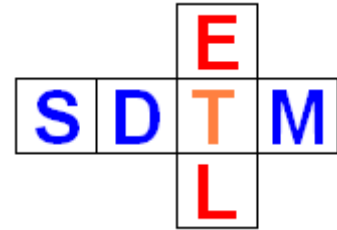


# SDTM-ETL 4.6: Summary of New Features

Author: Jozef Aerts, XML4Pharma

Last update: 2024-08-15



## Summary

This document contains a summary of the most important new features of SDTM-ETL 4.6 and bug fixes.

There are many minor improvements and new features that are not described in this document, but that can be found in other manuals / tutorials of SDTM-ETL 4.6.

## Table of Contents

"Caching" RESTful Web Services results.....	1
Improved algorithms for automated assignment of VISITNUM for unscheduled visits.....	2
Decommission of Pinnacle21 validation.....	2
Extended CLI (batch) execution.....	3
Other minor updates.....	3
Bug fixes.....	3
Limitations.....	4
Further development of SDTM-ETL.....	4

## "Caching" RESTful Web Services results

SDTM-ETL comes with several functions that use RESTful Web Services. These can be divided into 2 categories:

- RESTful Web Services for automated assignment of SDTM/SEND variable values for LBTESTCD, LBTEST, LBSPEC, and LBMETHOD from the LOINC code
- RESTful Web Services for automated unit conversions, also between "US conventional" units and "SI" units<sup>1</sup> and vice versa. This functions use RESTful Web Services made available by the [National Library of Medicine](#) NLM.

These functions are extremely useful as they allow to automate a lot of things in a more reliable way than just using "hard-coded" mappings, and are described in the documents "[Using the LOINC-CDISC mapping for LB datasets](#)" and "[Performing Unit Conversions in SDTM-ETL](#)".

Although these RESTful Web Services are very fast, some users want it even faster, and asked us to be able to "cache" the queries and results. Another reason is that some users may still want to be able to use these functions even when no internet connection is available<sup>2</sup>.

As of SDTM-ETL version 4.6, it is now possible to cache mappings between LOINC codes

---

<sup>1</sup> Remark that the designation "SI units" is essentially wrong, but have become popular by the publications of the FDA. Essentially one should speak about "molar concentration units" or "substance concentration" (like mol/dL), whereas "US conventional units" is essentially about "mass concentration" (like g/dL).

<sup>2</sup> Remark that none of these functions ever sends subject data over the internet.

and LB variable values in a separate file, and also to "cache" unit conversion factors in a separate "cache" file. In both cases, the system will always try to first use the information in the "cache" file, and only use the RESTful web service when the information cannot be obtained from the "cache" file.

Version 4.6 also comes with a number of tools to generate these "cache" files.

More details about "caching" can be found in the separate tutorials "Caching LOINC to SDTM Information" and "Caching Unit Conversion Information".

## Improved algorithms for automated assignment of VISITNUM for unscheduled visits

SDTM-ETL v.4.6 introduced the feature to automatically generate VISITNUM for unscheduled visits. As required by FDA, such visits are to be assigned a fractional number (e.g. 3.1, 3.2) depending on between which visits they occur (e.g. between visit 3 and 4)<sup>3</sup>. We now further improved the algorithm.

The old algorithm was not very accurate when for specific domains/datasets, some data was not collected in specific visits. For example, if lab data was collected in the regular visit 3, but not in the regular visit 4, then the algorithm would still assign VISITNUM=3.1 when lab data was collected in an "unscheduled" visit after visit 4 (for which no lab data was collected). The algorithm was now improved to also take into account the visit dates in dataset SV (subject visits) by comparing the visit dates in SV with the collection dates for the unscheduled visits. This however requires that also SV is loaded, and the mappings are defined before the mappings for the "findings" (and other non-SV) datasets. It further requires that for SVSTDTC (visit start date) is always being populated, and that also --DTC (collection date/time) is being mapped in the dataset definition for the other, non-SV, datasets. This is further explained in the document "[Handling unscheduled visits](#)".

## Decommission of Pinnacle21 validation

We decided to completely decommission Pinnacle21 validation in SDTM-ETL v.4.6. More and more, Pinnacle21 Community is known for its declining quality, with very many "false" positives, and its very bad user-friendliness. Until recently, there however were no real alternatives.

Now that CDISC has developed CORE validation (CDISC Open Rules Engine) and that also FDA has committed to CDISC CORE (other regulatory authorities are expected to follow soon), there is no reason anymore to further support Pinnacle21 validation.

Another reason is that Pinnacle21 validation, that started as an "open source" project, is now essentially "closed source", and that the new owner of Pinnacle21 (Certara) is more and more restricting the use of the software by CROs and sponsors.

Therefore, everything that has to do with Pinnacle21 validation has now been removed from the software, and replaced by CDISC CORE validation. The latter has many advantages such

---

<sup>3</sup> Essentially, this should not be necessary. It looks however that FDA reviewers are unable to sort information by (collection) date/time, so they require us to provide this additional information.

as less "false positives", clear messages, and validation logic that has been published by CDISC in the "open source" domain. Another advantage is that it allows the user to select which rules are to be applied in a validation. In near future, it will also be possible for companies to add their own (e.g. quality checking) rules to the engine, a feature that will also be supported by SDTM-ETL.

## Use of Oracle-SQL for generating and populating relational databases

SDTM-ETL comes with features for generating (SQL-based) databases with sets of SDTM/SEND tables. Unfortunately, there are still many "dialects" and "specializations" of SQL. The one that is implemented by default will work for most users, and is e.g. compatible with MySQL.

We do now also support Oracle-SQL. If you want to use Oracle-SQL, just add a line:

```
useoraclesql=true
```

to the properties.dat file, and the software will generate Oracle-compatible SQL commands when using one of the SQL features of the software.

## Extended CLI (batch) execution

SDTM-ETL can also be used to execute developed mappings in non-GUI, i.e. "batch" mode, using a "Command Line Interface". We have now extended the parameters used for the CLI, especially due to new features that have been introduced in versions 4.4 and 4.5, such as [automated generation of VISITNUM for "unscheduled" visits](#).

## Other minor updates

The algorithm for sorting according to the "Define-XML keys" has been improved. For each key variable, whether sorting is done alphabetically or numerically is now decided on by the datatype for the variable in the define.xml, rather than on the variable name.

## Bug fixes

The tooltip on the button "round" (for the "round" function) in the "Mapping Script Editor" was incorrect. It said that this function was for rounding to the nearest integer. This is incomplete: the function also rounds to the provided number of characters after the decimal point (second argument of the function). For example:

```
$pi = 3.1415927
```

```
$a = round($pi, 0); => results in "3"
```

```
$b = round($pi,4); => results in "3.1416".
```

The tooltip has been updated to reflect this.

Under certain circumstances, the "Dataset Label" was not written to the "header" in SAS-XPT datasets. This has been corrected.

## Limitations

As CDISC is currently updating the [Dataset-JSON standard](#) (to version 1.1) not all newer features have been implemented for generation of SDTM or SEND datasets in Dataset-JSON format. We recommend users to not use the Dataset-JSON export for production at the moment, also as it is still using Dataset-JSON v.1.0.

Once Dataset-JSON 1.1 is final and published, we will again adapt the software for the new standard, which we expect also to become the submission format of choice of the FDA in near future.

## Further development of SDTM-ETL

It is now very clear that the FDA is committed to push the replacement of the SAS Transport 5 (XPT) format by the by CDISC developed [Dataset-JSON format](#): the pilot with the FDA was extremely successful, and a CDISC working group (of which we are part) is currently refining the new standard, which will be name Dataset-JSON 1.1.

It is also very clear that FDA fully supports the [CDISC CORE project](#) for validation of submission datasets. For example, FDA has asked CDISC to add all "FDA business rules" to CORE. Essentially, this means that for validation, Pinnacle21 is expected to be replaced by CDISC CORE, also at the agency.

Both these have implications for the future versions of the SDTM-ETL software: the next major version (5.0) will have Dataset-JSON 1.1 as the primary format for the generated datasets and SAS Transport 5 as the secondary. Support for the old Dataset-XML format will probably be terminated.

As one of the co-developers of CDISC-CORE we will further support and extend the use of this open-source validation software, so that the world can finally get rid of this buggy and user-unfriendly P21 software.

We are also thinking about features such as "validate as you map", providing immediate feedback for the CORE validation engine for each case that a mapping was added or changed. Also, as soon as it is available, we will add and support the feature that users (or companies) can add their own validation rules to the engine in a secure way<sup>4</sup>.

---

<sup>4</sup> also meaning that any user- or company-specific validation rules are not shared with CDISC nor anyone else outside the company.