Validating CDISC Data with the SAS® Clinical Standards Toolkit

Nicolai Haag
HMS Analytical Software GmbH, Heidelberg, Germany

HMS Analytical Software

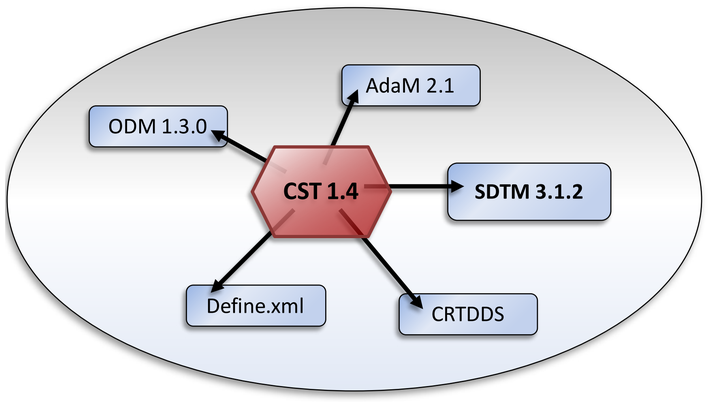
Introduction

Pharmaceutical companies increasingly apply industry-wide clinical data standards to meet the requirements of the US Food and Drug Administration when submitting New Drug Applications electronically and to ease the data exchange with partners and Clinical Research Organisations. A global standards organization is the Clinical Data Interchange Standards Consortium that published the Study Data Tabulation Model of human clinical study data tabulations for submission to regulatory authorities. The FDA stores CDISC SDTM data plus their accompanying metadata in a so-called Janus data warehouse. The metadata of the content and structure of the submitted clinical data are described in a machine readable XML document named Case Report Tabulation Data Definition Specification (CRT-DDS or Define.xml). The XML schema for the define document is also based on an extension of another standard, the CDISC Operational Data Model (ODM).
Before loading clinical data and the Define.xml into the Janus warehouse several validation checks are performed on the data and metadata verifying the SDTM and CRT-DDS conformance of the submission deliverables deploying the WebSDMT software. After successfully loading the files into Janus additional validation checks are employed within the Janus reviewer tools.

CDISC

Synopsis

By providing The SAS Clinical Standards Toolkit, SAS supports companies beeing compliant with regaluroty stanards within the CDISC environment. The framework is SAS Macro based and free available for SAS customers.

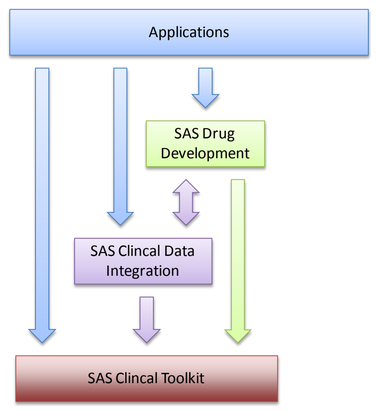


The SAS Clinical Standards Toolkit initially provides a set of standards and functionality aimed to generate the Define.xml and to perform validation checks against implemented standards. Currently the SAS CST 1.4 supports the CDISC-SDTM 3.1.2, CDISC-CRTDDS 1.0, ODM 1.3.0, ADaM 2.1, Adam 1.1 of Validation Checks, and CDISC-Terminology-201003 standards and has implemented most WebSDM, the Janus and several SAS-developed checks to back up the SDTM compliance.

Fields of application

Exceeding these standards, the toolkit has also the capability to be extensible and configurable to add new validation checks, new versions of standards, custom standards and upcoming standards. To manage that, the modular designed CST is built as a framework plus various pluggable standard modules that contribute to the centrally managed process runs. By providing an extensive process library of utility programs and data the toolkit enables the user to build robust processes to accomplish extensions to the allready provided tasks and standards.

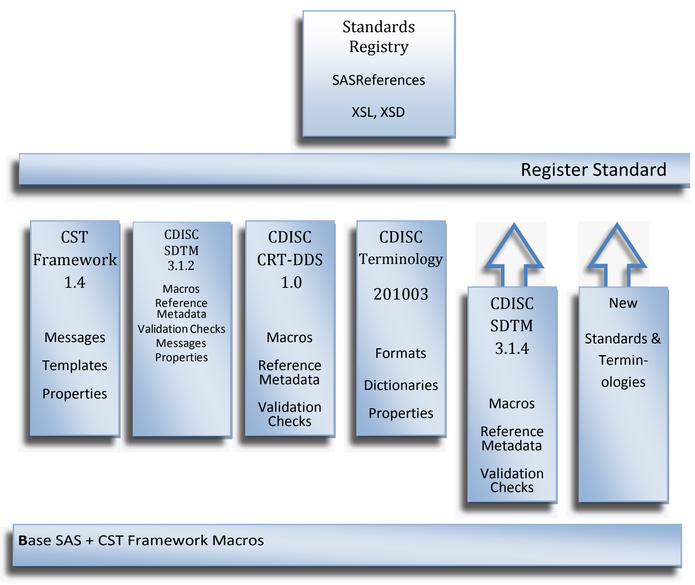
sas®



Of course, the SAS Clinical Standards Toolkit can be used in the context of other SAS Solutions. It can be integrated inwithin the SAS DI Studio in order to exploit the metadata management and code generation. It is also possible to use the SAS CST within the context of the SAS Drug Developemnt to use a regulatorey compliant repository.

Structure

The Global Standards Library consists of a registry where standards are registered and a number of already registered standards. It is the central concept of the SAS CST to use the Standards Registry in order to register oder unregister new or out of date standards.



The leftmost four standards are delivered with the current version of the CST. Additional standards as they become available or modifications of existing standards or even completely company defined standards can be registered within the framework.
The CST framework is also registered as a standard in order to integrate framework messages, templates and properties with those from the real standards. Nevertheless, framework macros reside in the usual "sasroot" location.

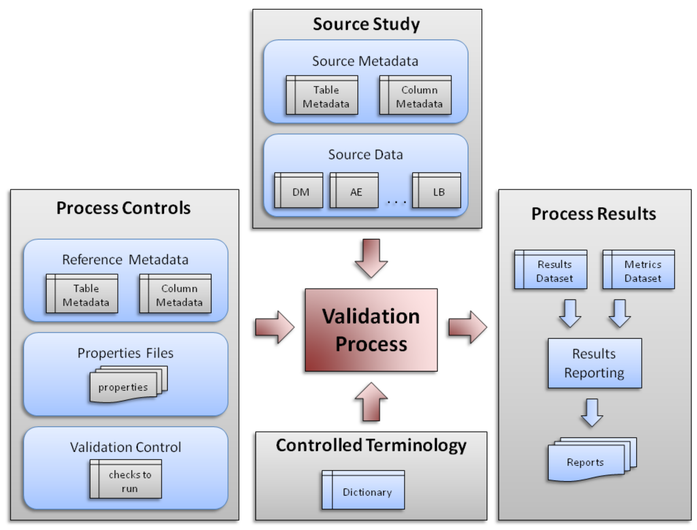
For maintenance tasks (registration of standards, retrieval of data and metadata from the library) there are a number of utility macros which operate purely on the Global Standards Library without involvement of user data or metadata. It is recommended to interact with the Global Standards Library only by use of those macros.

Validate SDTM Data

One of the most widely used funciontalities of the SAS CST is the validation of SDTM Study data.

In order to use the SAS Macros of the CST there must be created several SAS Datasets and referenced within the central reference Dataset itself. For the Study itself metadata of the columns and tables has to be provided additionally to the data itself. Within the process controls content there has to be reference table and column metadata. The validation checks to be performed are selected by creating a dataset containing a subset of all available validation checks (and custom validation checks). Additionally contolled terminologies may be referenced wirhin the validation process (i.e. Medra).

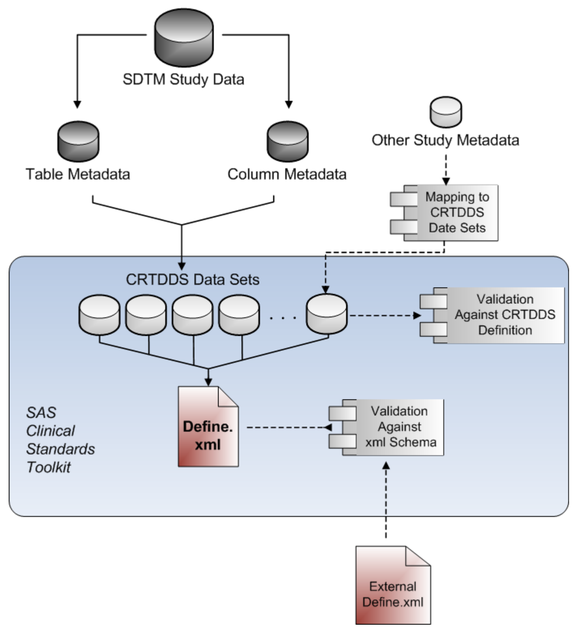
All informations of the locations and versions of the provided standards, metadata and datasets have to be collected in the central dataset named SASReferences. The structure of this dataset is predefined within the CST (same as the strucute of the metadata).



After the validation proess there will be two main result datasets. The result file itself with the findings of the validation process and the metrics file providing optional summary data of the validation results.

Create Define.xml

Creating the Define.xml is a central feature of the SAS CST. The Toolkit supports the creation ans validation of the Define.xml is order to submit a study to regulatory authorities



The main underlying technique for generating and validating the Define.xml is transferring the hierarchical structure of the Define.xml into a relational structure of SAS files with foreign key relationships. Roughly speaking that procedure yields SAS datasets that embody the Define.xml elements and sub-elements whereas the attributes are realized by variables - the result is also known as the SAS representation of the CDISC CRT-DDS standard.
In the next step the verification of datasets representing the Define.xml standard against the underlying relational model scrutinizes the referential integrity, expected variables and value conformance. Finally, the Define.xml is generated from the CRT DDS representing datasets via a further intermediate three level deep XML file that in turn is transformed into the target Define.xml. This transformation is carried out with the help of Java components using element specific XSL files that reside in the XSL repository of the global standard. Optionally a schema validation of the generated (or of an external) Define. xml can be done too.

Nicolai Haag
Dipl. Dok. (FH)
Senior Software Engineer & Trainer
HMS Analytical Software GmbH
Rohrbacher Str. 26 • 69115 Heidelberg
Phone: +49 176 16051-114
E-Mail: nicolai.haag@analytical-software.de

> *Requirements*

The only system requirement is to have Base SAS installed.

> *How to Get the CST*

For SAS 9.2 and above, the toolkit is a separately orderable component that is available at no additional charge to currently licensed SAS customers. Normally, you should contact the SAS representative in your company to get the CST. For SAS 9.1.3, the CST can be downloaded directly from the SAS website.