PharmaSUG 2015 - Paper SS04

Begin with the End in Mind – Using FDA Guidance Documents as Guideposts when Planning, Delivering and Archiving Clinical Trials

David C. Izard, Accenture, Berwyn, Pennsylvania, USA

ABSTRACT

Clinical professionals, including Statisticians and Programmers, work tirelessly contributing to protocol & SAP development, CRF & clinical database design, analysis dataset design & implementation and the generation of tables, figures and listings in support of clinical trials. If the compound under study successfully navigates clinical and regulatory hurdles, most of these items will make their way into an U.S. Food & Drug Administration (FDA) submission. Historically there have been many sources of information on how to organize and present these assets for FDA consideration, many of which seemingly contradicted each other, making the requirements for regulatory submission cloudy at best. The FDA has recently changed this landscape, releasing final versions of two guidance documents with a companion technical specification document as well as a data standards catalog. The documents cover the requirement to file applications electronically, use and provide standardized study data as part of regulatory submissions, and provide clear and comprehensive technical considerations for the inclusion of both clinical & non-clinical data in regulatory filings.

This paper will examine the assets that are created during the execution of a clinical trial and hold them up against the submission requirements that are dictated in these recently finalized documents. It will also examine the paradigm shift occurring at the FDA, moving from a reactive to proactive submission planning and review environment and the impact that has, not just on preparing your clinical trial assets for submission, but on steps you must take during planning, execution and archiving of your trial to ensure you meet the FDA's expectations come submission time.

A BRIEF HISTORY OF REGULATORY GUIDANCE FOR CLINICAL DATA ASSETS

The US Food and Drug Administration (FDA) first articulated loose requirements for the organization and submission of both data and documents in electronic form via the guidance document “Guidance for Industry: Providing Regulatory Submissions in Electronic Format – General Considerations” [1], issued in January 1999. It was the culmination of activities surrounding the publishing of electronic records and electronic signatures regulation (21 CFR Part 11) in March 1997 and progress made by both the Centers for Drug (CDER) and Biologic (CBER) Evaluation and Research in 1998 to articulate what documents could be submitted electronically as well as the process for submitting electronically. While the bulk of this document addressed the nuances of preparing documents in the relatively new Portable Document Format (PDF), it devoted 2 pages to the introduction, format, content and provision of SAS® transport (XPT) files to hold clinical data when providing materials to the agency for a regulatory review. While SAS transport files had been existence for a number of years, the SAS Institute had recently made the transport file and open data exchange standard with the publication of a technical support document [2].

In October 2005 FDA issued the first version of their finalized guidance document on the electronic Common Technical Document (eCTD) [3] related to the International Conference on Harmonization’s efforts to define this concept. This guidance document made the first callout to the Study Data Specifications [4] via the following excerpt from Section III.E.4, “See the associated document “Study Data Specifications” for details on providing datasets and related files (e.g., data definition files, program files).” This guidance, first issued in July 2004 along with the draft eCTD guidance, would be revised eight times over the next seven years to include updates based on new standards coming available, increased file size values based on revised agency capability and other incremental improvements. This document evolved over time from a document that put some constraints on sponsor defined data structures to becoming a more vocal proponent of standards available from the Clinical Data Interchange Standards Consortium (CDISC). It served as the anchor by which our industry developed and submitted data and related documentation to the FDA during the years that CDISC standards grew in terms of depth, breadth and acceptance with sponsors, service providers and regulatory authorities.

Other documents were produced by the agency to reinforce their thinking about standards. The CDER Common Data Standards Issues Document [5], first issued in May 2011 and updated in December 2011, was the FDA’s initial formal reaction to the use and misuse of standards by sponsors and service providers alike. This document, not really a guidance document, more of a collection of thoughts on a subject, described a number of challenges the agency was having when reviewing data that was theoretically standardized and provided items to consider and agency preferences, particularly where the various CDISC standards were not completely clear.

A second document, a draft version of the Standardized Study Data [6] guidance document, was first produced during this time period, in February 2012. This early version was designed to share the initial thoughts of the agency on what study data standardization meant to them, and basic concepts that would appear in future guidance documents such as a Study Data Standardization Plan, considerations for legacy data conversion, expectations for validation & traceability and the benefits to electronically submitting data based on standards, both in terms of structural standards (e.g., SDTM or ADaM) and content standards (e.g. controlled terminology).

This collection of documents along with feedback delivered during regulatory review cycles and presentations made by FDA representatives laid the ground work for what was to come next, the eventual requirement to provide data in standardized format to the agency for the vast majority of regulatory submissions.

THE NEW GUIDANCE DOCUMENTS

The Food and Drug Administration Safety and Innovation Act (FDASIA) [7], signed into law in July 2012, expands the FDA's authorities and strengthens the agency's ability to safeguard and advance public policy by, among other things, reaffirming the authority to collect user fees to fund product reviews and promoting innovation to speed patient access to safe and effective products. This fifth reauthorization of the Prescription Drug User Fee Act, better known as PDUFA V, includes a number of objectives related innovation as documented in the "PDUFA V Information Technology/Informatics Plan" [8], which includes a number of specific objectives related to requiring the electronic submission of data in standardized formats:

Objective	Milestones	Target Date
Objective 1: Require the electronic submission of data in standardized formats.	**Milestone 1.1**: Publish final guidance requiring regulatory submissions in electronic format – Submissions Under Section 745A(a).	Q2 FY15
	Milestone 1.2: Publish final guidance requiring regulatory submissions in electronic format – Standardized Study Data.	Q2 FY15
	Milestone 1.3: Publish final Data Standards Catalog	Q2 FY15
	Milestone 1.4: Publish final Study Data Technical Conformance Guide	Q2 FY15
	Milestone 1.5: Publish therapeutic area standards initiative project plan, v2.0, for public comment	Q3 FY14
	Milestone 1.6: Require NDA, certain BLA and ANDA submissions of data in standardized formats.	Q2 FY17

Table 1. Excerpt from PDUFA V Information Technology/Informatics Plan

Reference for table contents – PDUFA V Information Technology/Informatics Plan [8], Section 4.1, Table 3

The FDA has met or beaten each of these target dates, having made the Data Standards Catalog [12], Milestone 1.3, available as early as April 2014 via the Study Data Standards Resources web page [15]. The three new guidance documents (Milestones 1.1, 1.2 & 1.4) were issued as draft documents in February 2014 and as final documents on December 17, 2014. One of the documents, the Study Data Technical Conformance Guide (Study Data TCG) [11], has since been updated, Version 2.1 was made available to the public on March 18, 2015.

The parent guidance in this series of documents is "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Submissions Under Section 745A(a) of the Federal Food, Drug and Cosmetic Act" [9]. The primary objective of this guidance is to affirm that, as soon as December 2016, you will need to submit most if not all INDs, NDAs, ANDAs and BLAs electronically as opposed to filing on paper. The general exceptions to this guidance are devices typically regulated by CBER as biological products, such as those for screening donated blood for transfusion-transmissible diseases and non-commercial INDs. As you can see, the exception list is very, very small.

The 2nd guidance document in succession is "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Study Data" [10]. Following on to the requirement that most if not all submissions must be electronic, this guidance goes on to state that studies initiated in the relatively near future must utilize specific data standards for the collection, analysis and delivery of clinical and non-clinical trial data and results as endorsed by the

FDA as documented in the Data Standards Catalog [12]. This requirement kicks in for studies that would support an NDA, ANDA or BLA on the 2 year anniversary of the guidance document becoming final (December 17, 2016) and one year later for INDs. Two noteworthy observations:

1. This start date requirement is for the initiation of a trial, not for the inclusion of all trial data in a regulatory submission. You will still have the opportunity to submit data in older, legacy formats after these dates based on the standards that existed at the time the trial was initiated. Of note, the definition of initiated is the earliest date a subject signed informed consent for that study.
2. The scope of this guidance is clearly both your clinical and non-clinical studies, meaning your non-clinical studies must both capture and report data in a manner consistent with the CDISC SEND model.

The 3rd guidance document is the "Study Data Technical Conformance Guide" [11]. While the first two guidance documents highlighted are binding guidances, meaning that the requirements laid out within are not optional, this guidance is truly a guidance, describing the technical details and practical considerations for implementing this set of guidance document requirements. This document replaces the Study Data Specifications [4] and CDER Common Data Issues Document [5], which have since been removed from the Study Data Standards Resource web page [15] but can still be found by performing a general search from outside the FDA web site.

SUBMISSION PACKAGE COMPONENTS AND CONCEPTS DICTATED BY THE NEW GUIDANCE DOCUMENTS

Piecing together the requirements stated throughout the new guidance documents, particularly the Study Data Technical Conformance Guide, dictates both the items that must be included in a regulatory data submission package as well as characteristics that those items should possess. This section will ask a number of questions and then provide answers in the context of the guidance documents and related companion guide and catalog.

DEFINITIONS OF DATA STANDARDS, SUBMISSION DATA PACKAGE TYPES AND CONTENT

What submissions will require standardized study data?

- New Drug Applications (NDAs)
- Abbreviated New Drug Applications (ANDAs)
- Certain Biologics License Applications (BLAs)
- Certain Investigational New Drug Applications (INDs)

  Reference – Standardized Study Data [10], Section II.A, 1st paragraph

What is the acceptable set of standards that can be utilized for these submission types?

- The set of acceptable standards that the agency accepts can be found in the Data Standards Catalog [12]

  Reference – Standardized Study Data [10], Section II.C, 1st & 2nd paragraphs

What types of data are subject to these standards?

- Tabulation datasets for clinical and non-clinical studies
- Analysis datasets for clinical studies

  Reference - Study Data TCG [11], Section 4.1, 2nd paragraph

What is the standard that should be utilized for clinical tabulation datasets?

- Tabulation datasets should be prepared based on the CDISC Study Data Tabulation Model (SDTM) (Study Data TCG [11], Section 4.1.1.1)

What is the standard that should be utilized for non-clinical tabulation datasets?

- Non-clinical tabulation datasets should be prepared based on the CDISC SDTM model as defined by the Standard for the Exchange of Non-clinical Data (SEND) Implementation Guide (IG) (Study Data TCG [11], Section 4.1.3.1)

- The scope of non-clinical tabulation data defined by SEND standards, as of today, includes single dose general toxicity, repeat dose general toxicity and carcinogenicity studies (Study Data TCG [11], Section 4.1.3.1)

What is the standard that should be utilized for analysis datasets?

- Analysis datasets should be prepared based on the CDISC Analysis Data Model (ADaM) (Study Data TCG [11], Section 4.1.2.1)

What makes up a tabulation data submission package?

- SDTM domain datasets prepared based on the SDTM IG (Study Data TCG [11], Section 4.1.1.1) for clinical data or the SEND IG (Study Data TCG [11], Section 4.1.3.1) for non-clinical data as SAS transport (XPT) files (Study Data TCG [11], Section 3.3.1)
- A data definition file in define.xml format (Study Data TCG [11], Section 4.1.4.5)
- A blank case report form (CRF) annotated to the tabulation data for clinical studies (Study Data TCG [11], Section 4.1.4.6)
- A Study Data Reviewer's Guide (Study Data TCG [11], Section 2.2)

What makes up an analysis data submission package?

- Analysis datasets prepared based on the ADaM IG (Study Data TCG [11], Section 4.1.2.1, also Standardized Study Data [10], Section II.C.2) as well as other necessary analysis datasets (Study Data TCG [11], Section 4.1.2.2) as SAS transport (XPT) files (Study Data TCG [11], Section 3.3.1, also Standardized Study Data [10], Section II.C.1)
- A data definition file in define.xml format (Study Data TCG [11], Section 4.1.4.5)
- An Analysis Data Reviewer's Guide (Study Data TCG [11], Section 2.3)
- Software programs used to create the analysis datasets, tables & figures (Study Data TCG [11], Section 4.1.2.8)

If I need to submit my legacy source data, what would be included in this submission package?

- Legacy format source datasets prepared based as SAS transport files, following the rules and conventions as highlighted in the section "Study Data Exchange Format for Data – SAS Transport Files" later in this document and augmented by the requirements present in Study Data Specifications [4], Section 2.5.
- A data definition file in define.pdf format (Study Data Specifications [4], Section 3.1.2.2)
- A blank case report form (CRF) annotated to the tabulation data for clinical studies (Study Data TCG [11], Section 4.1.4.6) NOTE: The information present in the Study Data Technical Conformance Guide supersedes the information on this same topic present in the Study Data Specifications [11], Section 5.4.

If I need to submit my legacy analysis data, what would be included in this submission package?

- Legacy format source datasets prepared based as SAS transport files, following the rules and conventions as highlighted in the section "Study Data Exchange Format for Data – SAS Transport Files" later in this document and augmented by the requirements present in Study Data Specifications [4], Sections 2.5, 2.7.1, 2.7.2 & 2.7.4.
- A data definition file in define.pdf format (Study Data Specifications [4], Section 3.1.2.2)
- Software programs used to create the analysis datasets, tables & figures (Study Data TCG [11], Section 4.1.2.8)

CHARACTERISTICS FOR ALL SUBMISSION DATA PACKAGES

Study Data Exchange Format for Data – SAS Transport Files

- SAS transport files should be created using PROC COPY with the XPORT engine in order to preserve the open SAS transport file standard published by SAS [2], not PROC CPORT/CIMPORT (Study Data TCG [11], Section 3.3.1, 2nd paragraph)

- SAS transport files should be smaller than 1 GB and should be split if they exceed 1 GB (Study Data TCG [11], Section 3.3.2)
    - See Study Data TCG [11], Section 4.1.1.3 for examples of how to use –CAT and –SCAT variable as the basis for splitting SAS transport files
    - See Study Data TCG [11], Section 7 for guidance on where to place the split files within your eCTD
- Study Data TCG [11], Sections 3.3.3 through 3.3.7 provide specifics on how dataset and variable names and label should be appropriately assigned
- Dataset labels must be unique across the submission packages for a single study (Study Data TCG [11], Section 4.1.2.7). For example, if you have adverse event data present in both your SDTM domains (AE) and your ADaM analysis datasets (ADAE), these datasets must have different labels ("Adverse Events" and "Analysis – Adverse Events", for example).

Data Definition Files

- External dictionary usage should be documented both in the data definition file as well as the Trial Summary (TS) SDTM trial design domain if using SDTM v1.2 / SDTM IG v3.1.2 or greater, SEND IG v3.0 or greater (Study Data TCG [11], Section 4.1.4.5, 1st paragraph)
- Provide high quality dataset labels that clearly describe the content of the dataset (Study Data TCG [11], Section 4.1.4.5, 1st paragraph)
- Provide separate data definition files in XML format as a properly functioning define.xml for SEND, SDTM & ADaM (Study Data TCG [11], Section 4.1.4.5, 2nd paragraph)
- If the data definition file is provided as a define.xml based on v2.0 or later of the standard ([13]) then a define.pdf version of the data definition file does not need to be provided, otherwise a PDF version of your data definition file must be produced and provided along with the data definition file in define.xml format (Study Data TCG [11], Section 4.1.4.5, 2nd paragraph)

Annotated Case Report Form

- The annotated case report form (annotated CRF or aCRF) should be provided as a PDF file and named "acrf.pdf" (Study Data TCG [11], Section 4.1.4.6, 1st paragraph)
- The CDISC Metadata Submission Guidelines ([14]) should be followed when producing this document (Study Data TCG [11], Section 4.1.4.6, 1st paragraph)
- The annotated CRF should include all available forms used to collect source data, including treatment assignment forms (Study Data TCG [11], Section 4.1.4.6, 2nd paragraph)
- The text "NOT SUBMITTED" should be annotated onto the case report form where data are recorded on the CRF but not submitted, and an explanation for why data was not submitted should appear in the Study Data Reviewer's Guide (Study Data TCG [11], Section 4.1.4.6, 3rd paragraph)

Study & Analysis Data Reviewer's Guides

- The Study Data Reviewer's Guide should describe any special considerations or directions that may facilitate an FDA reviewer's use of the submitted data and help the reviewer understand the relationship between the study report and the case report tabulation data (Study Data TCG [11], Section 2.2, 1st paragraph)
- The Study Data TCG does not define a specific template for the Study Data Reviewer's Guide, but an example of a Study Data Reviewer's Guide, including a template, completion guidelines and examples, can be found on the PhUSE website at http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer's_Guide (Study Data TCG [11], footnote #11 at bottom of Page 4)
- If submitting to CBER you need to continue to prepare the Data Interpretation and Validation Report (DIVR). It should be incorporated into the Study Data Reviewer's Guide. Information on the DIVR can be found at http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/ucm209137.htm.
- The Analysis Data Reviewer's Guide provides the FDA reviewers with context for analysis datasets and terminology; in purposefully duplicated limited information found in other submission documents in order to

provide FDA reviewers with a single point of orientation to the analysis datasets (Study Data TCG [11], Section 2.3, 1st paragraph).

- The Study Data TCG does not define a specific template for the Analysis Data Reviewer's Guide, but an example of an Analysis Data Reviewer's Guide, including a template, completion guidelines and examples, can be found on the PhUSE website at http://www.phusewiki.org/wiki/index.php?title=Analysis_Data_Reviewer's_Guide. (Study Data TCG [11], footnote #13 at bottom of page 5)

Software Programs

- Software programs can be provided as either ASCII text files or PDF documents (Study Data TCG [11], Section 4.1.2.8)
- Software programs should include sufficient documentation to allow the reviewer to understand the submitted programs (Study Data TCG [11], Section 4.1.2.8)
- If a submitted software program created by an analysis tool uses a filename extension other than ".txt" (ASCII) or ".pdf" (PDF) then the file name should include the native extension (Study Data TCG [11], Section 4.1.2.8). For example, if you intended to include the program "build_dm.sas" in your submission, an acceptable file name for this program in your submission could be "build_dm_sas.txt" or "build_dm.sas.txt".

SPECIAL CONSIDERATIONS

There are a number of details regarding specific implementation considerations for the CDISC SEND, SDTM & ADaM models sprinkled throughout the Study Data Technical Conformance Guide but primarily located in Section 4. Many of these details focus on concepts and items that, in the literal sense, the CDISC models and implementation guides regard as "permissible" or "optional" but, in the eyes of the FDA, are essential for an efficient and effective review. Some examples of items included in these sections are the importance of comprehensive, well formed SDTM Trial Design domains, the appropriate use of custom and supplemental qualifier SDTM domains, and the need for numeric versions of a number of concepts in ADaM analysis datasets that appear in character form in SDTM domain datasets such as date, timing and coded representation of categorical variables.

The following is a list of topics that are not specific submission deliverables but are key characteristics of the clinical data that will be captured in a trial which will require both special care during the execution of the trial as well as documenting and archiving information in support of future submission work.

Use of Controlled Terminology

- The FDA acknowledges that clinical trials are run over time and that terminology versions will likely differ from one study to the next as a result (Study Data TCG [11], Section 6.1.2, 1st paragraph)
- The FDA expects sponsors to use the most current version of an FDA-supported terminology available at the time of coding (Study Data TCG [11], Section 6.1.2, 1st paragraph)
- Any special coding considerations should be documented in the Study Data Reviewer's Guide (Study Data TCG [11], Section 6.1.2, 2nd paragraph)
- The FDA expects sponsors to utilize controlled terminology wherever available and, if the controlled terminology does not contain the information needed by the sponsor for submission, it is incumbent upon the sponsor to work with the controlled terminology maintenance organization with enough lead time to ensure that necessary controlled terminology is available at the time of regulatory filing (Study Data TCG [11], Section 6.1.3)
- The FDA expects sponsors to utilize the following controlled terminologies for the following items and documented in the following locations with the SDTM domain datasets:

Concept	Controlled Terminology for Concept	Location in SDTM
Adverse Events Medical History	MedDRA	AE domain MH domain
Concomitant Medications	WHODrug	CM domain

Concept	Controlled Terminology for Concept	Location in SDTM
Study Medication	FDA Unique Ingredient Identifier (UNII)	TS trial design domain • Investigational product (TSPARM=TRT or TRTUNII) • Active comparator (TSPARM=COMPTRT) • Background Treatments (TSPARM=CURTRT)
Pharmacologic Class	Veterans' Administration's National Drug File – Reference Terminology (NDF-RT)	TS trial design domain • Pharmacologic class (TSPARM=PCLAS)
Indication	SNOMED Clinical Terms	TS train design domain • Indication (TSPARM=INDIC) • Diagnosis Group (TSPARM=TDIGRP)

Table 2. FDA Expected Use of Controlled Terminology

Reference for table contents – Study Data TCG [11], Section 6, specifically Sections 6.3 through 6.6 with subsections

Validation

- The FDA recognizes two types of validation rules (Study Data TCG [11], Section 8.2.1):
    - Conformance validation - conformance of the data and related documentation to a standard
    - Quality checks – ensuring the data will support meaningful analysis
- The FDA has published sets of data validation rules – one each for SEND and SDTM – that encompass both conformance and quality checks (Study Data TCG [11], Section 8.2.2, 1st paragraph)
- Sponsors should validate their data prior to submission using the published validation rules and either correct the validation errors or explain in the Study Data Reviewer's Guide why certain validation errors could not be corrected (Study Data TCG [11], Section 8.2.2, 2nd paragraph)

Traceability & Legacy Data Conversion

- Understanding the provenance of data (i.e. traceability of the sponsor's results back to the collected data) is an important component of regulatory review (Study Data TCG [11], Section 8.3.1, 1st paragraph)
- During the transition period from legacy data standards to the collection, analysis and submission of clinical assets based on standardized study data, the FDA recognizes that some study data many not conform to FDA-supported study data standards and may need to be converted (Study Data TCG [11], Section 8.3.1, 3rd paragraph)
- "Based upon reviewer experience, establishing traceability is one of the most problematic issues associated with legacy study data converted to standardized data. If the reviewer is unable to trace study data from the data collection of subjects participating in a study to the analysis of the overall study data, then the regulatory review of a submission may be compromised." (Study Data TCG [11], Section 8.3.1, 2nd paragraph)
- In order to mitigate risk associated with legacy data conversion, sponsors should:
    - Prepare and submit a Legacy Data Conversion Plan & Report (Study Data TCG [11], Section 8.3.2.2, item #1 under the 1st paragraph). NOTE: A template with completion guidelines and examples is under development by PhUSE as part of the PhUSE/FDA CSS Working Groups Initiative and should be available by the end of 2015.
    - Prepare and submit a CRF annotated to the legacy source data (Study Data TCG [11], Section 8.3.2.2, item #2 under the 1st paragraph)
    - Incorporate the Legacy Data Conversion Plan & Report into the Study Data Reviewer's Guide in order to record significant data issues, clarification and explanations of traceability (Study Data TCG [11], Section 8.3.2.2, item #3 under the 1st paragraph)

- Prepare and submit the legacy data (Study Data TCG [11], Sections 8.3.2, 1st paragraph, and 8.3.2.2, item #4 under the 1st paragraph)

Pooling & Harmonization in Support of Integrated Analysis

- While this concept applies to data within a study, it is important to evaluate this situation when creating your sets of pooled data – When submitting clinical and non-clinical data, sponsors should not mix different versions of the SEND, SDTM and/or ADaM standards within their set of pooled data (Study Data TCG [11], Section 4.1.4.4)
- "Regardless of the specific versions [of controlled terminology] used for individual studies, pooled analyses of coded terms across multiple studies (e.g., for an integrated summary of safety) should be conducted using a single version of a terminology" (Study Data TCG [11], Section 6.1.2, 2nd paragraph)

DEPARTMENTAL / SERVICE PROVIDER RESPONSIBILITIES AND OBLIGATIONS

The FDA Guidance Documents [9,10,11] thoroughly describe both assets that must be included in your regulatory submission as well as principles by which these assets are created. In light of these requirements, coupled with the requirement to plan and discuss the use of standardized study data elements early in the development of a product (Study Data TCG [11], Section 2.1), the agency expectation is that you put considerable effort into creating these assets per their requirements and that you protect these assets in terms of appropriately archiving these items as your study progresses so you are prepared to call on these assets come submission time. It is in these steps that sponsors typically fall down on their responsibilities. In their efforts to reduce costs and accelerate timelines associated with the conduct and reporting of each individual clinical trial, sponsors tend to lose sight of the need to draw on these assets come submission time.

A second consideration regarding accounting for clinical assets during study planning, execution and archiving involves engaging service providers to provide some if not all of these assets on a sponsor's behalf. Frequently contracts and timelines are written, again, with the focus on timely study execution and clinical results without the appropriate level of regard to contractually describing each of these deliverables and characteristics these deliverables should have.

Here is an asset by asset guide to consider when planning, executing and archiving a clinical trial in order to have necessary items available come submission time:

STUDY STARTUP – PROTOCOL AND STATISTICAL ANALYSIS PLAN (SAP)

The following consideration should be made with respect to both of these documents:

- Ensure that any key lists, characteristics or analyses articulated by either of these documents utilize terminology and concepts consistent with CDISC controlled terminology and standard dictionaries as highlighted in the Study Data Technical Conformance Guide [11]. Pay particular attention to any of the concepts expected to be articulated in the Trial Summary (TS) SDTM trial design domain.

DATA COLLECTION / DATA MANAGEMENT

Characteristics that should be present / questions to ask

- Request that CDASH principles be employed when designing the data capture devices used for your study
- Request the CDISC controlled terminology as well as other dictionaries mandated by the FDA are utilized appropriately when designing both the database and the data capture devices used for your study.

Data Management Plan (DMP)

- Make sure the DMP adequately documents a number of the metadata items required for submission, in this case primarily versions of controlled terminology as well as how the version of a particular controlled terminology might change during the course of study execution and how the final version will be noted in both the DMP as well as the final assets that are delivered.

Data Transfer Specifications

- See to it that you receive detailed specifications for data that is not collected directly on a case report form but will eventually be incorporated into your set of source data for the clinical trial. Some typical examples of this type of data would be laboratory data coming from a central laboratory, ECG data coming from an ECG

vendor and PK data (both concentration and parameter data) coming from a PK service provider. The specifications will be very important when documenting these items in your data definition file.

Blank CRF

- Ensure you have archived the final version of the CRF as well as any previous versions of the CRF used in the clinic that had an impact on how data was collected or analyzed. A primary and frequent example of this is when the inclusion / exclusion criteria change during the course of a trial, you need to know what version of these criteria were used to evaluate a subject for study participation. You will need to include all relevant pages from the versions of CRFs over time in your submission packages.
- Acquire the Blank CRF as a native PDF document generated by the application used to produce the CRF. Avoid scanned copies.
- Make sure you acquire all pages used to capture data related to your clinical trial, including but not limited to the pages used in the clinic, pages from any patient reported outcomes devices and pages used to randomize and/or stratify subjects in the trial.

Blank CRF annotated to the source data

- Even if your study was designed based on CDISC standards, particularly CDASH for data collection, make sure you get the blank CRF annotated to the source data as it came out of the clinical database management system. This was you will be able to both defend what was truly captured in the clinic versus what steps were taken to make the data SDTM compliant when producing the tabulation submission data package.

TABULATION (SDTM) SUBMISSION PACKAGE DEVELOPMENT

SDTM domain datasets

- You should receive and archive the final version of your SDTM domain datasets associated with each reporting effort / regulatory submission for your study. Whether they are initially delivered to you as SAS datasets, SAS transport files or the data format native to your CDMS is not as important as ensuring you are receiving a final product.
- Ensure that your set of source data in SDTM format is inclusive of all data expected to be included in your SDTM submission package. One key example is to make sure you have received both your pharmacokinetic concentration (PC) and parameter (PP) data as SDTM domains for studies where pharmacokinetic data was captured in the clinic.

SDTM mapping specifications

- Even if you are capturing your source data based on CDASH standards the SDTM mapping specifications will be essential when it comes time to populate your comments and metadata in the data definition file you produce for your study. Ensure they are developed and maintained throughout the course of your study and that they are in synch with the final SDTM domains at the conclusion of your study.

SDTM domain dataset metadata

- Consider all of the metadata information that is required for your data definition file and ensure that you have received it as part of the DMP or via other means. Examples would include versions of external dictionaries, ISO specifications and CDISC controlled terminology utilized for your study.

SDTM domain dataset validation results

- The validation results, be they generated by a standard tool such as OpenCDISC or proprietary software written by your organization taking the FDA's validation rules into account, should be executed and captured at the time of study data finalization in order to accurately record the thinking present at the time the data was being generated.

Blank CRF Annotated to the SDTM domain datasets

- Ensure that the same blank CRF described above, including all unique pages relevant to the capture of data in the clinic, has been annotated to the SDTM domain datasets following the standards laid out in the CDISC Metadata Submission Guidelines [14].

Data Definition File for SDTM domain datasets

- Make sure this item has been assembled and note the version of the define.xml standard that has been used to confirm if you need to also produce a PDF version of the data definition file.

Study Data Reviewer's Guide

- Again, the best time to produce this document is at the time the study is finalized as you have the opportunity to capture the current thinking with respect to both study conduct as well as validation issues that were captured, the primary content item communicated by this document.

Programs Used to Create SDTM Domains (Legacy Data Conversion)

- If you have generated your SDTM domain datasets based on source data from your CDMS that is not CDASH conformant you should request, receive and archive the software programs used to construct your SDTM domain datasets. In this scenario the agency is equally as likely to request these programs much like they require sponsors to submit their analysis dataset programs to review &/or verify logic employed by these programs.

ANALYSIS SUBMISSION DATA PACKAGE PLUS REPORTING

ADaM analysis datasets

- Similar to the SDTM domain datasets, you should receive and archive the final version of your ADaM analysis datasets associated with each reporting effort / regulatory submission for your study. Whether they are initially delivered to you as SAS datasets or SAS transport files is not as important as ensuring you are receiving a final product.

ADaM mapping specifications

- Again, similar to the SDTM domain datasets, the ADaM mapping specifications will be essential when it comes time to populate your comments and metadata in the data definition file you produce for your study. Ensure they are developed and maintained throughout the course of your study and that they are in synch with the final ADaM analysis datasets at the conclusion of your study.

ADaM analysis dataset metadata

- While typically smaller in scope, ensure that you capture any metadata necessary for producing your data definition file somewhere, be it in your protocol, SAP or some other direct means, prior to or at the conclusion of your study.

ADaM analysis dataset validation results

- The validation results, be they generated by a standard tool such as OpenCDISC or proprietary software written by your organization taking the FDA's validation rules into account, should be executed and captured at the time of study data finalization in order to accurately capture the thinking present at the time the data was being generated.

Data Definition File for ADaM analysis datasets

- Make sure this item has been assembled and note the version of the define.xml standard that has been used to confirm if you need to also produce a PDF version of the data definition file.

Analysis Data Reviewer's Guide

- Again, the best time to produce this document is at the time the study is finalized as you have the opportunity to capture the current thinking with respect to both study conduct as well as validation issues that were captured, the primary content item communicated by this document.

Software Programs Used to Create Analysis Datasets, Tables & Figures

- It is important that you receive and archive these software programs at the time the study as finalized as their location tends to be very well known at that time. Do not worry about the format of the programs as they are delivered versus the format and naming requirements for inclusion in your eCTD, that can be addressed as the submission publishing is executed.

- Ensure you receive and archive both driver programs as well as other programs called by these driver programs including common utility macros, particularly if these utility macros contain the logic for any key derived variables.

CONCLUSION

It took the FDA fifteen years to get from accepting clinical data in electronic format to setting expectations that require both clinical and non-clinical data to be conceived, recorded, analyzed, provided and archived based on specific endorsed standards. While the industry has two +/- years to enact plans to support these activities they have, in reality, known for nearly a decade that the agency would want, at some point, data and documentation that was well thought-out, properly documented formatted and clearly traceable from collection through to provision. Despite this lead time the industry will be scrambling to undo the challenges introduced by habit, convention, cost and the time necessary to enact this change properly. Hopefully this paper will serve to show that, by embracing this change at the point that data is conceived and keeping track of all of the assets created during the design, execution, analysis and archiving of a clinical trial, meeting FDA expectations will become a routine part of life sciences development work.

REFERENCES

[1] US Department of Health and Human Services, U.S. Food and Drug Administration. "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – General Considerations", Issued January 1999. Available at http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm072390.pdf, verified March 31, 2015. (Really! You can still get it!)

[2] SAS Institute, Inc. "TS-140: RECORD LAYOUT OF A SAS® VERSION 5 OR 6 DATA SET IN SAS® TRANSPORT (XPORT) FORMAT". Available at http://support.sas.com/techsup/technote/ts140.pdf, verified April 23, 2015.

[3] US Department of Health and Human Services, U.S. Food and Drug Administration. "Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications", Revision 2, Issued June 2008. Available at http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm072349.pdf, verified April 16, 2015.

[4] US Department of Health and Human Services, U.S. Food and Drug Administration. "Study Data Specifications", Issued July 18, 2012, Version 2.0. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM312964.pdf, verified March 31, 2015. (Still out there!)

[5] US Department of Health and Human Services, U.S. Food and Drug Administration. "CDER Common Data Standards Issues Document", Version 1.1, Issued December 2011. Available at http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf, verified April 16, 2015. (Also still out there!)

[6] US Department of Health and Human Services, U.S. Food and Drug Administration. "Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Standardized Study Data", Preliminary draft, Issued February 2012. Available at http://www.fda.gov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/General/UCM292169.pdf, verified April 16, 2015. NOTE: This is a link to the D*R*A*F*T version of this guidance document as issued in February 2012. The recently finalized version is available via a reference appearing later in this list.

[7] 112th Congress of the United States of America, 2nd Session, commencing January 3, 2012, S.3187, short title of "'Food and Drug Administration Safety and Innovation Act". Available at http://www.gpo.gov/fdsys/pkg/BILLS-112s3187enr/pdf/BILLS-112s3187enr.pdf, verified April 22, 2015.

[8] US Food and Drug Administration, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. "PDUFA V IT/Informatics Plan FY 2013 – FY 2017", update issued September 2014. Available at http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM416711.pdf, verified April 22, 2015.

[9] US Department of Health and Human Services, U.S. Food and Drug Administration. "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Submissions Under Section 745A(a) of the Federal Food, Drug and Cosmetic Act", Issued December 17, 2014. Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM384686.pdf, verified March 31, 2015.

[10] US Department of Health and Human Services, U.S. Food and Drug Administration. "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Study Data", Issued December 17, 2014. Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf, verified March 31, 2015.

[11] US Department of Health and Human Services, U.S. Food and Drug Administration. "Technical Specifications Document: Study Data Technical Conformance Guide", Version 2.1, Issued March 18, 2015. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf, verified March 31, 2015.

[12] US Department of Health and Human Services, U.S. Food and Drug Administration. "Data Standards Catalog", Version 4.0, Issued December 10, 2014. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx, verified March 31, 2015.

[13] Clinical Data Interchange Standards Consortium. "CDISC Define-XML Specification", Version 2.0, Issued March 5, 2013. Available at http://www.cdisc.org/define.xml, ZIP file containing printable standard named http://www.cdisc.org/system/files/members/standard/define_xml_2_0_releasepackage20140424.zip, verified April 13, 2015.

[14] Clinical Data Interchange Standards Consortium. "Study Data Tabulation Model Metadata Submission Guidelines (SDTM-MSG)", Version 1.0, Issued December 31, 2011. Available at http://www.cdisc.org/sdtm, ZIP file containing printable standard named http://www.cdisc.org/system/files/members/standard/application/zip/final_metadata.zip, verified April 13, 2015.

[15] US Department of Health and Human Services, U.S. Food and Drug Administration. Study Data Standards Resources web site. Available at http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm#study_data_stnd_dev, verified April 23, 2015.

ACKNOWLEDGMENTS

The author would like to thank Accenture, particularly Tom Lehmann, Matt Stroschein and Ellen Semple, for their leadership roles which allow me to participate in work that makes it possible for me to develop and present on material of this nature, and many of my coworkers, including but not limited to Fred Wood, Sandra Minjoe, Jerry Salyers, Kristin Kelly, Brian Hermann, Denise Esposito and Ashli Davis, for keeping me honest and on track in these endeavors!

RECOMMENDED READING

- Everything listed in the "References" section of this paper, but especially the Study Data Technical Conformance Guide (reference #11 above)
- FDA presentations on any of these subjects, which extend the thinking of the agency. Key presentations from the FDA's recent participation in the FDA/PhUSE Computational Sciences Symposium meeting, held March 16th & 17th, 2015, are available here: http://www.phuse.eu/CSS-Presentations2015.aspx .

CONTACT INFORMATION

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

Name:	David C. Izard
Enterprise:	Accenture
Address:	1160 Swedesford Road
City, State ZIP:	Berwyn, PA 19312
Mobile Phone:	(484) 467-0193
E-mail:	david.c.izard@accenture.com