

## A Road Map to Successful CDISC ADaM Submission to FDA: Guidelines, Best Practices & Case Studies

Vikash Jain, Accenture Life Sciences, Berwyn, PA  
Sandra Minjoe, Accenture Life Sciences, Berwyn, PA

### Abstract:

Submitting a filing to the FDA (Food and Drug Administration) using Clinical Data Interchange Standards Consortium (CDISC) data standards has become a norm in the pharmaceutical industry in recent years. This standard has been strongly encouraged by the agency to help expedite their review process, plus it also helps the sponsors and service providers have a means of efficient collaboration using these common industry standards.

This paper elaborates on the following fundamental and core components to be considered for the ADaM (Analysis Data Model) piece of submission:

- 1) Use of ADaM or ADaM-like data
- 2) Checking for ADaM compliance
- 3) Use of Define and other Metadata

We also present handful of case studies based on real time CDISC submission project experiences while collaboration with our sponsors.

### Introduction:

The Analysis Data Mode (ADaM)<sup>1</sup> supports efficient generation, replication, and review of analysis results. It includes fundamental principles and standards to follow in the creation of analysis datasets and associated metadata. Metadata are “data about the data” or “information about the data.”

The design of analysis datasets is driven by the scientific and medical objectives of the clinical trial. A fundamental principle is that the structure and content of the analysis datasets must support clear, unambiguous communication of the scientific and statistical aspects of the trial.

In adopting the principles and standards of ADaM when constructing analysis datasets and their associated metadata, it cannot be emphasized enough that early and effective communication between reviewers or other recipients of the data and sponsors is essential if the full benefits of analysis datasets are to be achieved.

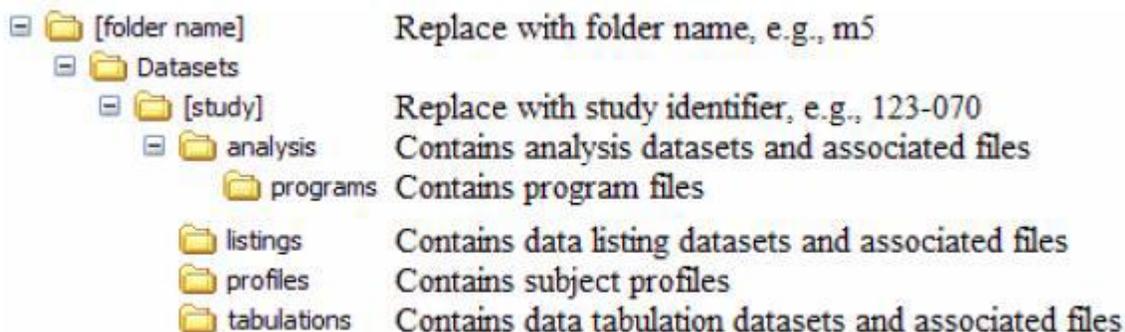
Data standards bring about more than just savings in time and money, companies will see additional benefits when they implement CDISC standards:

- Communication among project teams and partners is easier
- A greater level of accuracy and less training with a constant process
- Decision making is simplified
- Scientists can do the science rather than being concerned with the data
- Easier transfer of data between partners
- Opens up a wider choice of tools/technology (as long as they are standard-compliant)
- Mergers and Acquisitions are less daunting while integrating data
- Help Regulatory to review an application faster because of common standards and data structure being the same

In parallel with the development of clinical data submission guidance, the FDA has adopted the International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use standards for regulatory submissions and has issued a guidance document on the electronic Common Technical

Document (eCTD) as its framework for electronic submissions of pharmaceutical product applications. Revision 2 of this guidance<sup>2</sup> was posted in 2008.

According to FDA guidance documents on the eCTD<sup>2</sup>, submitted data can be classified into four types: 1) data tabulations, 2) data listings, 3) analysis datasets, and 4) subject profiles. These are collectively referred to as Case Report Tabulations (CRTs). The specification for organizing datasets and their associated files in folders within the submission is summarized in the following figure, from the Study Data Specifications document<sup>2</sup>.



As of April 18<sup>th</sup>, 2014, 2013, there were 448 documents on [fda.gov](http://fda.gov) that mention CDISC, including

- the Study Data Standards Resources website<sup>3</sup>, which references many CDISC models, including ADaM
- CDER Common Standards Issues document (December 2011)<sup>4</sup>, which includes references to CDISC SDTM and ADaM
- Several draft documents out for public comment

Two of the main issues FDA is trying to address by using standards are

- Speed of individual submission review
- Ability to combine data for cross-company safety reviews

Many of the sponsors we've worked with have had FDA reviewers request, at some point during their submission process, that their submission analysis data be in the ADaM structures. The most important question is: How do we determine whether the submission data is truly CDISC ADaM compliant, or if it will put the submission at risk for FDA review and approval?

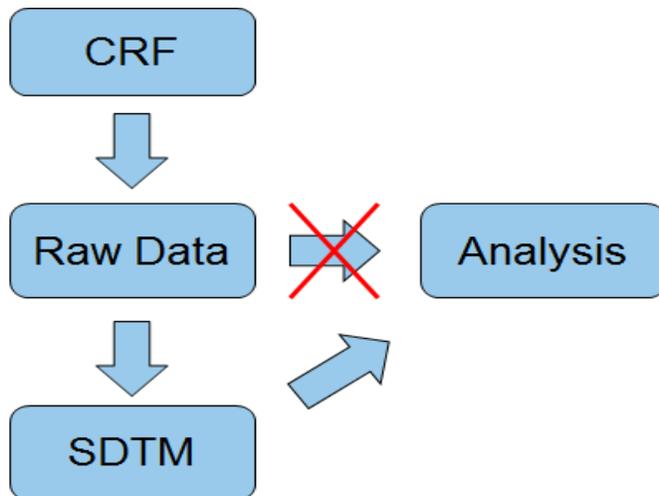
### What does FDA expect in submissions?

Some of the points made in the CDER Common Standards Issues document (December 2011)<sup>4</sup> document are:

- Analysis datasets must be submitted in addition to SDTM
- ADaM is the preferred standard structure
- ADSL is expected
- Fix or explain any OpenCDISC errors and warnings
- Denote in the define file the dataset(s) containing primary efficacy
- Include a Data Reviewers Guide to explain things a reviewer needs to know

- ADaM datasets should be derivable from SDTM (Not from other internal raw data)

The FDA document titled “CDER/CBER’s Top 7 CDISC Standards Issues”<sup>5</sup> includes the following logical data flow diagram, describing what to do/what not to do when creating analysis datasets:



It seems that some sponsors have submitted analysis datasets that were derived from their raw data, not their SDTM data. As the diagram shows, data should be mapped from SDTM to the analysis datasets.

### **What Is “True ADaM Data”?**

The following criterion must be met to be called true CDISC ADaM data:

- Meets ALL specifications as mentioned in the ADaM documents<sup>1</sup>
- Each of the datasets must contain all the variables needed for analysis and usability
- The data passes compliance as per the CDISC validation checks
- Each dataset structure is one of the ADaM classes, currently ADSL, BDS (includes time-to-event), or ADAE. Very few datasets will be exceptions with a class of OTHER, and are only those that can't be done using an ADaM structure.

In addition, “True ADaM Data” meets all the ADaM fundamental principles of analysis datasets and related metadata as described in the ADaM document<sup>1</sup>:

**Analysis datasets and their associated metadata must:**

- facilitate clear and unambiguous communication
- provide traceability between the analysis data and its source data (ultimately SDTM)
- be readily useable by commonly available software tools

**Analysis datasets must:**

- be accompanied by metadata
- be analysis-ready

### **What Is “ADaM-like Data”?**

“ADaM-like” data probably follows fundamental principles, described above, but does not meet some structure specifications. For example, ADaM-like data might deviate from ADaM in any of the following ways:

- ADSL includes extra variables, such as efficacy results
- Data class is BDS but uses SDTM variables instead of parameters
- Missing some required variables
- Contains parameters not used in analysis
- Incorrect labels used
- Fails compliance checks

Because of these non-conformities, ADaM-like data likely:

- Doesn't allow use of CDISC-specific tools
- Fails many OpenCDISC checks that then need to be explained
- Isn't in a familiar structure ... to anyone
- Is likely to confuse (and maybe even alienate) a reviewer

### **ADaM vs. ADaM-like Data: Pros/Cons**

In earlier years, as companies started trying to apply SDTM in-house, they often used an “SDTM-like” structure rather than true SDTM. This seems to have caused more problems than is solved, as described in the CDER Common Data Standards Issues document<sup>4</sup>.

CDER has accepted SDTM datasets since 2004; however, due to differences in sponsor implementation of the standard, CDER has observed significant variability in submissions containing “standardized” electronic clinical trial data. CDER has received numerous “SDTM-like” applications over the past several years in which sponsors have not followed the SDTM Implementation Guide. Furthermore, aspects of particular sponsor implementations have actually resulted in increased review difficulty for CDER reviewers. In addition, some sponsors have wrongly

It is most important to give the review what is needed to perform the review: data that can be easily understood, reproduced, traced back to SDTM, and used to generate analysis results. Data that is “ADaM-like” but doesn't meet

these reviewer needs may end up requiring additional work post-filing, such as submission of additional datasets, programs, and other documentation.

### **“ADaM-Like” Case Study Summary:**

Issue:

- Convert two Phase II legacy studies to SDTM and ADaM
- Complete Analysis & Reporting for two Phase III studies and ISS/ISE
- Compile NDA electronic submission

Scope:

- SDTM domains and ADaM datasets for legacy studies required extensive follow-up to 3<sup>rd</sup> party vendor to resolve data inconsistencies due to “SDTM-like” and “ADaM-like” data structures
- Redo TFL reconciliation to make sure the numbers match the previously reported analysis results
- Redoing work, including reconciling differences, is a huge added time and cost expense

Recommendation:

- It would be much smoother to proactively create SDTM, ADaM datasets and TLFs for Phase III studies and Integrated Summaries.
- Much easier to use the data in standard structures from the beginning

### **Checking for ADaM compliance**

Part of the process of creating ADaM data is to confirm that it is compliant. Determining ADaM compliance is a multi-step process, and usually involves some sort of automated component ADaM compliance represents traceability of analytical results back through data elements that were used and a set of rules which are designed to establish commonality in the way data are described.

The ADaM Compliance Team developed a list of compliance checks which can be implemented via common software. This was accomplished by focusing on objective criteria which can be assessed solely from datasets. For lack of a better term, we refer to these as “machine-readable checks.” The question of “which machine” and “which software” isn’t explicitly addressed, but instead generic statements are provided that can be easily coded into any software language. For example, SAS® is commonly used within clinical research, but isn’t the only software choice. Also, with the introduction of XML, HL7 standards and cutting edge technology, the landscape is likely change. With all this in mind, the checks were developed to be technology-neutral.

Automated ADaM Compliance checks allow us to easily assess the following:

- Overall ADaM compliance
  - Fundamental Principles: Traceable, Analyzable, Useable, Readable
  - Same SDTM variable name = same meaning/content/attributes
  - Co-dependency of variables (e.g., PARAM/PARAMCD)
- Dataset structure compliance
  - ADSL
  - BDS
  - ADAE

- Variable content compliance
  - Type (num or char)
  - Controlled terminology

Compliance checking can be done using OpenCDISC Validator, a company's home grown compliance tool, and manual review. A combination of all three will give the most robust compliance.

### **Compliance Check Case Study Summary:**

Scope: Sponsor shares submission level CDISC ADaM datasets with Programming group for compliance review, which were been developed by 3<sup>rd</sup> party CRO

Examples of some issues discovered:

- Variable must be a max of 8 characters, start with a letter A-Z and not contain anything other than letter, numbers, underscore (\_)
- Variable Label in the dataset should match the variable label described in ADaM Implementation Guide
- Variable Type in the dataset should match the variable type described in ADaM Implementation Guide.
- At least one analysis value (AVAL or AVALC) is not present in BDS
- Variables described in ADaM as Required must be included in the dataset
- \*DT variable SAS format in the dataset must be YYYYMMDD. or YYMMDD10. or DATE9.

Resolution:

- Worked with 3<sup>rd</sup> party vendor who developed the ADaM datasets to correct metadata for compliance and readability

### **Uses of Define and Metadata**

As mentioned in the introduction of this paper, metadata is "data about data". The term is ambiguous, as it is used for two fundamentally different concepts (types). Structural metadata is about the design and specification of data structures and is more properly called "data about the containers of data"; descriptive metadata, on the other hand, is about individual instances of application data, the data content.

CDISC metadata often uses HTML or XML and can be stored in Metadata Repository (MDR).

ADaM metadata facilitates communication by providing specification of details and links between the general description of the analysis (as found in the protocol's data analysis section, SAP, or the reported analysis methods), the analysis results, and the data used in the analysis, and the SDTM domains.

ADaM metadata can be categorized into 4 levels:

- Analysis Metadata
- Variable Metadata
- Parameter Value level Metadata
- Results Level Metadata

The first 3 levels of metadata are focused on the data. However, results-level metadata is focused on the analysis output itself.

The metadata structures described in the ADaM document<sup>1</sup> are based on the Case Report Tabulation Data Definition Specification Standard, version 1.0.0 (CRT-DDS).

**define.XML version and background:**

- Version 1.0 (Case Report Tabulation Data Definition Specification)
  - Has been out since 2005
  - Includes only dataset-level and variable-level metadata, and all else requires extension
  - Was designed and used for SDTM data transport
  - Is accepted by FDA
- Version 2.0<sup>6</sup>
  - Came out in Feb, 2013
  - Also includes parameter (value) –level metadata
  - Can be used for ADaM data transport
  - Accepted by FDA

Many companies use Microsoft Excel® to document and create metadata, in the forms of specifications and define documents. For better control of the metadata, in terms of both traceability and reusability, companies now seem to be moving to a metadata repository.

**Define and Metadata Case Study #1 Summary:**

Scope: Prior to FDA submission, client asked us to review analysis datasets for compliance and review define file for readability

Issues:

- Compliance issues were discovered in variable metadata as per ADaM implementation guide rules
- The Source/Derivation/Comment column within define file was not readable from reviewers point of view (specs vs. description)

Resolution: Worked with 3<sup>rd</sup> party vendor who developed the ADaM datasets to correct metadata for compliance and readability

Lesson Learnt: ADaM dataset define file adjustments required extended timeline, additional resourcing, and rework

**Define and Metadata Case Study #2 Summary:**

Scope: During submission critical path, client asked us to reconcile define metadata with actual ADaM Analysis Dataset

Issue: Found some variables in ADaM analysis datasets were not captured in the define file

Resolution: Recommended the define file be updated to include the missing variables

Lesson Learnt: Reconcile metadata with data throughout the development rather than at publishing stage.

**Define and Metadata Case Study #3 Summary:**

Scope: After delivery of ADaM deliverables, client reviewed ADaM and asked us to:

- Re-order the variables alphabetically, both in datasets and Define
- Show only ADSL variables once in the Define document
- Remove the ADSL variables from the description of all other datasets, even though they are present in the dataset

Issue: The above request contradicts the recommendation provided by ADaM data model and study data specification documents

Resolutions:

- We should not order the variables alphabetically in the analysis dataset/s
- The define file must match the dataset. That means if a variable exists in the dataset, it must be in the define
- The ADaM document<sup>1</sup> talks about variable order. Here is a quote from section 4.1.3 on page 11:  
*“Ideally, the ordering of the variables in the analysis dataset follows a logical ordering (not simply alphabetic). Refer to the FDA “Study Data Specifications” [7] for more information regarding the ordering of variables in the analysis dataset. It is recommended that the sponsor define a convention for ordering of variables within a dataset and then apply this ordering consistently for all analysis datasets. The ordering of the variables within a dataset should match the order of the variables as presented in the define file.”*
- Regarding variables that should be included in analysis datasets, FDA’s Study Data Specifications v2 document<sup>7</sup> says in section 2.7.2 on pages 7-8:  
*“Core variables should be listed after the key variables (USUBJID and visit) and included on each analysis dataset. Core variables include study/protocol, center/site, country, treatment assignment, sex, age, race, analysis population flags (e.g. ITT, safety) and other important baseline demographic variables.”*

## **Recommendations**

The following are our recommendations for ADaM implementation.

1. Ensure you at least meet ADaM Fundamental Principles
  - a. Traceable, Useable, Analysis-Ready, Metadata
2. Consider phased approach for structure adoption
  - a. ADSL first, as most companies are already creating a similar dataset
  - b. ADAE next, since it closely resembles SDTM AE, with a few additional variables
  - c. BDS last, since it requires a more ADaM knowledge
3. Run compliance checks on all ADaM structures, using OpenCDISC and/or internally-written checks, plus manual reviews
4. In addition to data, submit a define file and Data Reviewers Guide
5. Get familiar with industry standards in use/development
  - a. CDISC, including ADaM and other standards<sup>8</sup>
  - b. Computational Sciences Symposium Working Groups<sup>9</sup>
6. Consider contributing by providing feedback on drafts during public review periods and/or joining a team to help develop future standards
7. Keep up with FDA standards and updates at the FDA Study Data Standards website<sup>3</sup>, and sign up for email updates

## **References:**

- <sup>1</sup>. Analysis Data Model (ADaM), Implementation Guide (IG) and appendices - <http://www.cdisc.org/adam>.

2. FDA, 2008, "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Human pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications"  
<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064994.htm>
3. FDA Study Data Standards - <http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>.
4. CDER Common Data Standards Issues document -  
<http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf>.
5. CDER/CBER's Top 7 CDISC Standards Issues -  
<http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM291752.pdf>.
6. Define-XML 2.0 - <http://www.cdisc.org/define-xml>.
7. Study Data Specifications v2.0 -  
<http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM312964.pdf>
8. CDISC Standards - <http://www.cdisc.org/standards-and-implementations>.
9. Computational Sciences Symposium Working Groups -  
[http://www.phusewiki.org/wiki/index.php?title=CSS\\_Working\\_Groups](http://www.phusewiki.org/wiki/index.php?title=CSS_Working_Groups).

## **Contact:**

Your comments and questions are valued and appreciated. The authors can be reached at:

Vikash Jain  
Accenture  
1160 W. Swedesford Rd. Bldg. One  
Berwyn PA 19312  
vikesh.jain@accenture.com

Sandra Minjoe  
Accenture  
sandra.minjoe@accenture.com

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