# CDISC: Why SAS® Programmers Need to Know

Victor Sun, Novartis, East Hanover, NJ

## ABSTRACT

As CDISC advances global clinical data standards, it has been widely and rapidly expanded in pharmaceutical industry both domestically and globally. The pilot studies with CDISC based electronic submissions to FDA have been conducted from about ten companies. CDISC attention, however, has not been addressed enough in STAT group even though its implementation is bringing up a significant impact, especially to SAS® programmers in terms of work skills and job security. After briefly introducing CDISC development and outlining its implementation approaches, future expectations of challenge and opportunity facing to SAS® programmers are discussed. The suggestions provided are intended as a useful reference for SAS® users in understanding a potential trend of electronic submission with CDISC and leveraging job skills to satisfy this new change.

## INTRODUCTION

CDSIC (Clinical Data Interchange Standards Consortium), as one of new data standards for clinical development has been developing for last six years. Due to the core concept of open, multidisciplinary nature, its sponsorship, membership and participation have rapidly expanded across various organizations in the US, Europe and Japan. FDA has appointed Liaisons to CDISC, and completed a pilot study on patient profiler reviewing for electronic submission by using CDISC standards.

From "many one though impossible" to "the concept has become a reality", CDSIC is no longer the concern of only IT and DM, but of the Statisticians and SAS® programmers as well. The impact, however, of CDSIC implementation from top of company level to bottom of a trial team is unavoidable per current exercises. After a brief overview of CDISC, its rapid global expansion and implementing approaches are introduced. Then, the essential reasons to understand CDISC and new challenges facing SAS® programmers are addressed. Those comments are written to provide an insight to suggest what one should focus on during and after CDSIC is essential to be adopted in your organization soon so that one can take a competitive advantage by following up new technology, expanding job skills, heightening productivity and certifying the acquiring skills along career path even under today's tough job market.

## OVERVIEW

With the recently publications of the FDA guidelines for electric submissions (1-4), they require more data standards than ever before so the new required deliverables, such as define.pdf and blanckcrf.pdf, etc, can be clearly and conveniently provided to agency for reviewing. To facilitate this process, CDISC was founded to develop data standards that could be adopted by reviewing agencies, sponsors and vendors to support the electrical aquisition, exchange, submission and archiving of clinical data and metadata for drug development.

CDISC is an open, multidisciplinary and non-profit organization. The mission of CDISC is to develop of global, vendor-neutral, platform independent standards to improve data quality and accelerate product develop in our industry. During last six years, CDISC has established a number of working teams each with a specific area of focus below.

**ODM** – Operational Data Modeling team
to build the standard data exchange models that are being developed to support the data acquisition, exchange and archiving of operational data by using XML.

**LAB** – Laboratory Data Modeling team
to build a superset of data items that fully describes a clinical trial to the satisfaction of all the stakeholders included. The superset of fields constitutes the contents of the model. It will be operated by ODM.

**SDS** – Submission Data Standards team
to build the standard metadata models for supporting the data flow from the operational database to regulatory submission.

**ADaM**– Analysis Dataset Modeling team
(a) to build on metadata CRT models developed for safety domains, adding attributes and examples specific to statistical analysis;
(b) to use sample statistical results as guide for developing data models; and
(c) to initially focus on primary and secondary efficacy variables.

There are also a number of the task forces associated with the four operational teams mentioned above. They include the strategy task force, the public relations and communications task force, and the membership task force. The CDSIC group also cooperates with HL7 (Health Level 7) to harmonize existing CDISC work with the HL7 reference information model (RIM).

## EXPANSION

CDISC participants represent all stakeholders in the pharmaceutical industry. Within a total of 85 CDISC member companies, there are 47 sponsors member companies, 12 member companies and 26 associate member companies who have joined in so far and helped CDISC to achieve its mission.

Activities associated with electronic submission based on CDISC are briefly summarized below.

- Included 9 sponsors who submitted electric data and metadata based on CDISC SDS V2.0
- Facilitating FDA adoption of Patient Profiles tool and CDISC SDS Standards
- Demonstrated value of standard domain formats
- Evaluated the pilot data, and second phase pilot in planning (with additional data)

Due to open, multidisciplinary and cross-functional composition of CDISC, it not only widely recognizes in domestic, but also rapidly expands in global base, especially in year 2002. The milestones of CDISC development in Europe and Japan are highlighted below.

**In Europe**

- 1999- First CDISC presentation at DIA CDM Brussels
- 2002 (Jan) - European Repres. to CDISC BOD
- 2002 (Apr) - Regional CDISC Group proposed
- 2002 (May) - European CDISC Launch, Brussels
  - 10 participants from Pharma and CROs
- 2002 (Jun) - CDISC "Kick-off" meeting
  - 50 participants from 7 counties
  - Initiation of EU CDISC Coordinating Committee(e3C)
- 2002 (Nov) - Next action at DIA Clinical Data Management Conference in Edinburgh

**In Japan**

- 2002 (Jan) - First CDISC meeting
  - DIA Clinical Data Management meeting in Tokyo
- JPMA
- 2002 (May) - CDISC Meetings in Tokyo
  - 50 Pharma, CROs, Vendors
- 2002 (Jul) - CDISC Presentations by Japanese representatives to MHLW (Japan Regulatory Agency) and data manager group
- Currently planning Japan CDISC Coordinating Committee(j3C)

## IMPLEMENTATION

Based on the current exercises or pilot studies from some companies, various approaches at different levels are presented for implementing CDISC. DM and IT groups usually initiate the adopting activities at small scale. Only few know about CDISC from STAT group. In general, there are three approaches to implement CDISC. Each method with its major tasks is briefly described as follows.

**Forward**

- Start from CRF, work toward submission
- Decide what data groups are required
- Decide what data items to be collected
- Decide what data items to be collected
- Develop database standards
- Develop reporting dataset and display standards

**Backward**

- Start from regulatory requirements and CDSIC standards for submission
- Create continuity by using CDSIC variable
- names in clinical data management system
- Make sure CRFs collect data consistent with CDISC submission data modes

**Forward and Backward**

- Work on "principles" for CRFs, and database design, and analysis and database design, and analysis and statistical displays started simultaneously
- Forward: Specific standards developed from standards CRFs to database to reporting
- Backward: CDSIC submission data models used to develop standard CRFs

## CHALLEANGE VS. OPPORTUNITY

It is estimated that about 65% - 75% submission information is associated with safety data. A big volume of listings (e.g. for CRT dataset and patient profile, etc.) is always included even in electronic submissions. Only about 30% of programming time is used to generate statistical results with SAS®, and the rest of programming time to familiarize data structure, check data accuracy, and tabulate/list raw data and statistical results into certain formats(5). This non-statistical programming time will be significantly reduced after implementing the CDISC because of CDISC's uniform data structure, its useful functions of four models and some ready-to-be-used CDISC based tools (e.g. PPV – Patient Profile Viewer, etc) being completed or ongoing. Some listings are expected to be dropped and some certain tables are not necessary to be generated. It is reasonable to consider production of integrated, bookmarked and hyperlinked documents to be the final deliverables without too much redundancy. Those could be addressed in the FDA guidance on submission of CRT datasets in 2003, which the FDA recently announced it plans to be issued.

Because of these changes, new challenge to SAS® programmers is obvious and inevitable(6). At the same time, the changes also provide a great opportunity to people who are willing to reshape their job skills. A few suggestions are listed below.

**Familiarizing with CDISC**

With the new requirements of electronic submission, CRT datasets need to conform to a set of standards for facilitating reviewing process. They no longer are created solely for programers convenient. SDS will be treated as specifications of datasets to be submitted, potentially as reference of CRF design. Therefore, statistical programing may need to start from this common ground. All existing programs/macros may also need to be remapped based on CDISC so one can take advantage to validate submission information by using tools which reviewer may use for reviewing and to accelerate reviewing process without providing unnecessary data, tables and listings.

Although CDISC is not rigid, nor does it dictate what data a sponsor should collect, it will become a "Bible" which SAS® programmers can often refer to for understandings of CRT datasets, programming and preparation of submission.

**Understanding Metadata Concept**

Metadata, by a loosing definition under CDISC is data about data. Depending on the strategy of CDISC to be implemented and functionality of IT, DM and STAT groups in different companies, the requirement of understanding metadata concept to SAS® programmer may vary. Right now, the way we are applying SAS® is to use programming-intensive activity for creation, revision and reuse of reports without information about tables content or style that can go back to SAS® in an automated way. SAS® is now integrating with XML for more convenient reporting process(7). Four components of this underlying technology such as table content, table style, database map and publishable XML are mainly based on metadata concept.

Metadata as one of three data groups (CRT datasets, analysis datasets and metadata) are also required under CDISC submission or reviewing convenience.

**Knowing XML**

The language of the four models(e.g. ODM, LAB, SDS and ADaM) in CDISC is based on XML(eXtensive Markup Language), which is new emerging standard marketed as the ideal tool to help industries streamline business processes in terms of data management, electronic communications and contents managements. Both data and metadata are key components in XML. Therefore, SAS® and XML are now cooperating. XML Engine in SAS® v9.0 is built up so one can import a wide variety of XML documentation. SAS® does what is does best - statistics, and XML does what it does best - creating report-quality tables by taking advantage of the full feature set of the publishing software. This conversation can produce report-quality tables in an automated hands-off/light out process.

In addition, it is expected that future FDA standards for delivery of submission components may well specify XML as the format of choice. Some examples below indicate this intention. Those include XML-based FDA PPV tool, FDA XML standard for exchange of annotated ECG, and the updating guidance on submission of CRT datasets which will be released in 2003.

**Intensifying Analysis Skills**

Data analysis and result reporting are two major tasks to SAS® programers. With CDISC, XML and SAS® cooperation, the routine reporting process is expected to be more efficient than before. The task of analysis data, however, will be addressed more for selections of better methods on data analysis, tabulation on result reporting and visualisation on data presentation.

One way to strength of analysis skill, of course, is to closely follow up new features released from SAS® due to its continuous improvement along four cornerstones (e.g. scalability, interpretability, manageability, and usability). PROC GAM in v8.2 for data analysis and PROC GRADAR in v9.0 for data visualisation are just two examples for those new features.

Another way of enhancement of analysis skills is to try to understand clinical data more so more meaningful analysis can be performed at some aspects such as benefit/risk assessment, simultaneous evaluation adverse evens and laboratory value, etc. on safety information. Some exploratory analysis with graphic skills could be helps to obtain more information for concerning drug safety.

## CONCLUSION

With the new requirements from updating electronic submission and CDISC implementation, understanding only SAS® may not be good enough to fulfill for final deliverables. It is a time to expand and enhance the job skills from various aspects under new change so that SAS® programmers can take a competitive advantage, and continue to play a main role in both statistical analysis and reporting for drug development.

## REFERENCE


1. US Food Drug Administration. Guidance for Industry, Providing Regulatory Submission in Electronic Format-General Consideration. IT-2, January 1999.

2. US Food Drug Administration. Providing Regulatory Submission in Electronic Format-NDAs. IT-3, January 1999.

3. US Food Drug Administration. Guidance for Industry, Providing Regulatory Submission to the Center for Biologics Evaluation in Electronic Format-Biologic Marketing Applications. November 1999.

4. US Food Drug Administration. Example of an Electronic New Drug Application Submission. 2/17/99

5. Victor Sun and Greg Steffens, "Comparison Multiple Database Structures by Using SAS® Prior to Dealing with Clinical Trial Data", Paper presented in PharmaSUG2000. May 2000.

6. Steven Light, "The Changing Nature of SAS® Programming in the Pharmaceuticals Industry", Conference Proceedings at PharmaSUG2001, May 2001, 255-257.

7. Michael Palmer and Cecillia Hale, "Software: Replace SAS® Programmers with XML_Documents to Help People and Computers Be Happier with Each Other", Conference Proceedings at PharmaSUG2001, May 2001, 245-249.


**Useful Web Addresses**

http://www.cdisc.org/
http://www.xmltimes.com

## ACKNOLEGEMENTS


The author wishes to thank Dr. William Stewart and Mr. Raymond Heimbuch for their support and reviewing on this paper.


## TRADEMARK INFORMATION



## CONTACT INFROAMTION

Your comments and suggestions are valued and encouraged. Contact the author at


Victor Sun, Ph.D.
Novartis
One Health Plaza
East Hanover, NJ 07936
Work Phone: 862-778-6267
Email: victor.sun@pharma.novartis.com