist in the CDISC Library and displays all its available information:

CDISC Library CodeList Details

CDISC CodeList Details for CodeList with NCI code C74457

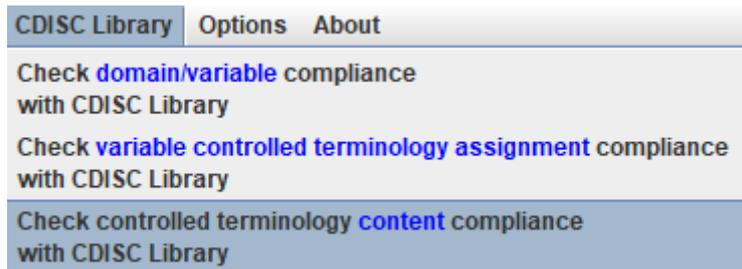
Short Name: RACE
 Long Name: Race
 Extensible: false
 Definition: Terminology codelist used to identify the race of an individual within the Clinical Data Interchange Standards Consortium Study Data Tabulation Model.
 Preferred term: CDISC SDTM Race Terminology
 Synonyms: Race

Codelist values:

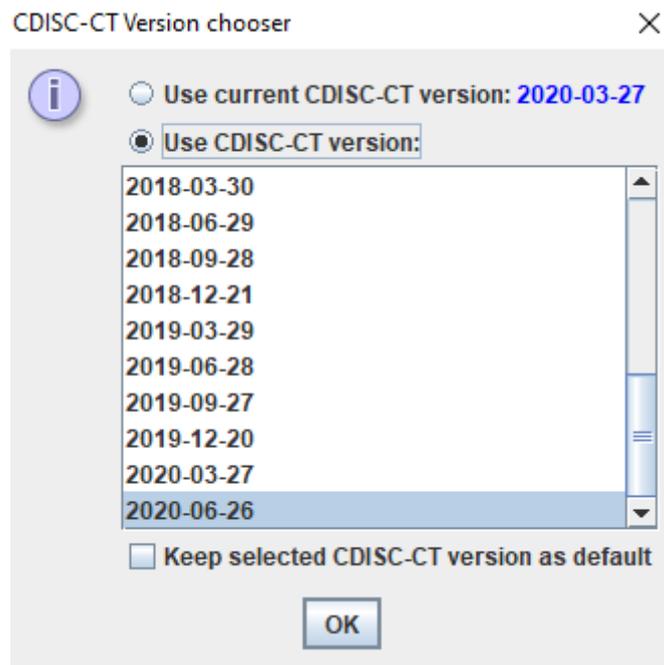
NCI Code	Submission Value	Definition	Preferred Term	Synonyms
C41259	AMERICAN INDIAN OR ALASKA NATIVE	A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. (FDA)	American Indian or Alaska Native	
C41260	ASIAN	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. (FDA)	Asian	
C16352	BLACK OR AFRICAN AMERICAN	A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American." (FDA)	Black or African American	
C41219	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	Denotes a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. The term covers particularly people who identify themselves as part-Hawaiian, Native Hawaiian, Guamanian or Chamorro, Carolinian, Samoan, Chuu. (FDA)	Native Hawaiian or Other Pacific Islander	
C43234	NOT REPORTED	Not provided or available.	Not Reported	Not reported

Seite 4 von

The next menu item "Check controlled terminology content compliance with CDISC Library":



Allows to check the contents of the assigned codelists (at least those that are correctly assigned) against the CDISC Library. First, it is asked against which version of CDISC-CT the codelist contents must be checked (default is the one selected when setting up the mappings for the first time). For example:



It is checked whether, according to the CDISC Library, the codelist is "extensible", and then the values in the codelist are checked against what is expected by the CDISC Library, taking the "extensibility rules" into account. An example report is e.g.:

CDISC Library Discrepancies Report

File

Discrepancies with CDISC Library

Dataset	Domain	Variable	Description	Expected Value
AE		AESER	The value 'N' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C49487'	C49487
AE		AESER	The value 'NA' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C48660'	C48660
AE		AESER	The value 'U' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C17998'	C17998
AE		AESER	The value 'Y' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C49488'	C49488
AE		AESER	The value 'MAYBE' is not allowed in the CDISC non-extensible CodeList version 2019-12-20 with NCI code C66742	
AE		AESCAN	The value 'N' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C49487'	C49487
AE		AESCAN	The value 'NA' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C48660'	C48660

Also here, one can click an hyperlink for further details. For example, clicking "C48660" for value "NA" leads to the detailed information from the CDISC Library:

CDISC Library Discrepancies Report

File

Discrepancies with CDISC Library

Dataset	Domain	Variable	Description	Expected Value
AE		AESER	The value 'N' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C49487'	C49487
AE		AESER	The value 'NA' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C48660'	C48660
AE		AESER	The value 'U' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C17998'	C17998
AE		AESER	The value 'Y' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C49488'	C49488
AE		AESER	The value 'MAYBE' is not allowed in the CDISC non-extensible CodeList version 2019-12-20 with NCI code C66742	
AE		AESCAN	The value 'N' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C49487'	C49487
AE		AESCAN	The value 'NA' is a valid CDISC CodeList value, but no NCI code has been provided and thus does not match the expected CDISC Library NCI code 'C48660'	C48660

CDISC Library CodeList Details

CDISC CodeListItem Details for Term with NCI code C48660

NCI Code:	C48660
Submission Value:	NA
Definition:	Determination of a value is not relevant in the current context. (NCI)
Preferred Term:	Not Applicable
Synonyms:	NA; Not Applicable

Further CDISC Library features will surely be added in future.