

Meeting CDISC Standards with SAS

Scott McGregor of S-cubed discussed how CDISC standards can be put to work using SAS Software

These days it is commonplace to see the acronym CDISC projected onto pharmaceutical sector meeting room walls. The Clinical Data Interchange Standards Consortium has developed a set of standards that “support the acquisition, exchange, submission and archive of clinical research data and metadata” (www.cdisc.org). This support was exactly what the industry needed and as such, CDISC standards with acronyms such as CDASH, SDTM and ADaM, and the implementation thereof is something that most companies have high on their project priority list.

Although the standards themselves come for free, the implementation of them does not. Rather, substantial initial investment of time and effort is required before the real benefits can be seen. This initial investment can involve reallocation and training of resources, changing of internal standards, reengineering of processes and systems and in most cases some outsourcing to companies specialising in these standards.

By using SAS Software at key points in the data handling and reporting process, you can reduce the initial costs required to implement these standards and achieve their full benefits.

A Little About CDISC Standards

The most developed and widely implemented of the CDISC standards is the Study Data Tabulation Model (SDTM), partly because the FDA (Food and Drug Administration) in the US recommends SDTM as the model for submission of raw clinical data. The model is not only useful for submission but also serves as an excellent standard for the interchange of raw clinical data. Due to this many companies have adopted SDTM as an internal standard for pre-analysis clinical data.

For statistical analysis data CDISC provides the ADaM model. It supports the clear and unambiguous communication of statistical results, covering both the derived and statistical data as well as result metadata. The FDA does not yet recommend or require the submission of analysis data in CDISC ADaM format, however, due to the strong links between ADaM and SDTM, using them together for a submission increases transparency, which may make this the logical next step. Several companies have already started submitting analysis data in the ADaM format.

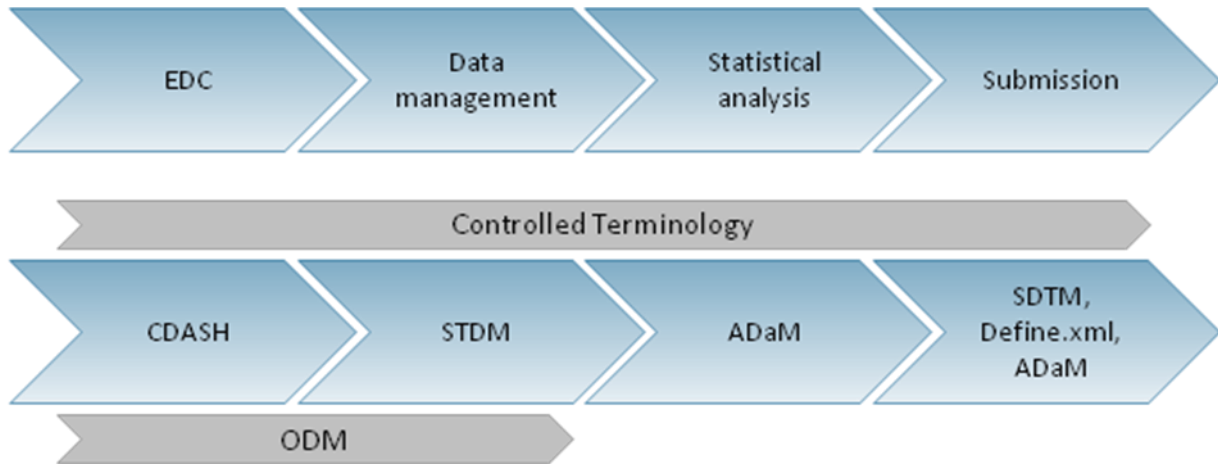
In October 2008 the Clinical Data Acquisition Standards Harmonization (CDASH) team released the first production version of a CDISC standard for data collection pages. The definition of this standard is still in progress, but this version already provides several benefits in the process as described by Joris De Bondt in the September issue of EPC.

The CDISC models are closely linked together and share the same controlled terminology, which is a constantly growing list of contents standards across all the models.

The Operational Data Model (ODM) is a CDISC model for the physical transfer of clinical data and associated metadata between various clinical data handling systems. For example, several widely used EDC systems use ODM for the transport of clinical data or metadata. The FDA’s long term strategy is to use HL7 messages for the submission of clinical data, so the ODM standard is not expected to play a role within an FDA submission; however it has tremendous value and potential as a vendor neutral mechanism for transport of data between systems.

Figure 1 illustrates how the CDISC models relate to the processes within data handling, reporting and submission of clinical data. By using SAS software at key points in this process, you can reduce the initial costs required and achieve the full benefits of the CDISC standards.

Figure 1: CDISC standards in the clinical data process



Using SAS to Import and Store CDISC Metadata

Currently, electronic descriptions of some of the CDISC models are available on the CDISC members' area of www.cdisc.org. The descriptions exist in MS Excel sheets, which can be imported by SAS and transferred to the data handling and reporting environment. This type of model metadata is very useful for data validation, setting up of EDC tools, and reporting and streamlining processes. By using SAS to import this information a dynamic link is created between the global, publically available, metadata and the internal systems within a metadata-driven environment.

At present the number of electronic model definitions available on www.cdisc.org is limited, and not truly in a platform-independent format. The extension of the concept of defining and sharing the CDISC metadata in a platform-independent format is expected to be a focus area for CDISC in the not too distant future. Currently work is ongoing within the CDISC teams to replace the existing MS Excel sheet versions of the model definitions with metadata specifications in ODM format.

Regardless of whether the model metadata is stored as ODM or MS Excel files, SAS can be used to create the link between the definitions and the internal system metadata.

Conversion of Data Using SAS

The three different CDISC standards - CDASH, SDTM and ADaM - are to some extent different representations of the very same clinical data. It is therefore relevant to ask why there are three different models for the same clinical data. The response is that the different models fulfil different purposes. CDASH is created to ease and streamline data collection, SDTM to be a solid standard to hold and interchange data and ADaM to analyze them. These needs are quite different, which is why it has not been feasible (or possible) to make a 'one-size-fits-all' model.

As a consequence, it is often necessary to convert data from one format to another. As the three models have evolved, it has always been a priority to make sure they are in sync with each other in order to increase transparency and to minimise the conversion effort. This means that the

differences between the three are no greater than necessary, and that a lot of the conversion can be undertaken using generic, data-driven SAS macros.

Such macros can, for example, turn the CDASH date/time variables into SDTM ISO8601 date/time, create SDTM sequence numbers etc. A more comprehensive description of such generic, data-driven, SAS macros can be found in PhUSE 2009 Paper #CD11 written by Niels Both, S-cubed ApS.

For legacy data conversion it is often required to re-annotate CRFs from previous studies, adding the appropriate SDTM standard field names to the CRFs. This is a laborious task that, if done manually, can take almost as long as the conversion programming itself. SAS can also be utilised here to reduce the time required to complete the task, by reading from mapping sheets containing page numbers and field names and activating PDF functionality to add the annotations to the appropriate pages.

Using SAS to Police Data Standard Compliance

The need to check that data adheres to certain models is obviously important within any standard driven environment. If meta- descriptions of the data are available within the environment, SAS Proc Contents or other SAS procedures can be used to compare the actual data with the defined description, to ensure that all required and expected variables exist.

In addition more advanced SAS macros can be built to further check the data, for example checking that the dates apply to the ISO8601 standard or that unique keys are in fact unique. SAS Proc CDISC offers the opportunity to do some basic checks on SDTM, but if the data are to be used for a submission it should be noted that the FDA are doing more advanced checks, and Proc CDISC is not built to replicate those checks.

Why wait until the FDA find issues with the submitted SDTM data, when you can build SAS to identify these errors before submission, removing unneeded turn-around times in an often short submission period?

Traditionally the main focus on policing a data standard has been on SDTM, but it is also possible to use SAS to check analysis data to determine the degree of compliance with the ADaM model for example.

Creating Scientific Output Using SAS

SAS software is especially good at producing outputs in the form of tables, figures and listings (TFLs). The TFLs can be created for example using Proc Report, Proc Tabulate and SAS/Graph. Using SAS ODS the output can be created in different formats i.e. PDF, RTF, HTML and many more. With CDISC standards in place, SAS systems can be created to produce the wide list of outputs required in connection with ongoing safety monitoring of the study or in the final creation of the clinical report. Both ADaM and SDTM data can be used as the source for this purpose however SDTM is mainly useful for simple, standardised, non-statistical, safety outputs, for example in connection with ongoing safety monitoring.

Building packages for submission to authorities can also be a laborious task. SAS can be used to build the submission package with SDTM or ADaM data at its centre and define.xml documentation describing the data structures within.

Conclusion

SAS is already well known and accepted within the Pharmaceutical industry and as such there are many experienced, knowledgeable SAS resources available. SAS can be used to build a validated, maintainable foundation of CDISC standard data and metadata through the creation of generic macros that can be re-used again and again.

Data and metadata can be verified using a SAS-built CDISC standards checker which dynamically reads the up-to-date version of the CDISC standard specification and identifies data errors quickly, improving accuracy and efficiency and reducing data cleaning time.

The combination of generic SAS Macros and CDISC standard source data means that you can put systems in place that work with **all** your data. Previously, for example, the creation of pooled analyses or integrated databases was a drawn out process involving much data cleaning, assumption and educated guessing. The initial investment in moving your data to this standardised environment pays dividends many times over for existing output requirements and allows more focus on going deeper into the analysis of your data.

By using SAS at key points of a CDISC standardised data handling and reporting process, precious and costly time and resources can be saved. Once the foundation of standard data and metadata is in place, SAS can really reap the benefits of standard input data by producing applications, detailed analyses and comprehensive reports in a stable and consistent manner. You really can put your CDISC standards to work using SAS software.

Scott McGregor is co-owner and SAS Specialist at S-cubed ApS. He began his career in the pharmaceutical industry as a SAS Programmer in 1990. Since then he has worked in a variety of sectors before founding S-cubed in 2007. Through his work in the pharmaceutical industry, Scott is regularly involved in data conversion, reporting and submission projects and works closely with the implementation and use of CDISC standards.