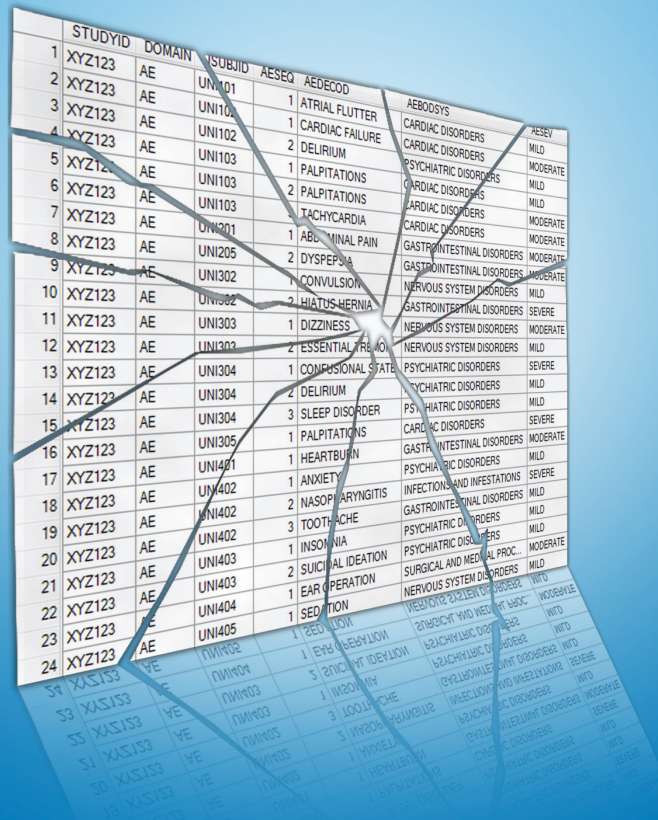


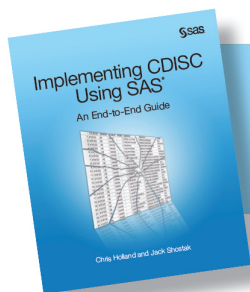
Implementing CDISC Using SAS[®]

An End-to-End Guide



	STUDYID	DOMAIN	SUBJID	ASEQ	AEDCOD	AEBODSYS	AEBODSYS	AESEV
1	XYZ123	AE	UNI001	1	ATRIAL FLUTTER	CARDIAC DISORDERS	MODERATE	
2	XYZ123	AE	UNI101	1	CARDIAC FAILURE	CARDIAC DISORDERS	MILD	
3	XYZ123	AE	UNI102	2	DELIRIUM	PSYCHIATRIC DISORDERS	MODERATE	
4	XYZ123	AE	UNI103	1	PALPITATIONS	CARDIAC DISORDERS	MILD	
5	XYZ123	AE	UNI103	2	PALPITATIONS	CARDIAC DISORDERS	MILD	
6	XYZ123	AE	UNI103	2	PALPITATIONS	CARDIAC DISORDERS	MILD	
7	XYZ123	AE	UNI201	1	TACHYCARDIA	CARDIAC DISORDERS	MODERATE	
8	XYZ123	AE	UNI201	1	ABDOMINAL PAIN	GASTROINTESTINAL DISORDERS	MODERATE	
9	XYZ123	AE	UNI205	2	DYSPEPSIA	GASTROINTESTINAL DISORDERS	MODERATE	
10	XYZ123	AE	UNI302	1	CONVULSION	NERVOUS SYSTEM DISORDERS	MILD	
11	XYZ123	AE	UNI302	2	HIATUS HERNIA	GASTROINTESTINAL DISORDERS	SEVERE	
12	XYZ123	AE	UNI303	1	DIZZINESS	NERVOUS SYSTEM DISORDERS	MODERATE	
13	XYZ123	AE	UNI303	2	ESSENTIAL TREMOR	NERVOUS SYSTEM DISORDERS	MILD	
14	XYZ123	AE	UNI304	1	CONFUSIONAL STATE	PSYCHIATRIC DISORDERS	SEVERE	
15	XYZ123	AE	UNI304	2	DELIRIUM	PSYCHIATRIC DISORDERS	MILD	
16	XYZ123	AE	UNI304	3	SLEEP DISORDER	PSYCHIATRIC DISORDERS	SEVERE	
17	XYZ123	AE	UNI305	1	PALPITATIONS	CARDIAC DISORDERS	MODERATE	
18	XYZ123	AE	UNI401	1	HEARTBURN	GASTROINTESTINAL DISORDERS	MILD	
19	XYZ123	AE	UNI401	1	ANXIETY	PSYCHIATRIC DISORDERS	SEVERE	
20	XYZ123	AE	UNI402	2	NASOPHARYNGITIS	INFECTIONS AND INFESTATIONS	SEVERE	
21	XYZ123	AE	UNI402	2	TOOTHACHE	GASTROINTESTINAL DISORDERS	MILD	
22	XYZ123	AE	UNI402	3	TOOTHACHE	PSYCHIATRIC DISORDERS	MILD	
23	XYZ123	AE	UNI403	1	INSOMNIA	PSYCHIATRIC DISORDERS	MODERATE	
24	XYZ123	AE	UNI403	2	SUICIDAL IDEATION	PSYCHIATRIC DISORDERS	MODERATE	
25	XYZ123	AE	UNI403	2	SUICIDAL IDEATION	SURGICAL AND MEDICAL PROC.	MILD	
26	XYZ123	AE	UNI403	1	EAR OPERATION	NERVOUS SYSTEM DISORDERS	MILD	
27	XYZ123	AE	UNI404	1	SEDATION	NERVOUS SYSTEM DISORDERS	MILD	
28	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
29	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
30	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
31	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
32	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
33	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
34	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
35	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
36	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
37	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
38	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
39	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
40	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
41	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
42	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
43	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
44	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
45	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
46	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
47	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
48	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
49	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	
50	XYZ123	AE	UNI405	1	DEPRESSION	NERVOUS SYSTEM DISORDERS	MILD	

Chris Holland and Jack Shostak



From *Implementing CDISC Using SAS[®]*. Full book available for purchase [here](#).

Contents

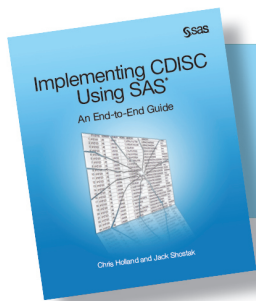
About This Book	xi
About The Authors.....	xvii
Acknowledgments	xix
Chapter 1: Implementation Strategies	1
Why “CDISC It”?	1
Which Models to Use and Where	4
Starting with the Clinical Data Acquisition Standards Harmonization (CDASH) Standard	5
Implementation Plans and the Need for Governance	5
SDTM Considerations	6
ADaM Considerations	8
Chapter Summary	11
Chapter 2: SDTM Metadata and Define.xml for Base SAS Implementation	13
SDTM Metadata	14
Table of Contents Metadata	15
Variable-Level Metadata	16
Codelist Metadata	19
Value-Level Metadata	22
Computational Method Metadata	24
Building define.xml	25
Define File Header Metadata	26
Define File Creation SAS Program	26
Chapter Summary	27
Chapter 3: Implementing the CDISC SDTM with Base SAS	29
Base SAS Macros and Tools for SDTM Conversions	30
Creating an SDTM Codelist SAS Format Catalog	31

Creating an Empty SDTM Domain Dataset	32
Creating an SDTM --DTC Date Variable	35
Creating an SDTM Study Day Variable.....	36
Sorting the Final SDTM Domain Dataset	37
Building SDTM Datasets.....	38
Building the Special-Purpose DM and SUPPDM Domains	38
Building the LB Findings Domain	42
Building a Custom XP Findings Domain	45
Building the AE Events Domain	47
Building the EX Exposure Interventions Domain	49
Building Trial Design Model (TDM) Domains.....	51
Chapter Summary	55
Chapter 4: Implementing CDISC SDTM with the SAS Clinical Standards Toolkit and Base SAS.....	57
SAS Clinical Standards Toolkit Background.....	57
Clinical Standards Setup for Study XYZ123	58
Building SDTM Datasets.....	61
Base SAS Macros and Tools for SDTM Conversions.....	61
Building the Special-Purpose DM and SUPPDM Domains.....	63
Building define.xml.....	66
Chapter Summary	71
Chapter 5: Implementing the CDISC SDTM with SAS Clinical Data Integration.....	73
SAS Clinical Data Integration Introduction.....	74
SAS Clinical Data Integration Metadata	74
Classifications of SAS Clinical Data Integration Metadata.....	74
Setup of SAS Clinical Data Integration Metadata	75
SAS Clinical Data Integration Study Setup.....	78
Define the Clinical Study and Subfolders	79
Register Source Datasets and Define Target SDTM Datasets.....	82
Setting SAS Clinical Data Integration Defaults	87
Creating SDTM Domains	89
Creating the Special-Purpose DM and SUPPDM Domain	89
Creating the AE (Adverse Events) Events Domain.....	97
Creating the XP Pain Scale Customized Findings Domain.....	99
Creating the EX Exposure Interventions Domain	100
Creating the LB Laboratory Findings Domain.....	100

Creating the Trial Design Model Domains	101
Using Customized Code in SDTM Production	101
Templating Your SDTM Conversion Jobs for Reuse	102
Using SAS Clinical Data Integration to Create define.xml	103
Chapter Summary	106
Chapter 6: ADaM Metadata and ADaM Define.xml.....	107
Metadata Spreadsheets	108
Variable Metadata in ADaM.....	108
Analysis Parameter Value-Level Metadata	109
Analysis Results Metadata	110
Building define.xml.....	113
Define.xml Navigation and Rendering.....	113
Chapter Summary	116
Chapter 7: Implementing ADaM	117
ADaM Tools.....	118
ISO 8601 Date and DateTime Conversions.....	118
Merging In Supplemental Qualifiers	120
ADSL – The Subject-Level Dataset	125
The ADaM Basic Data Structure (BDS)	128
ADAE – Adverse Event Analysis Datasets	131
ADTTE – The Time-to-Event Analysis Dataset.....	133
Chapter Summary	136
Chapter 8: SDTM Validation	139
SAS Clinical Standards Toolkit	140
SAS Clinical Standards Toolkit Setup	140
SAS Clinical Standards Toolkit Validation Program	142
SAS Clinical Data Integration.....	146
OpenCDISC Validator	150
Installing OpenCDISC Validator	151
Running OpenCDISC Validator (Graphical User Interface).....	151
Evaluating the Report File and Report Results	152
Modifying the Configuration Files.....	156
Running OpenCDISC Validator in Command Line Mode	158
Chapter Summary	162

Chapter 9: Validating ADaM Data	163
ADaM’s Published Validation Checks	163
OpenCDISC Validator	164
Running the OpenCDISC Validator as a Macro	165
Traceability Checks	167
Chapter Summary	170
Chapter 10: CDISC Data Review and Analysis	171
Safety Evaluations with JMP Clinical	172
Getting Started with JMP Clinical.....	172
Safety Analyses	176
Patient Profiles	180
Customizing JMP Clinical.....	181
One PROC Away with ADaM Datasets	183
Chapter Summary	185
Chapter 11: Integrated Data and Regulatory Submissions	187
Regulatory Guidance	188
Data Integration Challenges	188
Data Integration Strategies	189
Representing Subjects Who Appear in Multiple Studies in Subject-Level Datasets	190
Deciding on Which Data to Integrate	193
Coding Dictionary Issues.....	194
Summary of Data Integration Strategies	195
Data Integration and Submission Tools.....	195
Setting Variable Lengths Based on the Longest Observed Value.....	196
Converting from Native SAS to Version 5.0 Transport Files	199
Converting from Version 5.0 Transport Files to Native SAS	200
Getting Submission Ready	200
Chapter Summary	201
Chapter 12: Other Topics	203
Standard for Exchange of Non-Clinical Data	203
HL7 Messaging	204
BRIDG Model	205
Protocol Representation Model.....	205

FDA Janus Clinical Trials Repository	206
CDISC Model Versioning	206
Future CDISC Directions	207
Chapter Summary	208
Appendix A: Source Data Programs	209
adverse Dataset	210
demographics Dataset	211
dosing Dataset.....	212
laboratory Dataset	215
pain scores Dataset	221
Appendix B: SDTM Metadata.....	223
Appendix B.1 - Table of Contents Metadata	223
Appendix B.2 - Variable-Level Metadata	224
Appendix B.3 - Codelist Metadata.....	226
Appendix B.4 - Value-Level Metadata.....	228
Appendix B.5 - Computational Method Metadata	228
Appendix B.6 - Define Header Metadata	228
Appendix C: ADaM Metadata	229
Appendix C.1 - Define Header Metadata	229
Appendix C.2 - Table of Contents Metadata	230
Appendix C.3 - Variable-Level Metadata	230
Appendix C.4 - Parameter-Level Metadata	232
Appendix C.5 - Computational Method Metadata	232
Appendix C.6 - Codelist Metadata	232
Appendix C.7 - Analysis Results Metadata.....	233
Appendix C.8 - External Links Metadata	234
Appendix D: %make_define SAS Macro.....	235
%make_define SAS Macro.....	235
Additional SAS Code Comments.....	249
Index.....	253



From *Implementing CDISC Using SAS[®]*. Full book available for purchase [here](#).

Chapter 1: Implementation Strategies

Why “CDISC It”?	1
Which Models to Use and Where	4
Starting with the Clinical Data Acquisition Standards Harmonization (CDASH) Standard	5
Implementation Plans and the Need for Governance	5
SDTM Considerations	6
ADaM Considerations	8
Chapter Summary	11

Why “CDISC It”?

The decision to adapt to CDISC standards within an organization or for a particular clinical development program is not an easy one. Many large pharmaceutical companies have their own internal standards and tools built around those standards. For these companies, the decision about if and when to switch to CDISC was, and in many cases still is, fodder for heated debates. Upper management may cringe and balk at proposals to implement CDISC standards when they see the up-front price tag associated with it. Some organizations got burned by trying to implement the Study Data Tabulation Model (SDTM) too early, before FDA reviewers were prepared for it and before the standard had stabilized to a point where it was backwards compatible. Without an FDA mandate, the rationale for taking a wait-and-see approach seemed to have its merits.

Despite these concerns, set-backs, and growing pains for CDISC, the groundswell of support continued to rise. Mergers and acquisitions have persisted throughout the pharmaceutical industry, and behind the scenes of each merger are the data managers and SAS programmers who’ve worked at the same desk year after year, but have seen their employer name change three to four times. Throughout it all, with a change in the employer came the change in the case report form (CRF) designs, variable names, and data formats for the different compounds on which they worked. When it came time to integrate the data for a regulatory submission, a substantial amount of time was spent deciding on the structure and variable names to be used for the integrated database. And that was just the beginning. The time spent doing the actual conversions and integration is often much greater. As the programming hours piled up, those involved started to see the merits of having a standard across the industry.

2 Implementing CDISC Using SAS: An End-to-End Guide

Pharmaceutical and biotech companies weren't the only organizations undergoing mergers. During the late 1990s and early 2000s, many CROs consolidated as well. In addition to the numerous data standards they had to keep track of among their various clients, CRO SAS programmers also had to deal with different data formats being used internally due to consolidation with other CROs. Some at these CROs got to work on integration projects involving compounds that, at each new phase of development, had been passed from one organization and CRO to another. As a result, even the most basic key identifier of any clinical trial dataset, the subject ID, was sometimes uniquely named within each study. So as the programming hours piled up, the key decision makers at CROs started to see the merits of having a data standard across the industry.

Yet this grass-roots initiative to develop industry-wide standards would not have gotten off the ground without the support of the biggest consumer of clinical trial data of all, the US Food and Drug Administration. Although some FDA reviewers become accustomed to data formats they see from certain sponsors, most still have to deal with completely different data formats and structures from one sponsor to the next. This might not have been so cumbersome in the days before the Prescription Drug User Fee Act (PDUFA, commonly pronounced *puh-DOO-fa*) first became effective. Before PDUFA, a review clock was non-existent and 2-year reviews of New Drug Applications (NDAs) and Biologic License Applications (BLAs) were the norm. However, with the passage of PDUFA, review cycles were originally mandated to be 12 months (and are now down to 10 months). With those review clocks, along with increasing expectations to carefully inspect the electronic data that were packaged with NDA and BLA submissions, reviewers found themselves having to do more with less.

The aftermath of some pivotal events in 2004 put even more pressure on FDA reviewers. One was the investigation of suicidality risks among children on antidepressants. The other was the withdrawal of Vioxx from the market. Because of these two high-profile safety concerns, doctors, patients, and sponsors all suddenly had a vested interest in knowing whether the drugs they were prescribing, taking, or selling to treat depression, arthritis, or any number of spin-off indications were adding on risks that outweighed the benefits. The brunt of these class-effect determinations fell on the FDA clinical and statistical reviewers who were the only ones who had access to all the data that would allow them to make the informed decisions that the public, doctors, industry, congressmen, and media were all suddenly demanding. However, when the drug class under consideration involved 10 different compounds from as many different sponsors with as many different data formats, this was no easy task. "Wouldn't it be great," some FDA reviewers asked, "if we had all the data we need in one giant database?" Fortunately, within the FDA, certain reviewers, team leaders, and division directors all started to see the merits of having a data standard across the industry. To coin a common phrase of one particular FDA division director who had a penchant for promoting data standards at industry conferences, the mantra of the late 2000s became "just CDISC-It."

Evidence of FDA's support of data standards is not only found at conference podiums. Since the draft release of version 3.1 of the SDTM Implementation Guide (IG) in October, 2003, the FDA has issued a number of documents indicating their support of data standards. A summary of these appear in Table 1.1.

Table 1.1: FDA Documents and Events in Support of Data Standards

Time	Event
July, 2004	eCTD study data specifications reference the SDTM for tabulation data
March, 2006	“Development of data standards” listed as opportunity #44 in the FDA’s Critical Path opportunities list
September, 2006	SDTM/ADaM pilot project review completed and results presented at the CDISC Interchange
September, 2006	Old e-NDA guidance document is withdrawn (leaving the eCTD study data specifications as the only guidance relating to submission data)
December, 2006	Proposed rule to require electronic data with submissions is released in the Federal Register
May, 2008	First version of the PDUFA IV IT plan is released, making numerous commitments to the SDTM
October, 2009	Version 1.5 of the study data specifications released, making specific reference to the Analysis Data Model (ADaM) standard for analysis data
March, 2010	Version 1.0 of the CDER Data Standards Plan released, providing a commitment to CDISC standards
May, 2011	The “CDER Common Data Standards Issues Document,” version 1.0 is released, stating that CDER is “strongly encouraging sponsors to submit data in standard form”
February, 2012	Draft FDA guidance document <i>Providing Regulatory Submissions in Electronic Format -- Standardized Study Data</i> (a.k.a. “the e-Data Guidance”) released to establish “FDA’s recommendation that sponsors and applicants submit study data in a standardized electronic format”

Nowadays, the legion of CDISC implementers is tangible to any attendee of a SAS user conference struggling to find an empty chair in a session that has anything to do with CDISC. Managers are preaching the data standards gospel, software vendors are demonstrating their tools that make use of CDISC data, FDA presenters are promoting their preference for CDISC, and FDA documents are recommending the SDTM and ADaM models as the format to use for sponsors’ NDA and BLA submissions.

Despite all this, electronic data submissions are still (believe it or not) optional. Many compounds struggle to get beyond Phase I of development. Implementers have struggled with different interpretations of the standards and with study-specific data that don’t seem to have a clear-cut home in the SDTM IG. So while the decision to “CDISC-It” for a regulatory submission has become more clear-cut, the stage when the CDISC conversion takes place, and how far to go with it, is less lucid. Study phase, organizational size, standards know-how, and the amount of resources all factor in to the decision-making equation.

Which Models to Use and Where

The advantages of having universal data standards are largely geared toward users of the data for review or analysis. Across studies, medical and statistical data reviewers and analysts, whether they are on the sponsor side of the equation or on the regulatory review side, are starting to benefit from having nearly instant familiarity with how data is organized for any given study. This holds true whether the data are non-clinical SEND data, SDTM tabulation data, or ADaM analysis files.

However, for those who are responsible for putting the data in these standardized formats, there is much more work involved. Before data standards, there was often just one set of data provided with NDA and BLA submissions. Many of these datasets tended to be a hybrid between the raw CRF data and the analysis files. The structures and variable names often matched those of the original database tables and therefore required little manipulation. Now, not only do you have to worry about a potentially labor-intensive conversion process from the raw data tables to the SDTM domains, there is also the need to then create ADaM datasets from the SDTM domains.

For organizations with plenty of resources to devote to the implementation of standards, this process might be manageable. CROs who conduct a high volume of conversions for their clients have an opportunity to streamline their implementation process with each new iteration. Certain technologically advanced organizations such as software companies with proprietary electronic data capture (EDC) systems and expert knowledge of the data standards are capable of developing automated tools to assist with the conversion process from the raw CRF data to fully compliant SDTM domains and subsequent ADaM data files.

For those organizations with high volume and adequate resources, the question isn't always whether they are capable of implementing the SDTM and ADaM, it's more a question of whether the effort is worth it given the phase of development (the earlier the phase, the less likely a drug will ever advance to submission). For everybody else, the wait-and-see approach might be more appealing, given their lack of expertise and resources.

Eventually, however, certain medical products will advance in development and when they do, it is better to be prepared. As such, the objectives of this book are to provide the following:

- Considerations for deciding when to start implementing CDISC standards
- Advice for how to get started with CDISC implementation and how to move forward with it
- Tools based on SAS software to assist with the creation of CDISC data and metadata documentation (and instructions on how to use them)
- Information on how to check CDISC data for compliance
- Information about tools for using CDISC data for analysis and review

Starting with the Clinical Data Acquisition Standards Harmonization (CDASH) Standard

The best way to adapt to being a CDISC organization is to start implementing standards at the initial step of data acquisition—the CRFs. The Clinical Data Acquisition Standards Harmonization (CDASH) (<http://www.cdisc.org/cdash>) standard was created in response to opportunity #45 in FDA’s Critical Path Opportunities list, which was titled “Consensus on Standards for Case Report Forms.” Although part of the initiative was to standardize the look and feel of CRFs, a big part of the initiative in the eyes of CDISC implementers was to standardize the variable names of data elements being captured in the clinical database. Having such a standard that was consistent with SDTM terminology would make the conversion to SDTM much easier. With a CDASH structure behind any data management system, certain SDTM domains, like adverse events (AE), demographics (DM), and concomitant medications (CM), are almost instantly SDTM-ready with the initial data extract to SAS datasets. In total, 16 domains are covered by the CDASH standard, covering those that are common to most therapeutic areas and types of clinical research. For further reading, the CDASH-published document (version 1.1 was published in January, 2011) contains implementation recommendations, best practices, regulatory references, and loads of other information pertinent to CDASH.

For any organization starting from the ground up, implementing CDASH should be an easy decision because it precludes the need to develop a new organization-specific standard. However, unless the data management system happens to come pre-packaged with CDASH default templates, implementing an existing standard can still require a lot of work. Without these templates, one important element to a successful implementation is making sure the proper know-how is put to work. Just a basic knowledge of CDASH might not be enough. Having a breadth of knowledge that spans CDASH, SDTM, and ADaM can help prevent you from, for example, having variable names in the source data that conflict with variables in the SDTM or ADaM data. A careful deployment with proper attention to downstream standards can save you from unnecessary variable renaming later on.

Larger, older, or high-volume organizations with established data management and EDC systems that were put in place long before CDASH probably wouldn’t benefit as much from an immediate switch to CDASH because of the impact it would have on resources and timelines. For these organizations, a certain amount of creativity has had to go into the thoughts and ideas behind how and when to start with the CDISC implementation process.

Whatever the situation, whether the true source data is from an entirely CDASH environment or from something that resembles nothing of the sort, the source data can be considered just various shades of gray in the eyes of an SDTM implementer. Before delving into the programmatic conversion process, the very important step of mapping out a conversion plan needs to be discussed.

Implementation Plans and the Need for Governance

Before an actual CDISC implementation takes place, whether it is a conversion from CDASH to SDTM or the creation of ADaM data from the SDTM, it is often a good idea to document the precise mapping from one data source to another. The advantages of this are three-fold:

6 *Implementing CDISC Using SAS: An End-to-End Guide*

- it allows the work to be handed off from the planner(s) to programmer(s), thereby obviating the need to have these functions performed by the same individual.
- it provides a plan to the programmers that has been discussed, reviewed, and approved ahead of time. It will also prevent ad-hoc decisions by one programmer conflicting with those of another on the same project.
- it provides a specification that the final work product can be checked against and referred to along the way.

Anyone who has spent much time trying to implement a CDISC standard has probably quickly realized that, despite efforts to the contrary, much of it is subject to interpretation. Consequently, there is a strong likelihood that one person's interpretation is different from another's, and herein lies the foundation for another form of conflict relating to standards—the friction between two or more strong-minded individuals who each have their own interpretation on how to implement.

In order to handle this inevitable problem, many organizations have developed a form of governance where decisions relating to controversial issues are agreed upon by a group of experts. The process by which these issues are presented to and decisions are made by a governing board can vary. The board can either be responsible for reviewing and approving all document specifications developed within an organization or they can only get involved to weigh in on certain issues, especially the overarching ones that are likely to affect all projects.

For smaller organizations, use of a governing board might be unnecessary or impractical. Mapping decisions can either be made by senior personnel or by outside consultants. Whatever the size or status of an organization, in order to avoid conflicts later on, reviewing and approving mapping specifications before the actual work begins can, at the very least, prevent bad decisions from being made simply because they reflect what has already been done.

SDTM Considerations

As mentioned earlier, the decision about how and when to implement the SDTM is not always an easy one. Waiting until a Phase III study is unblinded and a pre-NDA meeting occurs can often mean having to convert a whole lot of data in a short amount of time. On the other hand, converting all studies, starting with the first-in-man Phase I, can mean spending a lot of effort on conversions for studies that might never even get into late Phase II or Phase III trials, where the benefits of SDTM conversions can really pay off.

Organizations struggling with these decisions should consider the following questions:

- Do you have the proper expertise and resources to implement the SDTM?

Proper and compliant implementation is important in order to ensure that tools that depend on standards work properly and users of your data (such as regulatory reviewers) have a pleasant experience doing so. Although the objective of this book is to help make the process easier, it will not teach the subtle details of the SDTM. The best reference for that is the most recent version of the

SDTM IG. It is full of details and instructions that should not be overlooked. For trickier problems, such as how to model data that doesn't seem to have an explicit domain, seek advice from consultants or online experts on any of the various message boards available. But even with the proper expertise, the conversion process can be a tedious one. Make sure you have sufficient resources to conduct a proper implementation.

- Do you have enough studies in the pipeline that would allow for an efficient and steep learning curve if every study were to be converted?

Like everything, practice makes perfect, and the less time you spend between implementations, the less you tend to forget how things were done the last time around. As such, a one-off SDTM conversion will not allow you to fine-tune the process with subsequent iterations.

- Do you have a stable environment to allow automation of certain parts of the conversion process from study to study?

Foundational changes, such as corporate mergers or your EDC vendor going out of business, are hard to prepare for. In some situations, however, you might be able to anticipate the likelihood of having different database designs across studies. If the designs do change, then you'll have trouble building an automated conversion processes across studies. The first conversion will be a learning experience regardless. But with each subsequent conversion, the more similarities there are with the raw CRF data across studies, the more opportunities you will find to make the conversion more efficient, such as using SAS macros or standard programs for converting certain domains.

- Do you plan on using any tools that could make downstream processes such as data cleaning, safety reviews, or analysis more efficient when used on SDTM data?

Certain off-the-shelf tools can make data review, particularly safety data review, easier if the data are SDTM-compliant. If you would like to produce patient profiles or other reports and summaries that review tools that leverage the SDTM can create, then you will certainly benefit from an SDTM conversion. Some of these review tools will be discussed in this book.

- What phase of development are you in?

Many regulatory guidance documents provide advice about how to incorporate safety data into a submission. They tend to differentiate between Phase I safety data from healthy volunteers and those from Phase II and III studies that are more pertinent to the population, dose, and treatment regimen being considered for approval. You must also consider the attrition rate of experimental therapies. Products that eventually make it to a regulatory submission are the exception rather than the norm, and when or if they do make it to submission, not integrating or converting the Phase I data might be an option to consider because such data, aside from potential PK results, PD results, or both, are less relevant to a review of a product's safety and efficacy. Products at later stages of development, however, might reap better rewards as a starting point for implementing the SDTM.

8 *Implementing CDISC Using SAS: An End-to-End Guide*

- Should I consider a staged approach?

Perhaps you or your organization lacks the resources or expertise for a full-blown SDTM conversion. You might still benefit from having certain key domains, such as adverse events, demographics, concomitant medications, and laboratory data in a format that, if not fully SDTM-compliant, is pretty close. Doing so will facilitate the development of standard programs, might be sufficient to use certain tools, and will make a full conversion, if required later on, that much easier. However, keep in mind that an FDA submission will likely require a fully compliant implementation.

ADaM Considerations

The first version of the Analysis Data Model (ADaM) model document was released in final form in December of 2004. It contained general considerations with respect to analysis datasets. Starting in April 2006, the ADaM team began working toward two significant goals:

- To define a standard data structure that would work well for many common analysis needs
- To create an ADaM Implementation Guide

Around the same time, the idea for a mock submission that FDA reviewers could use to see how well both the SDTM and ADaM data standards met their needs for a mock review started to gain some traction. This idea developed into the first SDTM/ADaM pilot project. During the course of a year, volunteers from industry worked feverishly to get this sample submission together, and volunteers from FDA worked diligently to closely evaluate the data, compile their comments, and discuss their findings. The constructive feedback assisted the ADaM team in its work on a new version of the model document and the first-ever implementation guide.

Drafts of the model document (ADaM version 2.1) and the implementation guide (ADaM IG version 1.0) were posted for public comment in May, 2008. Final versions of both documents were published in December, 2009 and serve as the basis for topics relating to ADaM in this book. They can be found on the CDISC web site at <http://www.cdisc.org/adam>.

The significance of this implementation guide should not be understated. Finally, a document existed that would allow industry to standardize the important analysis datasets that statistical and medical reviewers needed to verify sponsor results, assess the robustness of results, and explore trends in the data. Knowing that the format and structures of the data had, to some extent, been previously vetted among FDA reviewers gives implementers a certain degree of confidence that allows them to scrap their company or drug-specific formats for a new CDISC standard that will eventually have wide-spread familiarity.

The ADaM 2.1 model document highlights certain fundamental principles relating to ADaM data. As stated in the document, analysis datasets and their associated metadata must have the following characteristics:

- Facilitate clear and unambiguous communication
- Provide traceability between the analysis data and its source data (ultimately SDTM)
- Be readily usable by commonly available software tools

Further, analysis datasets must have the following characteristics:

- Be accompanied by metadata
- Be analysis-ready

The decision about whether or not to implement ADaM standards within an organization should be an easier and more straightforward one compared to the SDTM decision. First of all, an assumption with ADaM data is that there is corresponding SDTM data to which your ADaM data can be traced. So, the first question you have to ask is whether SDTM source data exists. If so, the next question is how extensive of a study, from an analysis perspective, are you dealing with. The effort to create ADaM data from SDTM data for small, safety- and PK-based Phase I studies should be balanced against a potentially limited benefit. This is because a) analyses for such studies are usually quite basic and b) analysis datasets from such trials are rarely expected for a regulatory submission. It is highly recommended, however, that at least the ADSL dataset be created. This one data set, which includes just one record per subject, is the minimum dataset requirement for an ADaM submission. Even for small, early-phase trials, it can be useful as a single source for capturing flags and certain key information relating to each subject's experience in the trial.

On the other hand, even with SDTM source data from a controlled clinical trial and a plan for inferential statistics, implementing ADaM might not necessarily be a foregone conclusion. One approach, although not recommended, is to instead implement what many refer to as SDTM+. With this approach, variables are added to the SDTM domains to assist with analysis. These variables can include treatment group identifiers, extra flags, conversions of SDTM dates to SAS date variables, and, perhaps, change from baseline calculations. While this approach might be convenient for some, it is not recommended for two primary reasons:

- Variables added to SDTM+ datasets are likely to vary across studies, submissions, and sponsors, and therefore do nothing to promote standards.
- SDTM+ datasets are usually ill-suited for providing the traceability necessary for complex computations and derivations, such as missing value imputations.

As such, the authors of this book recommend using ADaM whenever the following two conditions are true:

- A corresponding set of SDTM data exists.
- Non-trivial derivations, imputations, and/or inferences are calculated according to the statistical analysis plan for the given study.

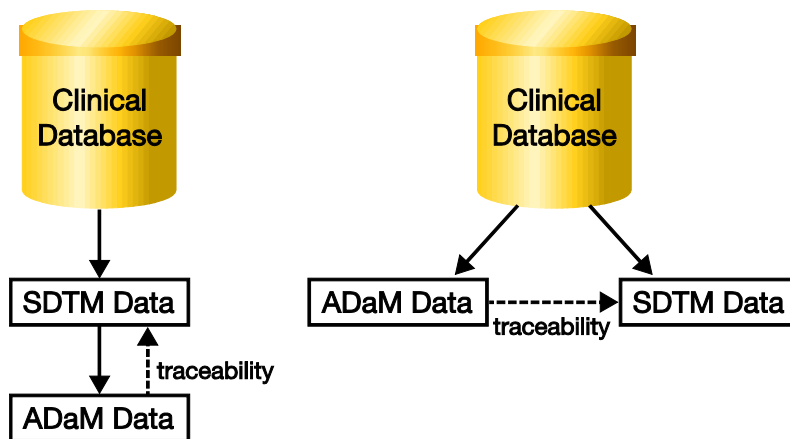
10 Implementing CDISC Using SAS: An End-to-End Guide

Not all analysis data will fit into one of the predefined analysis data structures, such as the ADaM Basic Data Structure (BDS). If you are *not* using the specific data structures mentioned in ADaM documents, you should at least consider adhering to the basic principles for analysis datasets mentioned previously.

The conundrum faced by institutions looking to implement CDISC standards for the first time, particularly for early-phase trials, is that while analysis data of some sort is almost always needed for producing summaries for a clinical study report, and resources and timelines are based upon this assumption, the SDTM data is, in many ways, primarily needed for submissions. As mentioned in the beginning of this chapter, the “old days”, when some organizations went straight from raw CRF data to analysis data, or a hybrid of the two, had that distinct advantage: one less time-consuming step to produce analysis data.

With this in mind, the authors concede one scenario in which you might choose to go directly from raw CRF data to ADaM, without an SDTM conversion in the middle. This one scenario would be when, due to short timelines and limited resources, the only other alternative is to not follow the ADaM standard at all and to not create SDTM data at all. Figure 1 displays both scenarios.

Figure 1.1: Scenarios for Implementing CDISC Standards



Option A (preferred):
Linear conversions from the clinical database to SDTM to ADaM.

Option B (not recommended):
SDTM is created at the same time as or after ADaM. Traceability back to SDTM is done retrospectively.

If you choose Option B, the primary risk you face is that when, or if, SDTM data are created, it could be rather difficult to ensure that records from the pre-existing ADaM datasets can be properly traced to these new SDTM data. In some cases, you might be better off creating ADaM data anew, although this

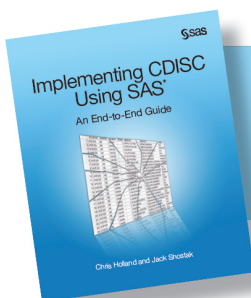
would result in duplication of work. It should also be noted that the so-called ADaM data would not truly be following the model until there were metadata and SDTM data to which they could trace back.

The authors hope that the tools based on SAS software covered in this book and the information on how to use these tools will make conversions to CDISC standards easy enough to balance out with the efficiencies gained by using CDISC data for exploration and analysis. In other words, we hope to provide readers with information and tools that make Option B an unnecessary one to consider.

Chapter Summary

The motivations for having industry-wide data standards are multi-fold. The decision about when and how to adapt to any CDISC standard is, however, a bit more complicated. In this chapter we presented some considerations to assist with these decisions. In this book, we will be providing many tools and examples to make the conversion process easier.

From *Implementing CDISC Using SAS®: An End-to-End Guide* by Chris Holland and Jack Shostak. Copyright © 2012, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



From *Implementing CDISC Using SAS®*. Full book available for purchase [here](#).

Index

A

ABLFL variable 130
ADAE (Adverse Event Analysis datasets) 131–132
ADaM (Analysis Data Model)
 about 3, 4, 117
 ADAE (Adverse Event Analysis dataset) 131–132
 ADSL (Subject-Level Analysis dataset) 125–128, 176, 190–192
 ADTTE (Time-to-Event Analysis dataset) 133–136
 Analysis Results metadata 110–113, 233
 Basic Data Structure (BDS) 10, 128–130
 Codelist metadata 232
 Computational Method metadata 231
 considerations 8–11
 datasets 183–185
 Define Header metadata 229
 define.xml 107–108
 External Links metadata 233
 metadata 107–108
 OpenCDISC Validator 150–167
 Parameter-Level metadata 231
 published validation checks 163
 spreadsheets 108–113
 Table of Contents metadata 229
 tools 118–124
 traceability checks 167–170
 Variable-Level metadata 108–109, 230–231
ADAMLIB parameter 168
ADEF (efficacy dataset), creating 128–130
&ADEFKEEPSTRING macro 130
ADSL (Subject-Level dataset) 125–128, 176, 190–192
&ADSLKEEPSTRING macro 127
ADTTE (Time-to-Event Analysis dataset) 133–136
adverse dataset source data programs 210

Adverse Event Analysis datasets (ADAE) 131–132
Adverse Events events domain (AE), creating 47–49, 97–98
ADY variable 119
AE (Adverse Events events domain), creating 47–49, 97–98
AEBODSYS variable 181
AEDECOD variable 181
ALT (alanine aminotransferase) 176
Amendment 1 to the Study Data Tabulation Model (SDTM) v 1.2 and the SDTM Implementation Guide: Human Clinical Trials V3.1.2 207
Analysis Data Model
 See ADaM (Analysis Data Model)
Analysis Parameter Value-Level metadata 109–110
Analysis Results metadata 110–113, 233
ANALYSISDATASET metadata 111
ANALYSISVARIABLES metadata 111
ARCHIVELOCATIONID metadata 16
ARM variable 66
ARMCD variable 66
ARM_DEMOGRAPHIC_TRT format 32
ARROW variable 250
AST (aspartate transaminase) 176
AVAL variable 115, 130
AVALCAT variable 130

B

Base SAS
 about 14, 30
 adverse dataset 210
 background 57–58
 building SDTM datasets 61–66
 clinical standards setup for Study XYZ123 58–61
 creating empty SDTM domain dataset 32–35

Base SAS (*continued*)

- creating SDTM codelist SAS format catalog 31–32
- creating SDTM study day variable 36–37
- creating SDTM--DTC date variable 35–36
- define.xml 66–71
- demographics dataset 211–212
- dosing dataset 212–214
- implementing CDISC SDTM with 29–55
- laboratory dataset 215–221
- macros and tools for SDTM conversions 30–38
- pain scores dataset 221–222
- sorting final SDTM domain dataset 37–38
- source data programs 210–222

BASE variable 130

BDS (Basic Data Structure) 10, 128–130

bilirubin (BILI) 176

BLAs (Biologic License Applications) 2

BRIDG (Biomedical Research Integrated Domain Group) 205

building

See also creating

- SDTM datasets 32–35, 38–54, 61–66
- special purpose DM 38–42, 63–66, 89–97
- SUPPDM domains 38–42, 63–66, 89–97

BY DISPLAYID NOTSORTED statement 250

BY variable 176

C

case report form (CRF) 1

Case Report Tabulation Data Definition Specification (CRT-DDS or define.xml) final version 1.0 (website) 15

CDASH (Clinical Data Acquisition Standards Harmonization) Standard 5

CDER Common Issues document 3, 196

CDISC

- benefits of using data 171
- data review and analysis 171, 172–183, 183–185
- future directions 207

implementing SDTM with Base SAS

29–55, 146–150

model versioning 206–207

reasons for using 1–3

scenarios for implementing standards 10–11

website 8, 19, 107

CDISC-Published metadata 74

&CFB macro 130

challenges, with data integration 188–189

character variable 21

CHG variable 130

CLASS statement 16

Clinical Data Acquisition Standards

Harmonization (CDASH) Standard 5

clinical standards setup, for Study XYZ123 58–61

clinical studies, defining 79–82

Clinical Trials Repository (CTR) 206

CNSR variable 133

CODEDVALUE metadata 20

Codelist metadata

ADaM (Analysis Data Model) metadata 232

SDTM (Study Data Tabulation Model) 19–21, 226–228

CODELISTDICTIONARY metadata 20

CODELISTNAME metadata 18, 20, 23

CODELISTVERSION metadata 21

coding dictionaries 194–195

command-line mode, running OpenCDISC Validator in 158–162

COMMENT metadata 17, 22, 108

company-specific metadata 75

COMPLETED STUDY variable 133

Computational Method metadata

ADaM (Analysis Data Model) metadata 232

SDTM (Study Data Tabulation Model) 24–25, 228

COMPUTATIONALMETHODOID metadata 108

COMPUTATIONMETHOD metadata 24

- COMPUTATIONMETHODOID metadata 18, 22, 24
- config= option 160
- CONFIG parameter 167
- config:define= option 160
- configuration files, modifying 156–158
- CONTENTS procedure 95
- converting between native SAS and version 5.0 transport files 199–200
- copying SDTM 3.1.2 standard to XYZ123 58
- CQ01NAM variable 131, 132
- creating
 - See also* building
 - Adverse Events events domain (AE) 47–49, 97–98
 - define file generation programs 67–71
 - define file metadata 66–67
 - define.xml 25–26, 66–71, 103–105, 113
 - efficacy dataset (ADEF) 128–130
 - empty SDTM domain dataset 32–35
 - EX (Exposure Interventions) domain 49–51, 100
 - LB (Laboratory Findings) domain 42–45, 100–101
 - SAS format catalog 31–32
 - SDTM domains 89–103
 - SDTM study day variable 36–37
 - SDTM--DTC date variable 35–36
 - Trial Design Model domains 51–54, 101
 - XP (Pain Scale Customized Findings) domain 45–47, 99–100
- CRF (case report form) 1
- CRIT1 text string 109, 115
- CRIT1FL variable 115, 130, 184
- CROs 2, 4
- CTR (Clinical Trials Repository) 206
- customized code, using in SDTM production 101–102
- customizing, in JMP Clinical 181–183
- D**
- data integration
 - challenges 188–189
 - strategies 189–195
 - submission tools and 195–200
- DATA step 250, 251
- DATA step code 102, 119, 130
- &DATALIST macro 198
- DATA_NULL_step 250
- dataset creator, emptying via SAS clinical standards toolkit 61–62
- datasets
 - ADaM (Analysis Data Model) 183–185
 - EMPTY_** 34
 - Empty SDTM domain 32–35
 - EMPTY_ADEF 130
 - FORMATDATA 84, 92, 93
 - Subject-Level (ADSL) 125–128, 190–192
 - SUPP 124
 - SUPPQUAL 124
 - Time-to-Event Analysis (ADTTE) 133–136
 - Validation Control 140–141
 - WORK 124
- DATASETS parameter 168
- date conversions 118–120
- date variable 17
- datetime conversions 118–120
- datetime variable 17
- Define Header metadata
 - ADaM (Analysis Data Model) metadata 229
 - SDTM (Study Data Tabulation Model) 228
- DEFINE parameter 167
- define.xml
 - ADaM (Analysis Data Model) 107–108
 - Base SAS 66–71
 - building 25–26, 66–71, 103–105, 113
 - creating generation programs 67–71
 - defining file creation SAS program 26
 - defining file header metadata 26
 - metadata 66–67
 - navigation 113–116
 - rendering 113–116
 - SAS Clinical Standards Toolkit 66–71
 - using SAS Clinical Data Integration to create 103–105

defining

- clinical studies and subfolders 79–82
- file creation SAS program 26
- file header metadata 26
- target SDTM datasets 85–87

DEMOGRAPHIC TRT variable 32

demographics dataset source data programs
211–212

DILI (drug-induced liver injury) 176

%DIRDELIM macro 199

DISPLAYFORMAT metadata 18, 22

DISPLAYID metadata 111, 112

DISPLAYNAME metadata 111

DOCLEAFID metadata 250

DOCTITLE metadata 250

DOCUMENTATION metadata 108, 111, 112,
250

DOMAIN metadata 17

domain sorting, via SAS clinical standards
toolkit 62

DOMAINKEYS metadata 15

DOMAINS parameter 121–124

dosing dataset source data programs 212–214

DO-WHILE loop 198, 251

drug-induced liver injury (DILI) 176

Drug-Induced Liver Injury: Premarketing
Clinical Evaluation (FDA) 176

%DTC2DT macro 126, 130, 132

E

eCTD (FDA) 3

EDC (electronic data capture) systems 4

editing

XYZ123 metadata 58–59

XYZ123 SAS clinical standards toolkit
control files 59–60

efficacy dataset (ADEF), creating 128–130

electronic data capture (EDC) systems 4

EMPTY_** dataset 34

Empty SDTM domain dataset, creating 32–35

EMPTY_ADEF dataset 130

evaluating Report file and Report results
152–156

EVNTDESC variable 133

EVS (NCI Enterprise Vocabulary Services) 19

EX (Exposure Interventions) domain, creating
49–51, 100

exchanging non-clinical data 203–204

EXIT command 161

External Links metadata 234

externalxml 69

Extract transformation 90–91

F

FDA

documents and events in support of data
standards 3

Janus 206

Safety and Innovation ACT (FDASIA) 206
use of OpenCDISC Validator by 150–151

FDASIA (FDA Safety and Innovation Act) 206

file creation SAS program, defining 26

file header metadata, defining 26

FILEOID metadata 26

FILES parameter 167

final SDTM domain dataset, sorting 37–38

float variable 17, 20

FORMATDATA dataset 84, 92, 93

%FROMEXP macro 200

G

governance, need for 5–6

Guidance for Industry: Integrated Summary of
Effectiveness (FDA) 188Guidance for Industry: Premarketing Risk
Assessment (FDA) 188**H**

Healthcare Link (website) 207

HL7 (Health Level Seven International)
204–205

HTMLENCODE statement 250

Hy's Law Screening routine 176–180

I

ICH M4E(R1)-Clinical Overview and Clinical Summary of Module 2 188

IDVAR variable 124

IHE (Integrating the Healthcare Enterprise) 207

implementation guide (ADaM 1G version 1.0) 8–11

implementation strategies

- ADaM considerations 8–11
- CDASH 5
- CDISC 1–3
- models 4
- need for governance 5–6
- SDTM considerations 6–8

implementing

- ADaM 117–137
- CDISC SDTM with Base SAS 29–55

IMPORT procedure 101, 250

INDEX variable 100

installing OpenCDISC Validator 151

integer variable 17, 20

integrated summary of efficacy 187

integrated summary of safety 187

Integrating the Healthcare Enterprise (IHE) 207

integration 187

ISO 8601 date conversions 118–120

ISREFERENCEDATA metadata 15

ITEMOID metadata 250

J

Janus 206

JMP Clinical

- customizing 181–183
- getting started 172–176
- patient profiles 180–181
- safety analyses 176–180
- safety evaluations with 172–183

K

Kaplan Meier estimates 184

**KEEPSTRING macro 34

L

LABEL metadata 15, 17, 22

laboratory dataset source data programs 215–221

LB (Laboratory Findings) domain, creating 42–45, 100–101

LBCAT variable 101

LBTEST variable 101

LBTESTCD variable 101

LEAFID variable 112, 250

LEAFRELPATH variable 250

LENGTH statement 17, 22

library 32, 38, 42, 45, 47, 49, 63

LIFETEST procedure 184

LLT (lower-level term) 194

Lookup transformation 92–94

M

macros

- &ADEFKEEPSTRING 130
- &ADSLKEEPSTRING 127
- &CFB 130
- &DATA LIST 198
- %DIRDELIM 199
- %DTC2DT 126, 130, 132
- %FROMEXP 200
- **KEEPSTRING 34
- make_codelist_format.sas 30
- make_dtc_date.sas 30
- make_empty_dataset.sas 30
- %make_define SAS 235–251
- %MAKE_EMPTY_DATASET 126, 127
- %MAKE_SORT_ORDER 127
- make_sdtm_dy.sas 30
- make_sort_order.sas 30
- MAXLENGTH 196–198
- %MAXLENGTH2 199, 201
- %MERGSUPP 121–124, 126, 200
- running OpenCDISC Validator as a 165–167
- %run_opendisc 201
- for SDTM conversions 30–38
- &SeverityVarMH 182

macros (*continued*)

- &TermVARMH 183
- %TOEXP 199, 201
- &VARLIST 198
- make_codelist_format.sas macro 30
- make_dtc_date.sas macro 30
- make_empty_dataset.sas macro 30
- %make_define SAS macro 235–251
- %MAKE_EMPTY_DATASET macro 126, 127
- %MAKE_SORT_ORDER macro 127
- make_sdtm_dy.sas macro 30
- make_sort_order.sas macro 30
- MANDATORY metadata 18, 23
- MAXLENGTH macro 196–198
- %MAXLENGTH2 macro 199, 201
- MedDRA codes 194–195
- merging supplemental qualifiers 120–124
- %MERGSUPP macro 121–124, 126, 200
- metadata
 - See also specific types*
 - in ADaM (Analysis Data Model) 108–109
 - Analysis Parameter Value-Level 109–110
 - Analysis Results 110–113, 233
 - CDISC-Published 74
 - Codelist 19–21, 226–228, 232
 - company-specific 75
 - file header 26
 - Reference dataset 141–142
 - SAS clinical data integration 74–78
 - SDTM (Study Data Tabulation Model) 14–15
 - Source dataset 141–142
 - Value-Level 22–24, 228
- MHBODSYS variable 181
- MHDECOD variable 181–183
- MHSEV variable 182
- MHTERM variable 182
- MHTOXGR variable 182
- model document (ADaM version 2.1) 8–11
- model versioning 206–207
- models 4
- modifying configuration files 156–158
- MONTH3DT variable 99

- MONTH6DT variable 99

N

- NAME metadata 15
- navigation, in define.xml 113–116
- NCI (National Cancer Institute) 156
- NCI Enterprise Vocabulary Services (EVS) 19
- NDAs (New Drug Applications) 2
- non-clinical data, exchanging 203–204
- NOXWAIT option 161
- number variable 21

O

- OID metadata 15
- OpenCDISC Validator
 - about 140, 150–151, 164–165
 - evaluating Report file and Report results 152–156
 - installing 151
 - modifying configuration files 156–158
 - running 151–167
 - running as a macro 165–167
 - running in command-line mode 158–162
- OPENCDISCPATH parameter 167
- ORIGIN metadata 17, 22, 108
- OUTLIB parameter 124

P

- PAIN ADVERSE EVENT variable 133
- PAIN RELIEF variable 133
- Pain Scale Customized Findings (XP) domain,
 - creating 45–47, 99–100
- pain scores dataset source data programs 221–222
- PAIN WORSENING variable 133
- PAIN3MO variable 99
- PAINBASE variable 99
- PAN6MO variable 99
- PARAM variable 109, 251
- PARAMCD variable 109, 251
- parameter invariant 130
- Parameter-Level metadata 232
- parameters

ADAMLIB 168
 CONFIG 167
 DATASETS 168
 DEFINE 167
 DOMAINS 121–124
 FILES 167
 OPENDISCPATH 167
 OUTLIB 124
 REFDT 119, 130, 132
 SOURCELIB 124
 SOURCES 167
 SRCLIB 168
 VALIDATORAJAR 167
 PARAMLIST metadata 111, 251
 patient profiles, in JMP Clinical 180–181
 PDUFA (Prescription Drug User Fee Act) 2
 PhaseForward's WebSDM 140–142
 Points to Consider on Application with 1. Meta-Analyses; 2. One Pivotal Study (EMA) 188
 Prescription Drug User Fee Act (PDUFA) 2
 PRM (Protocol Representation Model) 205–206, 207
 PROBLEM metadata 169–170
 PROBTYP metadata 169–170
 procedures
 about 183–185
 CONTENTS 95
 IMPORT 101, 250
 LIFETEST 184
 TRANPOSE 95, 100
 PROGRAM metadata 112
 PROGRAMMINGCODE metadata 112
 Protocol Representation Model (PRM) 205–206, 207
 PROTOCOLNAME metadata 26
 Providing Regulatory Submissions in Electronic Format - Standardized Study Data (FDA) 3
 PURPOSE metadata 15

Q

QLABEL variable 96

QNAME value 96, 124

QVAL variable 96

R

RACE variable 66
 RANDOMIZEDT 99
 RANK metadata 20
 REASON metadata 111
 REFDT parameter 119, 130, 132
 Reference dataset metadata 141–142
 Reference Information Model (RIM) 205
 registering
 source datasets 82–84
 XYZ123 standard 60–61
 regulatory submissions
 about 187–188
 data integration and 195–200
 guidance 188
 preparing 200–201
 RENAME clause 65
 rendering, in define.xml 113–116
 REPEATING metadata 15
 Report file/results, evaluating 152–156
 -report= option 160
 -report:overwrite= option 160
 RESULTID metadata 111
 RESULTNUM metadata 250
 RFENDTDC variable 65
 RFSTDTC variable 65
 RFXSTDTC domain 157–158
 RIM (Reference Information Model) 205
 ROLE metadata 18, 23, 108
 ROLECODELIST metadata 18, 23, 108
 running OpenCDISC Validator
 in command-line mode 158–162
 Graphical User Interface 151–152
 as a macro 165–167
 %run_opendisc macro 201

S

Safety and Innovation Act (FDASIA) 206
 safety evaluations, with JMP Clinical 172–183
 SAFETY flag 176

- SAS Clinical Data Integration tool
 - about 14, 146–150
 - creating define.xml 103–105
 - creating SDTM domains 89–103
 - implementing CDISC SDTM with 73–105
 - metadata 74–78
 - setting defaults 87–88
 - study setup 78–88
 - using to create define.xml 103–105
- SAS Clinical Standards Toolkit
 - about 140
 - background 57–58
 - building SDTM datasets 61–66
 - clinical standards setup for Study XYZ123 58–61
 - define.xml 66–71
 - domain sorting via 62
 - emptying dataset creator via 61–62
 - setup 140–142
 - validation program 142–146
- SAS Clinical Standards Toolkit 1.3: User's Guide 58, 60, 140, 146
- SAS format catalog, creating 31–32
- SCHEMALOCATION metadata 26
- SDTM (Study Data Tabulation Model)
 - about 1, 4, 38
 - AE (Adverse Events) events domain 47–49, 97–98
 - Base SAS macros and tools for conversions 61–62
 - building datasets 32–35, 38–54, 61–66
 - Codelist metadata 19–21, 226–228
 - Computational Method metadata 24–25, 228
 - considerations 6–8
 - copying 3.1.2 standard to XYZ123 58
 - creating domains 89–103
 - Define Header metadata 228
 - EX (Exposure Interventions) domain 49–51, 100
 - LB (Laboratory Findings) domain 42–45, 100–101
 - metadata 14–15
 - special purpose DM 38–42
 - SUPPDM domains 38–42
 - Table of Contents metadata 15–16, 223
 - TDM (Trial Design Model) domains 51–54, 101
 - templating conversion jobs for reuse 102–103
 - Terminology.txt file 156
 - using customized code in production 101–102
 - Value-Level metadata 22–24, 228
 - Variable-Level metadata 16–19, 224–226
 - XP (Pain Scale Customized Findings) domain 45–47, 99–100
- SDTM study day variable, creating 36–37
- SDTM validation
 - about 139–140
 - OpenCDISC Validator 150–162
 - SAS Clinical Data Integration 146–150
 - SAS Clinical Standards toolkit 140–146
- SDTM--DTC date variable, creating 35–36
- SE (Subject Elements) domain 101
- SELECTIONCRITERIA metadata 111
- SENDIG (SEND Implementation Guide) 203
- SET statement 65
- setting
 - SAS clinical data integration defaults 87–88
 - variable lengths 196–198
- setup
 - SAS clinical data integration metadata 75–78
 - SAS clinical data integration study 78–88
 - SAS Clinical Standards Toolkit 140–142
- &SeverityVarMH macro 182
- SEX variable 66
- SIGNIFICANTDIGITS metadata 17, 22
- sorting final SDTM domain dataset 37–38
- **SORTSTRING variable 62
- source 38, 42, 47, 49, 63
- source data programs
 - adverse dataset 210
 - demographics dataset 211–212
 - dosing dataset 212–214

- laboratory dataset 215–221
 - pain scores dataset 221–222
 - Source dataset metadata 141–142
 - source datasets, registering 82–84
 - source= option 160
 - sourcedata 67, 69
 - SOURCEDATASET metadata 21
 - SOURCELIB parameter 124
 - sourcemetadata 67
 - sourcerootpath 67, 69
 - SOURCES parameter 167
 - SOURCEVALUE metadata 21
 - SOURCEVARIABLE metadata 21
 - special purpose DM
 - building 38–42, 63–66
 - creating 89–97
 - spreadsheets (ADaM metadata) 108–113
 - SQL Join transformation 91
 - SRCDOM variable 136, 167–170
 - SRCLIB parameter 168
 - SRCSEQ variable 136, 167–170
 - SRCVAR variable 167–170
 - Standard for Exchange of Non-Clinical Data 203–204
 - STANDARD metadata 26
 - statements
 - CLASS 16
 - BY DISPLAYID NOTSORTED 250
 - HTMLENCODE 250
 - LENGTH 17, 22
 - SET 65
 - WHERE 91, 93, 176
 - strategies, data integration
 - about 189–190
 - choosing data 193–194
 - coding dictionaries 194–195
 - subject-level datasets 190–192
 - strategies, implementation
 - ADaM considerations 8–11
 - CDASH 5
 - CDISC 1–3
 - models 4
 - need for governance 5–6
 - SDTM considerations 6–8
 - STRUCTURE metadata 15
 - Study Data Tabulation Model
 - See* SDTM (Study Data Tabulation Model)
 - Study XYZ123
 - clinical standards setup for 58
 - copying 3.1.2 standard to 58
 - editing metadata 58–59
 - editing SAS clinical standards toolkit
 - control files 59–60
 - registering standard 60–61
 - study-based custom codelist format catalog 141
 - STUDYDESCRIPTION metadata 26
 - STUDYID variable 95
 - STUDYNAME metadata 26
 - STUDYOID metadata 26
 - studyoutputpath 67, 69
 - STYLESHEET metadata 26
 - subfolders, defining 79–82
 - Subject Elements (SE) domain 101
 - Subject Sequence Generator transformation 97–98
 - SUBJECT variable 90
 - Subject Visits (SV) domain 101
 - Subject-Level dataset (ADSL) 125–128, 190–192
 - SUBJID variable 66
 - submissions, regulatory
 - about 187–188
 - data integration and 195–200
 - guidance 188
 - preparing 200–201
 - SUPP dataset 124
 - SUPPDM domain 38–42, 63–66, 89–97
 - supplemental qualifiers, merging 120–124
 - SUPPQUAL dataset 124
 - SV (Subject Visits) domain 101
- T**
- TA (Trial Arms) 101

Table of Contents metadata
 ADaM (Analysis Data Model) metadata 230
 SDTM (Study Data Tabulation Model) 15–16, 223
 target SDTM datasets, defining 85–87
 TDM (Trial Design Model) domains, creating 51–54, 101
 TE (Trial Elements) 101
 templating SDTM conversion jobs for reuse 102–103
 &TermVARMH macro 183
 TESTCD variable 196
 text variable 17, 20
 TI (Trial Inclusion) 101
 time variable 17
 Time-to-Event Analysis dataset (ADTTE) 133–136
 TITLE metadata 250
 %TOEXP macro 199, 201
 tools
 ADaM (Analysis Data Model) 118–124
 data integration and submission 195–200
 for SDTM conversions 30–38
 traceability checks 167–170
 TRANSLATED metadata 20
 transport files, converting between native SAS and Version 5.0 199–200
 TRANSPOSE procedure 95, 100
 Transpose transformation 94–97
 Trial Arms (TA) 101
 Trial Design Model (TDM) domains, creating 51–54, 101
 Trial Elements (TE) 101
 Trial Inclusion (TI) 101
 Trial Summary (TS) 101
 Trial Visits (TV) 101
 trialdesign.xlsx 51
 TRTSEQA variable 192
 TRTSEQP variable 191
 TS (Trial Summary) 101
 TV (Trial Visits) 101
 TYPE metadata 17, 20, 22

U

UML (Unified Modeling Language) 205
 UNIQUEID variable 66, 90
 "up version" 189
 USUBJID variable 66, 95, 98, 101, 189

V

validation, ADaM 163–170
 validation, SDTM
 about 139–140
 OpenCDISC Validator 150–162
 SAS Clinical Data Integration 146–150
 SAS Clinical Standards toolkit 140–146
 Validation Control dataset 140–141
 Validation Properties file 142
 VALIDATORAJAR parameter 167
 Value-Level metadata 22–24, 228
 VALUelistOID metadata 18, 22, 110
 VALUENAME metadata 22
 variable lengths, setting 196–198
 VARIABLE metadata 17
 Variable-Level metadata
 ADaM (Analysis Data Model) metadata 108–109, 224–226, 230–231
 SDTM (Study Data Tabulation Model) 16–19, 224–226
 variables
 BY 176
 ABLFL 130
 ADY 119
 AEBODSYS 181
 AEDECOD 181
 ARM 66
 ARMCD 66
 ARROW 250
 AVAL 115, 130
 AVALCAT 130
 BASE 130
 character 21
 CHG 130
 CNSR 133
 COMPLETED STUDY 133
 CQ01NAM 131, 132

CRIT1FL 115, 130, 184
 date 17
 DEMOGRAPHIC TRT 32
 EVNTDESC 133
 float 17, 20
 IDVAR 124
 INDEX 100
 integer 17, 20
 LBCAT 101
 LBTEST 101
 LBTESTCD 101
 LEAFID 112, 250
 LEAFRELPATH 250
 MHBODSYS 182
 MHDECOD 181–183
 MHSEV 182
 MHTERM 182
 MHTOXGR 182
 MONTH3DT 99
 MONTH6DT 99
 number 21
 PAIN ADVERSE EVENT 133
 PAIN RELIEF 133
 PAIN WORSENING 133
 PAIN3MO 99
 PAINBASE 99
 PAN6MO 99
 PARAM 109, 251
 PARAMCD 109, 251
 QLABEL 96
 QVAL 96
 RACE 66
 RANDOMIZEDT 99
 RFENDTC 65
 RFSTDTC 65
 SEX 66
 **SORTSTRING 62
 SRCDOM 136, 167–170
 SRCSEQ 136, 167–170
 SRCVAR 167–170
 STUDYID 95
 SUBJECT 90
 SUBJID 66

TESTCD 196
 text 17, 20
 time 17
 TRTSEQA 192
 TRTSEQP 191
 UNIQUEID 66, 90
 USUBJID 66, 95, 98, 101, 189
 VISIT 101
 VISITDY 145
 VISITNUM 100, 174
 VSSTRESN 109
 VSTEST 109
 VSTESTCD 109
 &VARLIST macro 198
 VARNUM metadata 17
 VERSION metadata 26
 VISIT variable 101
 VISITDY variable 145
 VISITNUM variable 100, 174
 VSSTRESN variable 109
 VSTEST variable 109
 VSTESTCD variable 109

W

web forums (OpenCDISC.org) 158
 WebSDM 140–142
 websites

- ADaM Basic Data Structure for Time-to-Event Analyses document 133
- ADaM published validation checks 163
- Biomedical Research Integrated Domain Group (BRIDG) 205
- Case Report Tabulation Data Definition Specification (CRT-DDS or define.xml) final version 1.0 15
- CDASH (Clinical Data Acquisition Standards Harmonization) Standard 5
- CDISC 8, 19, 107
- Healthcare Link 207
- %MAXLENGTH2 macro 199, 201
- OpenCDISC 156
- OpenCDISC validation rules 157
- OpenCDISC Validator 150, 158

websites (*continued*)

PRM (Protocol Representation Model) 205

SAS transport files 199

SENDIG (SEND Implementation Guide)

203

tab-delimited file of NCI data 156

traceability checks 168

WHERE statement 91, 93, 176

WORK dataset 124

X

XP (Pain Scale Customized Findings) domain,

creating 45–47, 99–100

XPTESTCDVALUE metadata 110

XSYNC option 161

Z

Zimmerman, Hy, Dr. 176

From *Implementing CDISC Using SAS®: An End-to-End Guide* by Chris Holland and Jack Shostak.
Copyright © 2012, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

About The Authors



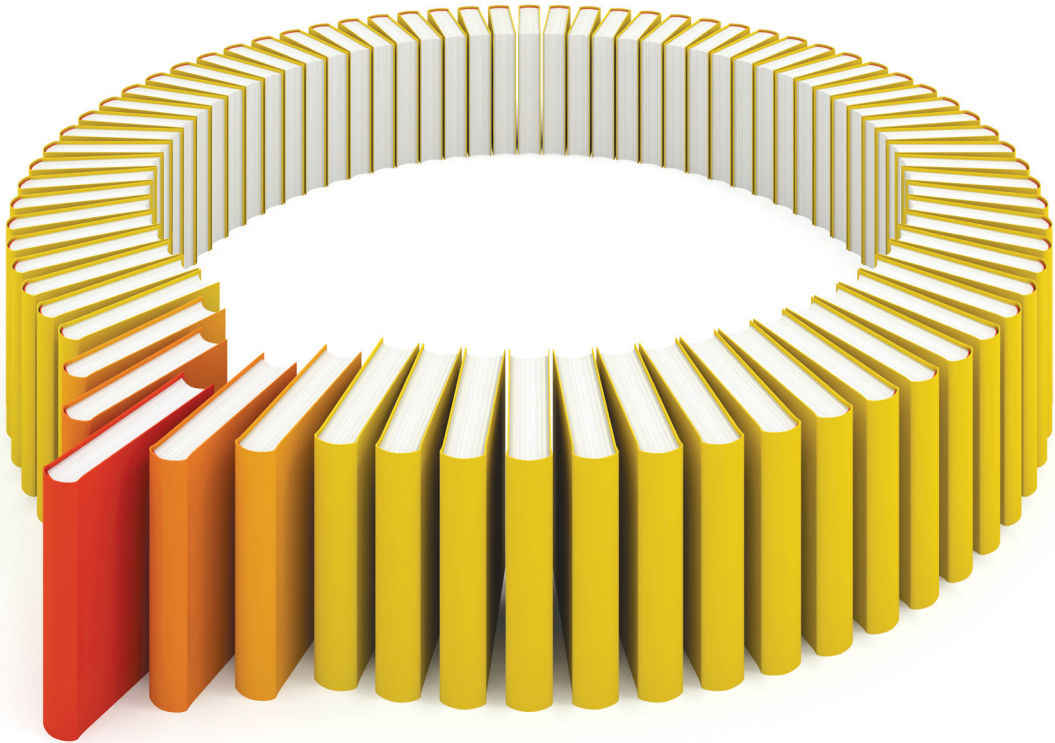
Chris Holland has been a SAS user since 1990. He currently works as a Biostatistics Director for Amgen and has prior experience heading up biostatistics departments for Sucampo Pharmaceuticals, MacroGenics, and Micromet. He has also worked as a statistical reviewer at the Center for Drug Evaluation and Research in the U.S. Food and Drug Administration. There he served as the technical lead for the SDTM/ADaM Pilot Project review team, which included FDA staff involved with assessing the suitability of CDISC standards for meeting the needs of medical and statistical reviewers. Holland continues to be active in the CDISC community, particularly with the ADaM team. He received an MS in Statistics from the University of Virginia, and a BS in Statistics from Virginia Polytechnic Institute and State University. He is an Accredited Professional Statistician™ by the American Statistical Association.



Jack Shostak, Associate Director of Statistics, manages a group of statistical programmers at the Duke Clinical Research Institute. A SAS user since 1985, he is the author of *SAS Programming in the Pharmaceutical Industry*, and coauthor of *Common Statistical Methods for Clinical Research with SAS Examples, Third Edition*, both published by SAS Press. Shostak has published papers for the Pharmaceutical SAS Users Group (PharmaSUG) and the NorthEast SAS Users Group (NESUG), and he contributed a chapter, “Reporting and SAS Tool Selection,” in the book *Reporting from the Field*. Jack is active in the CDISC community, contributing to the development of ADaM and the Statistical Domain Analysis Model, and he serves as a CDISC ADaM trainer. He received an MBA from James Madison University, and a BS in Statistics from Virginia Polytechnic Institute and State University.

Learn more about these authors by visiting their author pages, where you can download free chapters, access example code and data, read the latest reviews, get updates, and more:

- <http://support.sas.com/hollandc>
- <http://support.sas.com/shostak>



Gain Greater Insight into Your SAS[®] Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

 support.sas.com/bookstore
for additional books and resources.


THE POWER TO KNOW.®

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. © 2013 SAS Institute Inc. All rights reserved. S107969US.0413