

Paper 183-2010

The Past, Present, and Future of Clinical Data Standards

Chris Decker, d-Wise Technologies, Raleigh, NC

ABSTRACT

One of Albert Einstein's many great quotes during his life included the following: "The distinction between past, present and future is only a stubbornly persistent illusion". This quote is a perfect microcosm for data standards in the clinical research industry.

Over the last decade the clinical research industry has attempted to work toward a common data standard with the goal of accelerating drug development by improving the data collection, transformation, analysis, and submission process. The adoption of industry wide standards has been lethargic and led to a plethora of challenges. This paper will provide a history of clinical data standards, the role of SAS in these standards, and a forward view on how standards might evolve over the next decade.

INTRODUCTION

Over the last decade the clinical research industry has attempted to work toward a common data standard with the goal of accelerating the drug development process by improving the data collection, transformation, analysis and submission process. The adoption of industry wide standards has been slow going and led to a wide range of challenges. Decades of legacy processes and non standard data have led to internal company data standards that are inconsistent and differ wildly from company to company. Within a company there is a many levels of consistency depending on the enforcement of these standards. Across companies the exercise of trying to combine data becomes a bottomless pit of unusable data. Given the issues of drug safety over the last decade, both pharmaceutical companies and the FDA are accelerating the need for data standards across the industry. This paper will provide a history of clinical data standards, the current state of the union, and one person's peek into how standards might evolve over the next decade. The paper will also discuss the parallel role of SAS in these standards over time.

In the famous story, a Christmas Carol, Ebenezer Scrooge is visited by the ghost of Christmas past, present and future with the hope of showing him the error of his ways and positively changing his future. Using the Christmas Carol story, this paper will describe how the Ghost of Clinical Standards past, present and future can help us change our ways and provide hope for the future.

THE PAST

The Ghost of Christmas Past visited Ebenezer Scrooge and reminded him of the simpler times within his childhood. If the Ghost of Clinical Standards Past visited us today he would tell us a similar story. With the introduction of computers, the data collection process for clinical trials was a new fangled idea and the idea of data standards really didn't cross many people's minds. Each study had their own unique set of data and the perception was that there was no way you could reuse information across studies. Clinical studies and the associated data were 'special'.

IDENTIFYING CHALLENGES

As technology became more robust clinical programmers started to realize the inefficiencies in recreating processes and metadata from scratch every time as well as the overlap in data elements across studies. They also saw the many inconsistent methods for collecting specific data elements that seem simple on the surface. The simplest examples that conveys this challenge is the definition of gender of a subject (Male or Female) within a clinical study. At quick glance, this seems like a very a very non ambiguous data point. However, as you can see from the figure below even something this simple can lead to challenges.

Study 1 Variable: GENDER Values: M, F	Study 2 Variable: SEX Values: Male, Female
Study 3 Variable: SEX2 Values: 1, 2	Study 4 Variable: GEND01 Values: 0, 1

Figure 1. Inconsistencies in Clinical Data Collection

It's fairly obvious from this figure that the lack of data standards can lead to a myriad of issues when different people define codes, variables, and processes across studies. This led companies to begin defining data standards within their organizations.

COMPANIES BEGIN TO STANDARDIZE

Clinical research companies began to define companywide data standards to improve efficiencies in both the reusability of tools as well as the ability to combine data across clinical studies. While this seems like it would be a fairly straightforward process, companies soon find themselves buried in discussions around the best way to accomplish this task.

Many of the data managers, clinical programmers/statisticians, and clinicians had different perspectives primarily due to their specific needs of the data. These needs, while similar in some cases, differed more often than not. Data Managers were very focused on defining a data standard that optimized the data collection process and reduced the need to reconcile data issues. Clinical programmers and statisticians wanted data that was analysis ready to generate the tables, listing, and figures needed for submission to the FDA. The clinicians didn't necessarily understand the need for standards, but just wanted as much information they could have within their 'data'. These differences among users led to many late night discussions to hammer out standards.

In addition, there are a number of unique aspects of the clinical research process which make defining a rigid data standard complex. First, each disease studied within clinical research has its own methodology and testing associated with it. So the way you study heart disease is very different from how you study asthma. Therefore, the type of data, as well as the way you collect it, varies greatly. Second, advances in medical science occur very rapidly within a disease and thus lead to even more changes in how the data is collected, analyzed, and reported on. Finally, the clinical research industry is governed by rigorous regulation and the data collected in this process has many audiences. A drug development company uses the data in one way whereas regulatory agencies use it differently.

All of these differences lead to roadblocks in defining a common standard. These additional complexities create even more challenges with defining, using, and maintaining a data standard.

CDISC IS BORN

Over the last 25 years, data standards have slowly been adopted for the collection and transfer of clinical data. In the beginning the focus was on real time collected at hospitals, and the standard used for this exercise was Health Level 7 (HL7). However, this standard was more for the individual patient data points in the health care arena and could not easily be translated to the clinical research area.

In the late 90's a group of individuals decided to get together to see if they tackle the monumental challenge of defining a data standard across clinical research. The Clinical Data Interchange Standards Consortium (CDISC) was formed with the mission "to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research". Over the last decade a number of models have been developed within CDISC to support the needs of clinical trial data. The table below contains a list of the more critical models CDISC has developed over the years and their purpose.

Table 1. Summary of Relevant CDISC models

Model/Standard	Purpose
Operational Data Model (ODM)	XML specification supporting interchange of data, metadata or updates of both between clinical systems
Clinical Data Acquisition Standards Harmonization (CDASH)	Data model for a core set of global data collection fields (element name, definition, metadata)
Submissions Data Tabulation Model (SDTM)	Data model supporting the submission of data to the FDA

Model/Standard	Purpose
	including standard domains, variables, and rules
Analysis Dataset Models (ADaM)	Data model closely related to SDTM to support the statistical reviewer
Define.xml	XML Specification to contain the metadata associated with a clinical study for submission
Standards for the Exchange of Non-clinical Data (SEND)	Data model extending SDTM to support the submission of animal toxicity studies
Protocol Representation Model (PRM)	Metadata model focused on the characteristics of a study and the definition and association of activities within the protocols, including "arms" and "epochs".
Terminology	Standard list of terms across all the CDISC data models.

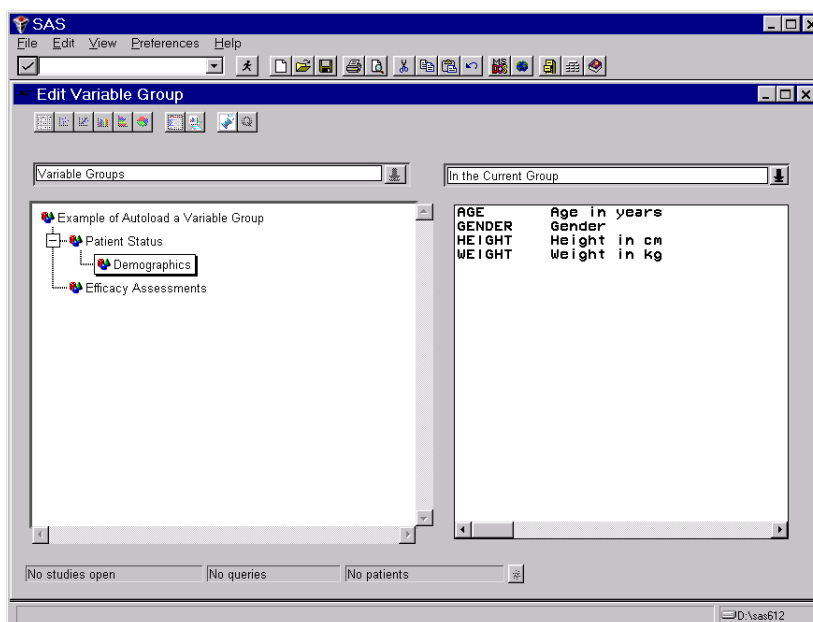
The standards described in this table have varying levels of maturity. The SDTM and define.xml are probably the most widely used CDISC components and has been referenced in various FDA documents. While CDISC has done an excellent job of laying the foundation for clinical data standards changing the large and extremely lethargic clinical research industry is a daunting task. It has taken 10 years to get organization on board with this effort and moving in the same direction.

ROLE OF SAS

SAS has always been a core component to the data processing and analysis of clinical data. In the beginning days of collecting data electronically, the clinical research industry, which consisted of predominantly non technical people needed a programming language that easy to use and could perform high end analytics. SAS was obviously a perfect fit, and this began a long and dominant use of SAS for processing clinical data. In 1999, the FDA identified the SAS V5 transport file as the mechanism for delivering data to the FDA. The FDA selected the format because it was an open format which means the structure was in the public domain and could be consumed by other technologies. By US law, the FDA must remain "vendor neutral" and cannot endorse or require use of any specific vendor's product.

In the mid 1990's, SAS began looking at building industry specific solutions and the pharmaceutical industry was an obvious target. SAS came out with two products to support data warehousing and clinical reporting. PH.DataWare was built on top of Warehouse Administrator and was SAS' first attempt at building an ETL specific tool for data transformations. PH.Clinical (Figure 2) was built as a SAS report generation tool as well as a clinical tool for viewing and exploring clinical data. Both products had some success but were not widely adopted. The ETL solution was too rigid and did not provide enough flexibility for the uniqueness of clinical data, a challenge commonly seen in ETL solutions. The PH.Clinical solution took too much programming out of the hands of hard core SAS programmers and made them work with point and click interfaces. Both products were slowly phased out in the mid 2000's.

Figure 2. Reporting Interface within PH.Clinical



PROC CDISC was developed in the early 2000's to help support the new emerging CDISC standard. It attempted to support the ODM standard within CDISC and provide tools to move data back and forth between SAS and the ODM xml specification. While it provided some basic capabilities for SAS programmers it was not fully supported by SAS and never had a production release. In addition, because the use of ODM was somewhat limited early on, there was no real demand for this capability.

The biggest challenge SAS faced in implementing solutions to support clinical data standards was the Clinical SAS programmers. Programmers within this industry have a long history of using BASE SAS for creating, in some cases, very elaborate SAS frameworks for dealing with clinical data standards. Replacing these home grown solutions that involved entrenched SAS programmers was sometimes difficult even though most of the time the home grown systems were not very robust.

THE PRESENT

In the Christmas Carol, the Ghost of Christmas Present visits Ebenezer Scrooge to show him a variety of scenes ranging from festive events to lonely orphans in an attempt to teach Mr. Scrooge a sense of responsibility for his fellow man. Again, this story provides a parallel analogy to the Clinical Standards process. The Ghost of Clinical Standards Present would tell us that industry wide standards have been adopted with mixed results and the technology to support those standards is all over the map. However, the same Ghost would tell us that we must all feel a sense of responsibility to our 'fellow' industry colleagues to help make the drug development process more efficient.

INEFFICIENCIES OF THE REGULATORY REVIEW PROCESS

Probably the one issue that has ignited the aggressive movement towards data standards is the need to make drugs safer as major issues have surfaced regarding the safety of drugs over the last decade. In the past, most companies have submitted data in their own proprietary standards, making the need to combine data across therapeutic classes of drugs an impossible task. Therefore, the FDA and other regulatory agencies cannot look at integrated data to identify safety issues before they occur. In addition, without standardized data they can use with standard review tools, the review cycle becomes slow and tedious. Figure 3 below from an FDA reviewer was taken recently and highlights the sophisticated review process. As you can tell, this is probably not the most efficient way to review data.

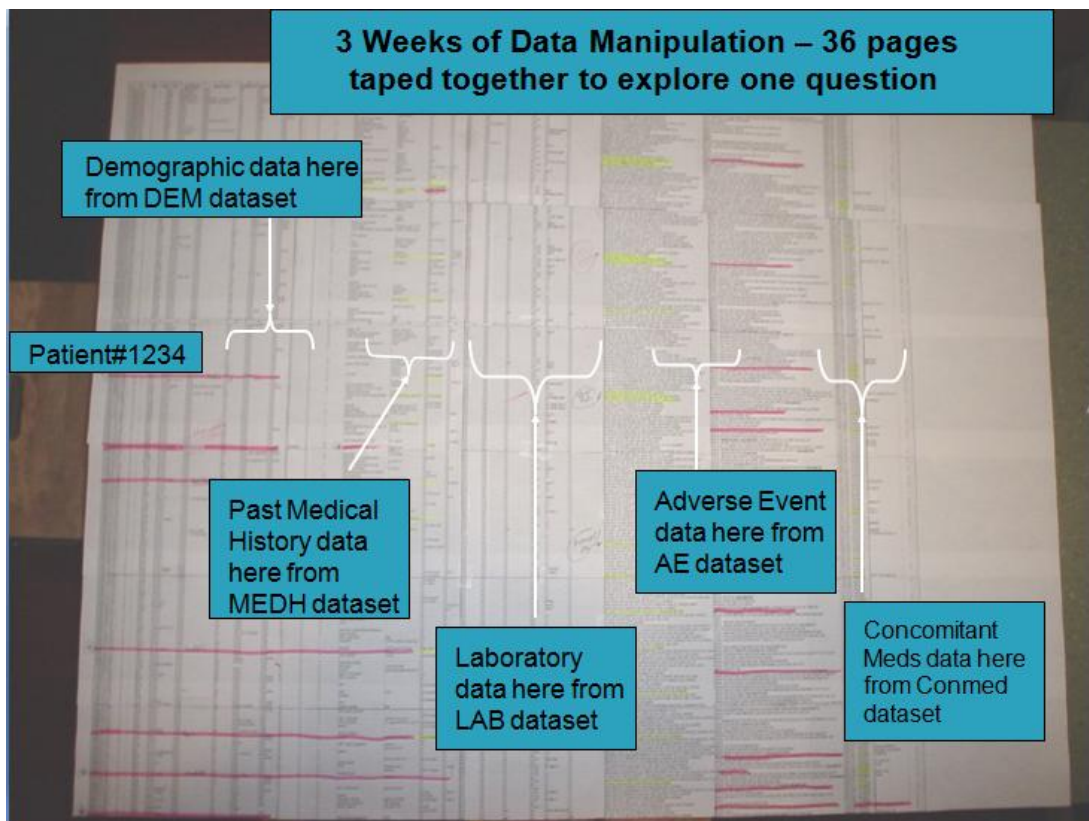


Figure 3. FDA Review Process¹

These issues have led to increased funding at the FDA to support the implementation and use of data standards including training, pilots, and the creation of the Computational Sciences Center (CSC) to drive the direction within the agency. The mission of the CSC is to create an integrated review environment including an informatics platform allowing easy access to review tools, use of data standards, and support for data management and review tool development.

Recently, the FDA gave a clear message that they want data submitted in the CDISC SDTM, ADaM, and define.xml formats even if the standards don't currently meet all their needs. The CDISC standards should be every company's baseline and it should begin during data collection. The FDA will continue to collaborate with industry and CDISC to refine the standards to meet their needs.

BARRIERS TO ADOPTION

As mentioned earlier, one of the biggest challenges in adopting standards and technology is the strong reluctance to change. "It isn't broken then don't fix it" is the comfortable cliché people use and reluctance to change is common behavior among most individuals.

Unfortunately, as companies begin to adopt data standards, they aren't defining it as an integral part of their operational process but after the fact as a necessary evil of submitting the data to the FDA. This issue began due to the widespread adoption of CDISC SDTM, a model defined for the raw data in a submission format. By defining this model first, CDISC started their standards development smack dab in the middle of the process, the creation of study data for submission in the model. At the time this made sense because the most important customer of the data was the FDA. Unfortunately, this creates challenges because it isn't the way data is collected or analyzed within the drug development process. As companies try to adopt the SDTM standard, they are very reluctant to change their processes and internal operational data standard. Therefore, if the standards are not integrated into their process and initiated much further upstream during study design and data collection, SDTM ends up being a very expensive and time consuming exercise at the end of a clinical trial.

However, companies are slowly starting to modify their internal processes to better support the standard. With the increased use of the ODM model for data transfer as well as the introduction and swift implementation of CDASH for data collection, the adoption of the standards should increase rapidly over the next decade. The data standards will now be used at beginning of the clinical trial, the data collection step, and more easily move through the data transformation and analysis steps.

Finally, the current standards do a very good job of defining the generic data structures because those data domains are very consistent across clinical studies. These include domains such as Demographics, Adverse Events, and Laboratory parameters, data usually categorized as safety data within a clinical trial. Figure 4 shows a sample of the standard SDTM domains.

Figure 4. Summary of CDISC SDTM Domains

- Special-Purpose Domains (defined in Section 5):**
 - Demographics — [DM](#)
 - Subject Elements — [SE](#)
 - Comments — [CO](#)
 - Subject Visits — [SV](#)
- Interventions General Observation Class (defined in Section 6.1):**
 - Concomitant Medications — [CM](#)
 - Substance Use — [SU](#)
 - Exposure — [EX](#)
- Events General Observation Class (defined in Section 6.2):**
 - Adverse Events — [AE](#)
 - Medical History — [MH](#)
 - Clinical Events — [CE](#)
 - Disposition — [DS](#)
 - Protocol Deviations — [DV](#)
- Findings General Observation Class (defined in Section 6.3):**
 - ECG Test Results — [EG](#)
 - Laboratory Test Results — [LB](#)
 - Questionnaires — [QS](#)
 - Vital Signs — [VS](#)
 - Microbiology Specimen — [MB](#)
 - PK Concentrations — [PC](#)
 - Inclusion/Exclusion Criterion Not Met — [IE](#)
 - Physical Examination — [PE](#)
 - Subject Characteristics — [SC](#)
 - Drug Accountability — [DA](#)
 - Microbiology Susceptibility Test — [MS](#)
 - PK Parameters — [PP](#)
- Findings About (defined in Section 6.4)**
 - Findings About — [FA](#)
- Trial Design Domains (defined in Section 7):**
 - Trial Arms — [TA](#)
 - Trial Visits — [TV](#)
 - Trial Summary — [TS](#)
 - Trial Elements — [TE](#)
 - Trial Inclusion/Exclusion Criteria — [TI](#)
- Relationship Datasets (defined in Section 8):**
 - Supplemental Qualifiers — [SUPPQUAL](#) or multiple SUPP-- datasets
 - Related Records — [RELREC](#)

However, as described earlier, diseases have unique data elements that must be analyzed leading to different collection mechanisms and the need to define disease specific data standards. The question from users is always, "Where do I put this data?" and this challenge leads to inconsistency in these disease specific standards. While the FDA has stated their support for the current standards they've asked CDISC to increase the speed at which they develop disease specific standards.

NOT A TWO DIMENSIONAL WORLD

The most critical gap in the existing standards is the two dimensional world they live in. The standards define specific data domains with variable names, definitions, and rules but clinical research is much more complex than rows and columns in a SAS data set. Tabular data structures are very limited in the information they can convey. The CDISC models strongly encourage the use of metadata to help define the traceability and transparency of the data. The define.xml is an xml specification which captures the metadata about the data submitted to the FDA. However, the xml and the SAS data sets are technically disconnected from each other. Therefore, the process of defining and managing metadata is very manual and prone to errors and inconsistent information.

In addition to the need for more tightly connected metadata and data, the standards must provide a better mechanism for tying together data around a patient instead of data associated with a domain. For example, currently Adverse Events are collected as a single domain within SDTM. However, a clinician wants to understand the complex relationships across multiple clinical endpoints within a patient, with Adverse Event being just one endpoint. There needs to be a better way to define metadata about the data in a much more transparent and hierarchical way so more dynamic relationships can be described.

In recent years the FDA has discussed the idea of moving the clinical research data standards into a more robust HL7 xml standard which is currently used for health care systems and electronic health records. However, over the last year, large gaps have been identified in attempting to move the current standards to this model. While this will continue to be investigated there is no timeline for an implementation. In lieu of a drastic shift to a standard such as HL7, the increased adoption of the ODM xml specification for transferring data between systems might provide a more realistic opportunity to shift towards a hierarchical data standard and tightly integrate metadata and data.

ROLE OF SAS

At the beginning of 2009 SAS put a new focus on developing solutions to support clinical data standards and transformations.

After years of attempting to develop tools to support clinical data standards, SAS has developed what appears to be a robust framework within BASE SAS to support the management of clinical data standards. The SAS Clinical Standards Toolkit is a framework of SAS macros, metadata, and configuration files including a representation of the SDTM metadata (Figure 4), a large set of validation checks, and the ability to create define.xml for submissions.

Figure 4. Study Metadata within Clinical Toolkit

The figure shows two overlapping SAS window screenshots. The top window, titled 'VIEWTABLE: Source Table Metadata', displays a table with columns: Table Name, Table Label, Observation Class within Standard, (Relative) path to xpt file, and Title for xpt file. The bottom window, titled 'VIEWTABLE: Srcmeta.Source_columns', displays a table with columns: SASreferences sourcedata libref, Table Name, Column Name, Column Description, Column Order, Column Type, Column Length, Column Required or Optional, Column Origin, and Column Role.

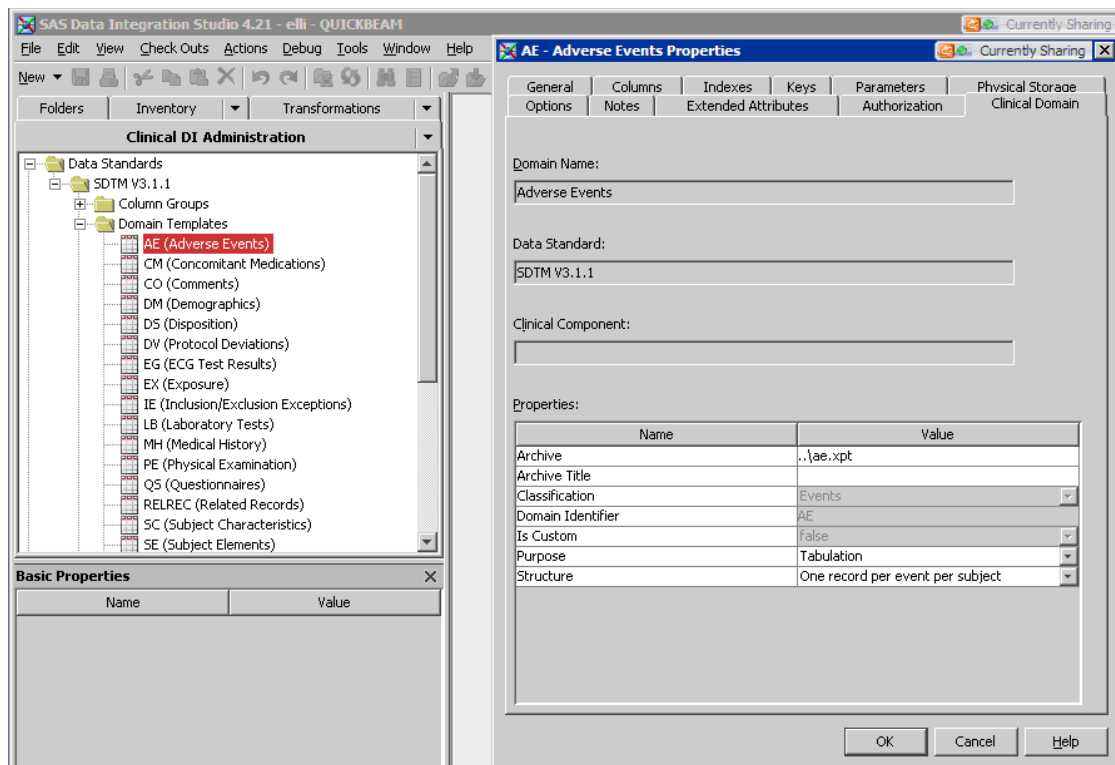
Table Name	Table Label	Observation Class within Standard	(Relative) path to xpt file	Title for xpt file
DM	Demographics	Special Purpose	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Demographics SAS transport
DS	Disposition	Events	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Disposition SAS transport file
DV	Protocol Deviations	Events	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Protocol Deviations SAS tran
IE	Inclusion/Exclusion Exceptions	Findings	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Inclusion/Exclusion Exceptio transport file
LB	Laboratory Tests	Findings	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Laboratory Tests SAS transport
MH	Medical History	Events	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Medical History SAS transport
PF	Pulmonary Function	Findings	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Pulmonary Function SAS tran
RELREC	Related Records	Relates	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Related Records SAS transport file
SUPPAE	Supplemental Qualifiers - AE	Relates	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Supplemental Qualifiers - AE transport file
SUPPALL	Supplemental Qualifiers - ALL	Relates	..\cdisc-sdtm-3.1.1\sample\cdisc-sdtm-3	Supplemental Qualifiers - ALL transport file

SASreferences sourcedata libref	Table Name	Column Name	Column Description	Column Order	Column Type	Column Length	Column Required or Optional	Column Origin	Column Role	
1	SRCDATA	AE	STUDYID	Study Identifier	1	C	40	Req	CRF	Identifier
2	SRCDATA	AE	DOMAIN	Domain Abbreviation	2	C	6	Req	Derived	Identifier
3	SRCDATA	AE	USUBJID	Unique Subject Identifier	3	C	40	Req	Sponsor Defined	Identifier
4	SRCDATA	AE	AESSEQ	Sequence Number	4	N	8	Req	CRF or Derived	Identifier
5	SRCDATA	AE	AEGRPID	Group ID	5	C	40	Perm	Sponsor Defined	Identifier
6	SRCDATA	AE	AEREFID	Reference ID	6	C	40	Perm	Sponsor Defined	Identifier
7	SRCDATA	AE	AESPID	Sponsor-Defined Identifier	7	C	40	Perm	Sponsor Defined	Identifier
8	SRCDATA	AE	AETERM	Reported Term for the Adverse Event	8	C	200	Req	CRF	Topic
9	SRCDATA	AE	AEMODIFY	Modified Reported Term	9	C	200	Perm	Sponsor Defined	Synonym Qualifier
10	SRCDATA	AE	AEDECOD	Dictionary-Derived Term	10	C	200	Req	Derived	Synonym Qualifier
11	SRCDATA	AE	AECAT	Category for Adverse Event	11	C	40	Perm	Sponsor Defined	Grouping Qualifier
12	SRCDATA	AE	AESCAT	Subcategory for Adverse Event	12	C	40	Perm	Sponsor Defined	Grouping Qualifier
13	SRCDATA	AE	AEOCCUR	Adverse Event Occurrence	13	C	1	Perm	CRF or Sponsor Defined	Record Qualifier
14	SRCDATA	AE	AEBODSYS	Body System or Organ Class	14	C	200	Exp	CRF or Derived	Record Qualifier
15	SRCDATA	AE	AELOC	Location of the Reaction	15	C	40	Perm	CRF or Derived	Record Qualifier
16	SRCDATA	AE	AESEV	Severity/Intensiv	16	C	40	Perm	CRF	Record Qualifier

The initial version of the Toolkit implements the standard JANUS and webSDM validation checks and provides the ability to generate the define.xml based on a SAS representation of the metadata necessary within this specification. By providing the Toolkit with a set of tools that are familiar to SAS programmers (e.g. macros and SAS data sets), SAS has finally delivered a viable solution for managing clinical standards.

Based on the traditional SAS tools for data integration, SAS has developed the Clinical Data Integration product, a solution to support the management of standards as well as the transformation of clinical data to a standard. Clinical Data Integration is a set of plug-ins sitting on top of SAS Data Integration Studio (Figure 5) which is a traditional ETL solution. The added capabilities contain specific functionality relevant to the clinical transformation process including the management of versions of standards, creating study specific components, creating customer specific standards, building SDTM custom domains, and reporting of the standards used. In addition, the solution uses the Toolkit described above under the covers to run validation checks and create the define.xml.

Figure 5. SAS Data Clinical Data Integration



With the SAS Clinical Standards Toolkit and SAS Data Integration solutions, SAS appears to be headed in the right direction with supporting the needs within the industry. They still face the challenge of the traditional SAS programmer who just wants to write code, but the gap is closing as efficiencies become more apparent with the use of these tools.

THE FUTURE

In recent years the FDA and other government organizations working with clinical data have seen the critical need for more robust data standards which, in the long run, will lead to better and more efficient science. However, in order for this to be realized people have to adopt standards, use standards, and continue to evolve to make the standards better. The iterative process of changing can be very painful but innovation does not usually come without pain.

In the Christmas Carol, the Ghost of Christmas Future shows Ebenezer Scrooge what the future will hold if he doesn't change his ways – a quiet demise with nothing but a nonexistent legacy. The Ghost of Clinical Standards Future would provide us a similar message. We must continue to change and adopt to ensure we deliver data standards that leave a legacy of streamlining this process and bringing drugs to market in record time.

MOVE TO A THREE DIMENSIONAL WORLD

As mentioned earlier, there are limitations in defining data standards in a two dimensional world. The data and metadata must better define the complex interdependent relationships between clinical research data which cannot completely be captured in the existing data standards.

The FDA has indicated the need to move to a more robust XML standard such as HL7 that supposedly would provide the ability to define these complex relationships across data and metadata. However, the current HL7 model is designed to handle a single point in time and does not support either the relationships between different clinical trial domains within a patient as well as the need to capture the traceability of derived data. Recently, there has been a push to move in this direction with a deadline of 2013 for the adoption of an HL7 message and the elimination of the SAS transport file. However, because of the backlash from industry the FDA has backed off this message and has indicated there is no timeline for this implementation. They will take their time and develop an alternative that works and can be easily adopted by industry.

Even though the short term seems to support use of the standards as they exist today, the industry cannot deny the need to move to a three dimensional standard if they expect to realize rapid efficiencies. This leads to many challenges in the future as to how data standards will evolve to meet the needs of both clinical research and regulatory agencies.

CONTINUED ADOPTION OF STANDARDS

While the current standards have limitations the industry must continue to work towards adopting the standards in their process even if it doesn't lead to immediate efficiencies in the short term. By jumping full throttle into the standards we can learn where the gaps are and work harder to close those gaps. This is easy to recommend in theory but leads to challenges as companies are under more pressure every day to get drugs submitted fast.

In the future, standards can be adopted more smoothly if the industry works harder at incorporating them earlier in the process. As CDASH matures we can work on collecting the data in a standard and thus make everything else downstream much easier since the standards are aligned. The standards can even go back further to the development of the protocol with the CDISC release of the Protocol Representation 1.0 Model which not only provides a standard for collecting metadata about a Protocol but was also developed with a three dimensional world in mind. By iteratively following this lifecycle of clinical data standards in the future (Figure 6) and improving the steps as we go along, standards will become an integral part of the process instead of a necessary evil.

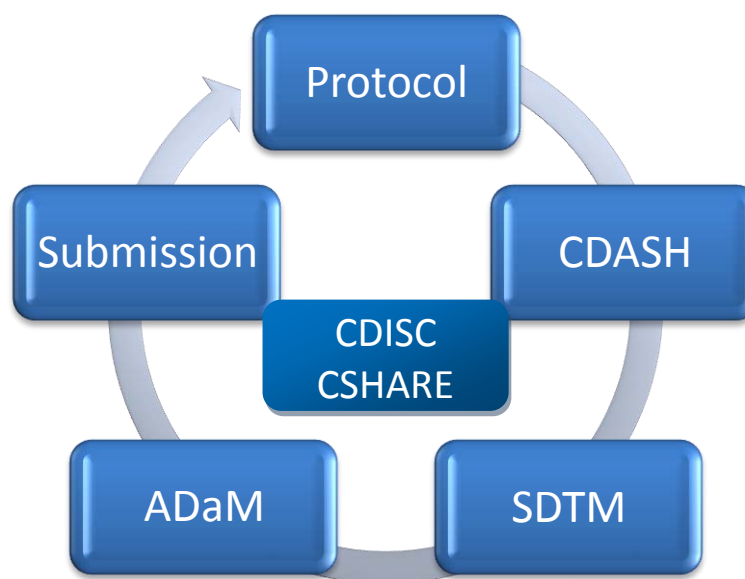


Figure 6. Life Cycle of CDISC Standards

In addition to the standards mentioned earlier, CDISC initiated the CSHARE project this year, a new and innovative project to improve the standards development process. In the past each of the standards defined by CDISC were developed in a silo and communicated through word documents. This led to a lot of inconsistency and repetition across models and slowed down the development of new standards. The goal of CDISC SHARE is to create a

global, accessible, electronic library, which through advanced technology, enables precise and standardized data element definitions that can be used within applications and across studies to improve biomedical research and its link with healthcare. This project has the potential to provide much needed consistency and more rapid standards development.

CROSSROADS OF CLINICAL RESEARCH AND HEALTH RECORDS

Over the last decade clinical researchers have always had this dream to access data held within electronic health records (EHRs) at hospitals, doctors, and medical research centers. On the flip side, site clinicians who participate in clinical trials have to deal with the cumbersome process of entering data multiple times.

The CDISC Healthcare Link project began in 2005 and focuses on the mission of interoperability between healthcare and clinical research. CDISC used a concept called Retrieve Form for Data-capture (RFD), which provides the ability for clinicians to access interfaces for entering their data into the electronic medical records (EMR) system and having the information populate the data elements required by clinical trials. This has many benefits including improvement in data quality, timeliness of data, and alleviating the pain researchers find when entering data multiple times.

This is an example of using the data standards to improve efficiencies and is yet another example of how these standards will be used in the future.

ROLE OF SAS

SAS seems to be heading in the right direction with developing tools that support the needs of industry. However, they need to understand the complexities of the data collected so they don't become relegated to just analysis within the clinical research. They need to understand that clinical data and its associated metadata cannot be captured in a rows and columns two dimensional world. As the standards continue to develop into something that is more dynamic and 'three dimensional' SAS must build tools that leverage those new standards.

CONCLUSION

The Ghost of Clinical Standards past, present and future has provided you with a whirlwind journey through the history of clinical data standards including the challenges, progress, and future hopes.

The Ghost of Clinical Standards Past described the challenges of working with clinical data as technology was introduced into the drug development process. This included the inconsistency in the data across studies and the need to reinvent the wheel every time a new study was initiated. He gave us hope by discussing the birth of CDISC and its potential to solve all the pains of the clinical data world. Finally, he shared with us the previous history of SAS' attempt at building solutions specific to clinical data standards and their mixed results.

The Ghost of Clinical Standards Present provided the current state of affairs. He described the gaps that still exist in the regulatory review process and how the FDA has not provided clear direction regarding their needs for reviewing submissions. He also explained that while the current standards are a step in the right direction they still have adoption barriers including the challenge of using the standards in a company's day to day process as well as the disconnect between the data and metadata. Finally, he presented an overview of the promising new SAS solutions for working with clinical data standards including the SAS Clinical Standards Toolkit and SAS Clinical Data Integration.

The Ghost of Clinical Standards Future gave us a glimpse into what the future of clinical data standards might hold. The industry must continue to work harder at adopting the standards earlier in their process and seeing it through the entire workflow while the FDA must work harder at providing a clear direction for their expectations. He also explained the need to move towards a three dimensional world to better describe the complexities within clinical research and the eventual merging of electronic health records and clinical trial data. Finally, he provided a challenge for SAS to continue to adapt to the changing standards and realize rows and columns are not going to suffice in the long term.

At the end of the Christmas Carol, Ebenezer Scrooge realizes the error of his ways, and pleads to the Ghost of Christmas Future: "I will honor Christmas in my heart, and try to keep it all the year. I will live in the Past, the Present, and the Future. The Spirits of all Three shall strive within me. I will not shut out the lessons that they teach. Oh, tell me if I may sponge away the writing on this stone!" If the clinical research industry does not continue to adapt and change and get out of the world of rows and columns we will continue to repeat past history and not realize the efficiencies absolutely necessary to improve the drug development process and thus save more lives.

I will leave everyone with a question: "Are these the shadows of the things that Will be, or are they shadows of things that May be?"

REFERENCES

- 1) Cooper, Chuck, M.D., Office of Translational Sciences, CDER, FDA. "Computational Science and Data Standards in CDER", 8th Annual DIA eCTD Conference, 2009.

RECOMMENDED READING

- CDISC Data Standards – <http://www.cdisc.org>
- HL7 Data Standards – <http://www.hl7.org>
- SAS Solutions for Life Sciences – <http://www.sas.com/industry/pharma/index.html>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Chris Decker
Enterprise: d-Wise Technologies
Address: 4020 Westchase Blvd Suite 527
City, State ZIP: Raleigh, NC 27607
Work Phone: 919-600-6234
E-mail: cdecker@d-wise.com
Web: www.d-wise.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.