Implementing CDISC Using SAS®

An End-to-End Guide

Revised Second Edition
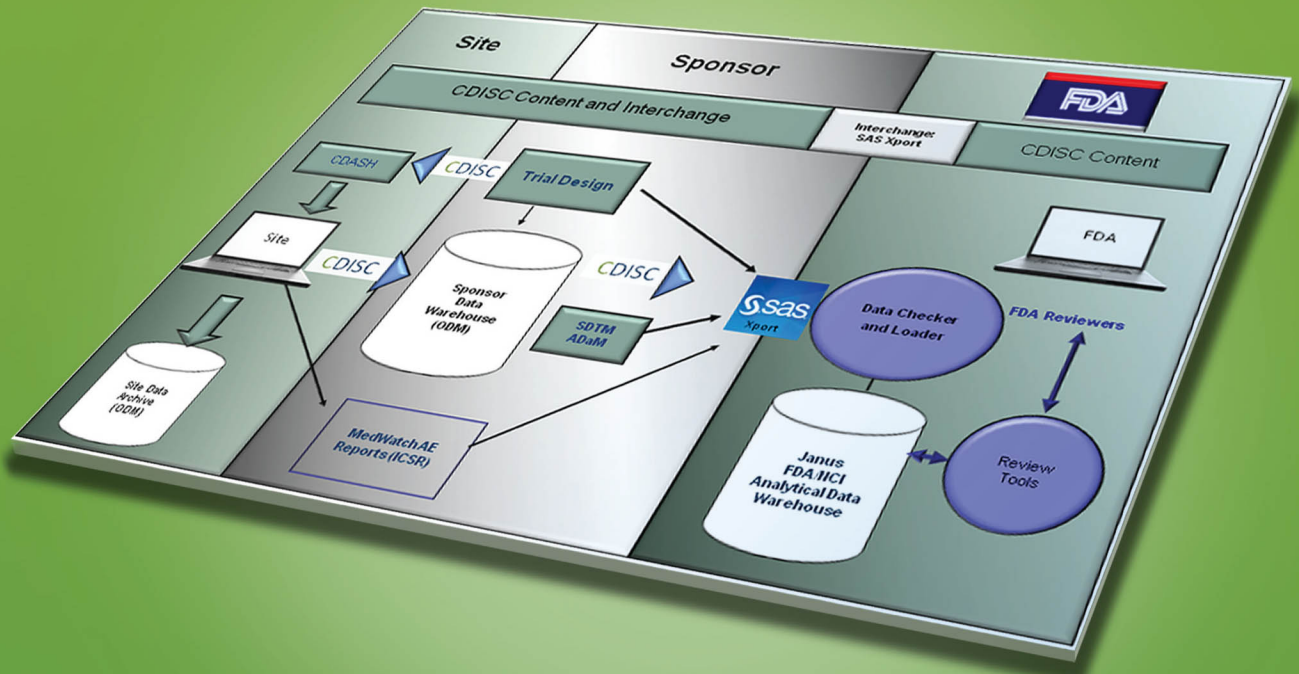
Chris Holland
Jack Shostak

About This Book

Background

The Clinical Data Interchange Standards Consortium (CDISC) started in 1997 as a global, open, multidisciplinary, non-profit organization focused on establishing standards to support the acquisition, exchange, submission, and archiving of clinical research data and metadata. The mission of CDISC is "to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare."

CDISC standards are promoted as being "vendor-neutral, platform-independent and freely available via the CDISC website." When one speaks of clinical research data, however, one particular vendor often comes to mind. SAS software has been used for analyzing, summarizing, and reporting clinical trial data since the early 1980s. The US Food and Drug Administration (FDA) has required that electronic data provided to them in marketing approval submissions be formatted in the SAS 5.0 transport file format. Although this format is now an open standard accepted by many types of software, it has played a role in establishing and maintaining SAS as the industry standard in the pharmaceutical industry for numerous tasks relating to clinical trial data.

The flagship standard developed by a CDISC working group is the Study Data Tabulation Model (SDTM). The term "tabulation" has its origins in the code of federal regulations (CFR) regarding what is required in New Drug Applications (NDAs). Specifically, 21 CFR 314.50 describes case report tabulations (CRTs) as "data on each patient in each study" that are "pertinent to a review of the drug's safety or effectiveness." Originally, these tabulations were generated as listings on paper. However, by the late 1990s, electronic data sets were playing an increasing role in new drug reviews. As a result, the SDTM was developed as a standard data format for the electronic submission of these required tabulations.

Since the inception of CDISC, its bond with SAS and SAS users in the pharmaceutical industry has been a natural marriage. Over the past decade, as CDISC standards have built up momentum in terms of stabilizing and gaining regulatory approval and preference, SAS programmers have played an increasing role in the complete life cycle of clinical trial data—from collection, to analyzing, reporting, documenting, validating, submitting and archiving data, and all the steps in between. As such, the time seems right to provide an end-to-end handbook for SAS users in the pharmaceutical industry who are working with CDISC standards.

What's New in This Edition?

When writing the first edition of this book, the FDA was still not requiring that marketing applications for new drugs and biologics be submitted electronically, much less that CDISC standards be applied to the data sets that accompanied fully electronic submissions. But the writing was on the wall, so to speak (or on the screens at industry conference presentations), that such changes were imminent. This was the motivation for that first edition—to help late adopters get on board with the someday-to-be-required CDISC implementation.

Then, in July of 2012, the FDA Safety and Innovation Act, or FDASIA, was passed, setting the wheels in motion for the requirement of electronic marketing applications including standardized study data. The Act deferred the details about the format of these soon-to-be-required electronic submissions to guidance documents, but it set a time table for them to take effect "beginning no earlier than 24 months after issuance of a final guidance." In December 2014, "Providing Regulatory Submissions in Electronic Format – Standardized Study Data" was issued as a final guidance document, thereby setting December 2016 as the due date for official conversion to CDISC standards for future submission. The guidance document, as expected, specifies the use of both SDTM and ADaM datasets.

But these requirements by themselves were not uniquely motivating for a new edition of our book. As we were writing the first edition, we knew that the rapid evolution of CDISC standards would render at least part of the book outdated before it even came out. Perhaps most notable among these changes was the release and finalization of the Define-XML 2.0 standard, a more extensible version of the original standard for dataset meta-data. Many other changes have also since been released, including a host of new SDTM domains, a new version to the ADaM Implementation Guide (1.2), a new data structure for occurrence-related analysis data sets (the OCCDS), new ADaM validation checks (now on version 1.3), new Pinnacle 21 software, new therapeutic area standards, and, of course, new versions to the SAS and JMP software used to implement these standards. Many of these updates have been reflected within their respective relevant sections within the book.

What hasn't changed is the process by which we recommend that these standards be implemented. The use of CDASH at the collection stage is more important than ever, although mostly outside the scope of this book. The subsequent "linear" conversions from the raw, collected data to the SDTM and ADaM standards are still very much a basic tenet of that process. We also continue to espouse the virtues of establishing the metadata before the creation of the actual data that they describe. So from that perspective, the "end-to-end" organization of the book has remained largely unchanged.

Based on user feedback, we have issued a revised version of the second edition to update the make_define macro that accompanies this book to add the capability to handle C-Codes. The metadata spreadsheets were updated accordingly. We also added the latest C-Codes from the most recent release.

Is This Book for You?

Any manager or user of clinical trial data in this day and age is likely to benefit from knowing how to either put data into a CDISC standard or analyze and find data once it is in a

CDISC format. If you are one such person—a data manager, clinical and/or statistical programmer, biostatistician, or even a clinician, then this book is for you.

Prerequisites

It is assumed that the readers of this book have some basic understanding of SAS programming. It is also advantageous if the reader has some familiarity with the CDISC SDTM and ADaM models. If you do not have this knowledge, then there will be some portions of this book that will be more difficult to follow.

Purpose

This book has two purposes. One is to introduce readers to the end-to-end process for implementing CDISC standards on clinical trial data. The second is to introduce readers to tools based on SAS that cannot only facilitate the implementation process but also facilitate the ultimate goal of analyzing your data once it follows a CDISC standard. Some of these tools are actual SAS products. Some are SAS macros provided by the authors.

In order to achieve these goals, the authors have created fictional clinical trial data that will be used to provide examples of how to apply SAS tools to the various steps in the process along the way.

Organization and Scope of This Book

With the end-to-end concept in mind, this book is organized with a start-to-finish mentality for clinical trial data. Chapter 1 starts with an overview of the primary standards covered in this book, the SDTM and the Analysis Data Model (ADaM), and a discussion of considerations as to when each should be implemented. Chapter 2 is focused on what should be the first step to any big project, the specifications. Although many currently think of the creation of metadata and data documentation as part of the process that comes after building the datasets, the authors hope that the tools and tips provided here will facilitate the creation of a define.xml file as a tool for implementing SDTM standards before any of the actual conversion work is performed.

Chapters 3, 4, and 5 all cover SDTM implementation, but with different tools (Base SAS, SAS Clinical Standards Toolkit, and SAS Clinical Data Integration, respectively). Chapter 6 relates to ADaM metadata while Chapter 7 covers ADaM implementations using Base SAS. With regulatory, quality, and compliance concerns in mind, Chapters 8 and 9 cover CDISC validation using SAS software and Pinnacle 21 software, respectively. Chapter 10 is about analyzing data that have been structured to follow CDISC standards, with a particular focus on JMP Clinical, which has been specifically developed to work on SDTM and ADaM data. Finally, concluding the "end-to-end" process, Chapter 11 provides information on how to integrate your data from various studies into one integrated database and other steps required for providing your data in a regulatory submission.

Although many CDISC standards have stabilized, there are new standards on the horizon. Chapter 12 looks at some of these uncertainties, such as the Janus clinical trial data repository and HL7 messaging for CDISC content. Other topics relating to CDISC standards are covered here as well, such as the SEND model for non-clinical data and the BRIDG model for relating HL7, CDISC, and other standards, models, and semantics together.

What is not covered in this book are details, guidelines, and examples on how to put certain types of data into the CDISC models. These are details that CDISC therapeutic area working teams are most adept at dealing with. The SDTM implementation guide (IG), contains a wealth of knowledge pertaining to how, when, and where to put clinical trial data in SDTM domains. Similarly, the ADaM IG and associated appendixes have more information to help implementers decide how to build ADaM datasets.

Software Used to Develop This Book's Content

As you might expect, the majority of the software used in this book is SAS software. The following software was used in the production of this book:

- Base SAS 9.4
- SAS Clinical Standards Toolkit 1.6
- SAS Clinical Data Integration 2.6
- JMP 11.2
- JMP Clinical 5.0
- Pinnacle 21 Community 2.1.0

Various versions of Internet Explorer and Chrome were used to render the define.xml files

Data and Programs Used in This Book

The majority of the source data for this book can be generated by the SAS programs found in Appendix A. There is some study metadata used to generate the Trial Design Model datasets for the SDTM that you can find in Microsoft Excel files on the author pages listed below. The metadata spreadsheets used for the ADaM and SDTM data can also be found on the author pages listed below. Finally, all programs used in this book can be found on the author pages as well.

Author Pages

You can access this book's author pages at:

http://support.sas.com/hollandc

http://support.sas.com/shostak

The author pages include all of the SAS Press books that these authors have written. The links below the cover image will take you to a free chapter, example code and data, reviews, updates, and more.

Example Code and Data

You can access the example code and data for this book from the author pages listed in the previous "Author Pages" section. On the author page, select "Example Code and Data" to display the SAS programs that are included in this book.

For an alphabetical listing of all books for which example code and data are available, see http://support.sas.com/documentation/onlinedoc/code.samples.html. Select a title to display the book's example code.

If you are unable to access the code through the website, send email to saspress@sas.com.

Additional Help

Although this book illustrates many analyses regularly performed in businesses across industries, questions specific to your aims and issues may arise. To fully support you, SAS Institute and SAS Press offer you the following help resources:

- For questions about topics covered in this book, contact the author through SAS Press:
  - Send questions by email to saspress@sas.com; include the book title in your correspondence.
  - Submit feedback on the author's page at http://support.sas.com/author_feedback.
- For questions about topics in or beyond the scope of this book, post queries to the relevant SAS Support Communities at https://communities.sas.com/welcome.
- SAS Institute maintains a comprehensive website with up-to-date information. One page that is particularly useful to both the novice and the seasoned SAS user is its Knowledge Base. Search for relevant notes in the "Samples and SAS Notes" section of the Knowledge Base at http://support.sas.com/resources.
- Registered SAS users or their organizations can access SAS Customer Support at http://support.sas.com. Here you can pose specific questions to SAS Customer Support; under *Support*, click *Submit a Problem*. You will need to provide an email address to which replies can be sent, identify your organization, and provide a customer site number or license information. This information can be found in your SAS logs.

Keep in Touch

We look forward to hearing from you. We invite questions, comments, and concerns. If you want to contact us about a specific book, please include the book title in your correspondence.

Contact the Author through SAS Press

- By e-mail: saspress@sas.com
- Via the Web: http://support.sas.com/author_feedback

xiv

Purchase SAS Books

For a complete list of books available through SAS, visit sas.com/store/books.

- Phone: 1-800-727-0025
- E-mail: sasbook@sas.com

Subscribe to the SAS Learning Report

Receive up-to-date information about SAS training, certification, and publications via email by subscribing to the SAS Learning Report monthly eNewsletter. Read the archives and subscribe today at http://support.sas.com/community/newsletters/training!

Publish with SAS

SAS is recruiting authors! Are you interested in writing a book? Visit http://support.sas.com/saspress for more information.

Acknowledgments

This book would not have been possible without the thorough review and thoughtful feedback of several individuals that we would like to thank and recognize for their contribution.

Internal SAS support:

- Editor: Stacey Hamilton
- Technical Publishing Specialist: Denise T. Jones
- Cover design: Robert Harris
- Copyeditor: Kathy Underwood (with marketing materials edited by John West)

Technical reviewers:

- Lex Jansen
- Nate Freimark
- Bill J. Gibson
- Geoffrey Mann

Jack would also like to dedicate this edition in memory of his father John and dog Susie. They taught him many lessons in how to do the right things and how to smell the roses, respectively.

Chris would like to dedicate this edition to all of the devoted CDISC volunteers who contribute so much of their time and energy to the worthy cause of developing data standards in our industry.

About These Authors

CHRIS HOLLAND has been a SAS user since 1990 in a variety of settings ranging from academia, contract research organizations, regulatory agencies, and both small and large pharmaceutical and biotechnology companies. He was first introduced to CDISC while working 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 FDA review team and invented an early version of the MAED Service, an adverse event review tool that made use of data standards and is currently in production at the FDA. Since leaving the FDA, 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, a BS in statistics from Virginia Tech, and is an Accredited Professional Statistician™ by the American Statistical Association.
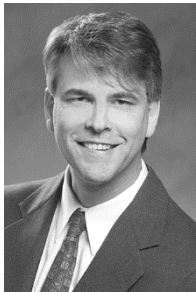
JACK SHOSTAK, Associate Director of Statistics, manages a group of statistical programmers at the Duke Clinical Research Institute. A SAS user since 1985, he is the author of *SAS® Programming in the Pharmaceutical Industry, Second Edition* and coauthor of *Common Statistical Methods for Clinical Research with SAS® Examples, Third Edition*, as well as *Implementing CDISC Using SAS®: An End-to-End Guide*. Shostak has published papers and given talks for various industry groups, including the Pharmaceutical SAS Users Group (PharmaSUG) and PhUSE. He is active in the Clinical Data Interchange Standards Consortium (CDISC) community, contributing to the development of the Analysis Data Model (ADaM), and he serves as an ADaM trainer for CDISC. Shostak received an MBA from James Madison University and a BS in statistics from Virginia Tech.

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more:
http://support.sas.com/publishing/authors/holland_chris.html

http://support.sas.com/publishing/authors/shostak.html

Chapter 1: Implementation Strategies

The Case for Standards

The decision to adapt to CDISC standards within an organization or for a particular clinical development program has gotten easier since Congress approved the FDA Safety and Innovation Act, or FDASIA, in July of 2012. As of December 2016, the implementation of CDISC standards, primarily the Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM), is required for certain studies contained within FDA marketing applications.   For the early adapters, this change will have no impact on their current processes. These are the organizations that clearly saw the benefits of adapting to CDISC as soon as possible. Mergers and acquisitions have persisted throughout the pharmaceutical industry for decades, and behind the scenes of each merger are the biostatisticians, data managers, and SAS programmers who've worked at the same desk year after year, but have seen their employer names 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.

Pharmaceutical and biotech companies weren't the only organizations undergoing mergers. During the late 1990s and early 2000s, many contract research organizations (CROs) consolidated as well. In addition to the numerous data standards that 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. For some time, FDA reviewers had 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 two-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 that they were prescribing, taking, or selling to treat depression, arthritis, or any number of spin-off indications were adding 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 clinical trial 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 appears 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 *Providing Regulatory Submissions in Electronic Format—Standardized Study Data* (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"
December 2014	Final FDA e-Data guidance. Started the 24-month clock for requiring NDAs, aNDAs, and certain BLAs and INDs to be made available electronically to CDER and CBER and in the required format as specified by the Data Standards Catalog (available at http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm)

Nowadays, the legion of CDISC implementers is tangible to any SAS user conference attendee who struggles 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 use CDISC data, FDA presenters are promoting their preference for CDISC, and FDA documents are requiring the implementation of SDTM and ADaM models for sponsors' NDA and BLA submissions. The question is no longer *whether* to implement CDISC standards. Rather it is now more a question of *when*, and how far to go with it.

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, benefit from having nearly instant familiarity with how data are 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, you also 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.

Many large organizations with high volume and adequate resources are now implementing CDISC standards as early as Phase 1, despite the historically low chance of a drug ultimately advancing to the marketing submission stage. For others, 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
- An introduction to SAS software tools that 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/standards/foundational/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 considerable work. Without these templates, one important element to a successful implementation is making sure that 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.

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:

- 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 opinion 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. Either the board can 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 be made either 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 work that 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 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 that 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 SDTMIG. It is full of details and instructions that should not be overlooked. For trickier problems, such as how to model data that don'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 that 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 difficult 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 with review tools that leverage the SDTM, 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.

- 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 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 the 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 website at http://www.cdisc.org/standards/foundational/adam. Although updated final versions to these models will be available by the time this edition is published, they will not impact this book.

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 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 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 dataset, 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.

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 following the basic principles for analysis datasets mentioned previously.

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.

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.