PharmaSUG 2016 - Paper SS02

Preparing Legacy Format Data for Submission to the FDA
When & Why Must I Do It, What Guidance Should I Follow?

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

ABSTRACT

The U.S. Food & Drug Administration (FDA) released a number of binding guidance documents and companion materials that require clinical studies initiated on or after December 17, 2016 to utilize FDA endorsed data standards at the time the study is planned and executed if you intend to include the study as part of a future New Drug Application (NDA), Abbreviated New Drug Application (ANDA) or Biologics License Application (BLA). These guidance documents spend considerable effort documenting these new requirements but give little consideration to the body of clinical data that currently exists in legacy format. Furthermore, all previous guidance documents have now been deprecated in favor of these new, forward-looking guidances, leaving a void for how a Sponsor or Service Provider should prepare legacy data and related documentation for regulatory submission let alone when the submission of legacy format data is required or expected.

This paper will examine the agency's thinking on the role legacy format clinical data should play in a submission, drawing on the limited information available in current guidance as well as feedback from questions to the FDA posed at conferences and via the CDER & CBER eData email support services. It will also examine what constitutes a legacy format data submission and how one should utilize both current and legacy format guidance documents to prepare these assets for inclusion in a filing.

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 two new guidance documents (Milestones 1.1 & 1.2) as well as the Study Data Technical Conformance Guide (Study Data TCG) [11] were issued as draft documents in February 2014 and as final documents on December 17, 2014. The Study Data TCG has since been updated a number of times, the current version (v3.0) was just released on March 31, 2016.

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 document is the "Study Data Technical Conformance Guide" [11]. While the first two documents highlighted are binding guidances, meaning that the requirements laid out within are not optional, this document 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.

It is also worth mentioning that a 3rd binding guidance document, called out in the "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Submissions Under Section 745A(a) of the Federal Food, Drug and Cosmetic Act" [9], specifically the "Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications" [16] was finalized and published to the FDA web site on May 15, 2015. This is 3rd revision to the eCTD guidance first published in 2005 and updated in 2008 [3]. It, among other things, reinforces how and when pursuing waiver criteria is appropriate and refers to the Standardized Study Data guidance [10] for all things data when included in an eCTD based regulatory submission. It also provides links to a number of useful resources to consult when planning, compiling and submitting both data and related documents to the FDA as part of a regulatory filing.

WHAT IS LEGACY FORMAT DATA?

The FDA does not explicitly state what legacy format data is; rather, the agency clearly articulates the following information regarding their endorsed data standards:

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)

In the agency's eyes, if a set of data (source, tabulation &/or analysis) is not developed based on these FDA endorsed data standards, then it is regarded as legacy format data. With the exception of Section 8 of the Study Data Technical Conformance Guide [11], which is largely dedicated to ensuring traceability when legacy format data is part of the clinical data chain of custody, there are very few references to legacy data throughout this document:

- Section 4.1.4.6, on the subject of annotated case report forms included in a data publishing package, emphasizes that the name of the blank annotated case report form should be 'acrf.pdf' regardless if the data is standards based or legacy format.
- Section 7, which clarifies where data and related documentation should be placed in an eCTD, includes folders and descriptions for legacy data folders.

All other parts of this document as well as the parent binding guidance documents only express concepts in terms of the FDA endorsed data standards.

Of note, the FDA currently expresses the set of FDA endorsed data standards based on the format of the tabulation (SDTM) and analysis (ADaM) data that you will include in your submission. The FDA has acknowledged that a CDISC standard for data collection, CDASH (Clinical Data Acquisition Standardization & Harmonization) does exist

but that it is not regarded as mature enough to be referred to as the agency's expected standard for clinical data collection. In response to a question during the FDA panel discussion at the Society for Clinical Data Management (SCDM) annual meeting in September 2015, Ron Fitzmartin stated that FDA has been monitoring the evolution of the CDASH standard at that the plan to formally evaluate the next release of the CDASH standard, v2.0, due out in the 1st quarter of 2016 (as of April 7, 2016 it has not been released), for incorporation into the set of FDA endorsed data standards.

The Study Data Technical Conformance Guide [11] does call out the CDASH standard in two places, once in Section 4.1.1.2 as an example of collecting data based on the SDTM standard, and a 2nd time in 2nd paragraph of Section 8.3.1, again calling out that following the CDASH standard should lead to capturing data in SDTM complaint format. They are careful, though, to say "for example, CDASH", not naming CDASH as an endorsed standard.

A final thought on what is regarded as legacy format data, the PhUSE initiative to define the template, completion guidelines and examples of the Study Data Standardization Plan (Study Data TCG [11], Section 2.1) have included a definition of what they regard as legacy data in the template document [17], which reads, "Study data that does not conform to the standards by the date of requirement specified in the published Data Standards Catalog [12]." While this definition was not generated by the FDA, members of the FDA are part of this effort and both contributed to and validated this definition at the 2016 FDA/PhUSE Computational Sciences Symposium meeting held March 14th & 15th, 2016.

WHEN AND WHY MUST I SUBMIT LEGACY FORMAT DATA?

Section 8 of the Study Data Technical Conformance Guide [11] is devoted to the concepts of validation and traceability as these are the cornerstones for the agency's position on how data should be assessed for quality prior to delivery as part of a submission package. The numerous examples in the tables contained within Section 8.3.2.1 identify the potential pitfalls one can encounter when converting from legacy format data to an FDA endorsed data standard. All of these examples, taken collectively, state that if legacy format data is on the critical path from data collection to analysis and interpretation of results, then it must be provided to the agency as part of the regulatory filing. This would apply for all of the following scenarios:

1. Legacy format source data was used as the basis for producing legacy format analysis data which in turn was used to produce tables, figures & listings for the clinical study report. In this case, both the legacy format source and analysis data would need to be included in the submission.
2. Legacy format source data was used as the basis for producing SDTM domain datasets and in turn ADaM analysis datasets that were the basis for the tables, figures & listings appearing in the clinical study report. In this case all three sets of data would be submitted.
3. Similar to the first scenario, legacy format source and analysis data was used to support the clinical study report but SDTM domain datasets were also produced from the legacy format source data to support the development of an integrated safety database that would be the basis for tables, figures & listings that would appear in the Integrated Summary of Safety as well as be referred to in the Summary of Clinical Safety within Module 2 of the eCTD. In this case your legacy assets, the study level SDTM domains and your pooled safety database would all be provided to the agency as part of the application.

The following is specific thoughts on legacy data and particularly legacy data conversion taken from Section 8 of the Study Data Technical Conformance Guide [11].

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. This report, template and completion guidelines will be incorporated into the Study Data Reviewer's Guide and should be available sometime in 3Q2016.

- 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)

WHAT GUIDANCE DOCUMENTS DICTATE HOW LEGACY DATA SHOULD BE PREPARED FOR REGULATORY SUBMISSION?

The Study Data Technical Conformance Guide [11] is written with the future in mind; virtually all of its focus, with the exception of Section 8, is on preparing and submitting data and related documentation based on FDA endorsed data standards. Section 8 digresses solely for the purpose of supporting sponsors and service providers as they bridge the gap between data that may have been captured and/or analyzed based on legacy format standards but will also serve as the basis for creating assets based on FDA endorsed data standards.

So, this begs the question, what guidance should I follow when preparing legacy format data for inclusion in a regulatory submission to the FDA? The answer lies is direct feedback that was first publicly recorded during the FDA panel discussion at the FDA/PhUSE Computational Sciences Symposium 2015 meeting and reinforced via a eData Room response [18] to this direct question (**NOTE**: the full text of the questions posed and the FDA response is included in the References section of this paper). Specifically, the agency states:

> "Where the Study Data Technical Conformance Guide does not explicitly re-define, restate, clarify, or otherwise alter the information provided in previous guidance, sponsors should adhere to the previous guidance. This primarily relates to "legacy" data information, as the Study Data Technical Conformance Guide focuses on "standardized" data."

The agency goes on to make the following specific statements about legacy data, the data definition table for legacy data and the annotated CRF for legacy data [18]:

> "Legacy data formatting should follow Sections 2.5 & 2.7 of the Study Data Specifications document if the Study Data Technical Conformance Guide does not explicitly provide information and/or specification for legacy study data submission."
>
> "The data definition for legacy data can be defined in either a pdf file or in an xml file which adhere to the define.xml standard. The previous guidance provided should be used for data definitions related to legacy data, if standardized data definitions are not used."
>
> "The annotated case report form submitted should adhere to the Study Data Technical Conformance Guide (Section 4.1.4.6), unless the Study Data Technical Conformance Guide does not explicitly re-define, restate, clarify or otherwise alter the information provided in previous guidance. When information is not superseded, sponsors should adhere to the previous guidance."

WHAT MAKES UP A LEGACY FORMAT DATA PACKAGE?

Based on this FDA clarification, the following is a list of requirements that distill the FDA's thinking on the subject of delivering legacy data to the FDA.

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 as well as tables & figures in support of primary and secondary efficacy (Study Data TCG [11], Section 4.1.2.10)

CONCLUSION

The FDA has set the future standard for which clinical and non-clinical data must meet in order to be accepted by the agency for review of a NDA, ANDA and/or BLA. The articulation of their expectations is just that, focused on the future, but leaves out a number of details on how Sponsors and Service Providers should prepare data and related documentation that does not meet the future standard. At this point in time the agency's penultimate goal is for Sponsors and Service Providers to embrace FDA endorsed data standards, their ultimate goal is traceability from collection through to analysis. Paying attention to the preparation and presentation of assets based on both FDA endorsed and legacy format data standards, understanding and embracing full transparency when establishing clinical data chain of custody can only help a Sponsor's cause when in front of the agency. Hopefully this paper has filled in the gaps left by the FDA's recently issued guidance documents with respect to preparing and presenting legacy format data to the FDA.

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 April 7, 2016. (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 7, 2016.

[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/ucm149705.pdf, verified April 7, 2016.

[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 April 7, 2016. (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 7, 2016. (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 7, 2016. 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 7, 2016.

[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 7, 2016.

[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 April 7, 2016.

[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 April 7, 2016.

[11] US Department of Health and Human Services, U.S. Food and Drug Administration. "Technical Specifications Document: Study Data Technical Conformance Guide", Version 3.0, Issued March 31, 2016. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf, verified April 7, 2016.

[12] US Department of Health and Human Services, U.S. Food and Drug Administration. "Data Standards Catalog", Version 4.4, Issued August 17, 2015. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx, verified April 7, 2016.

[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 (registration with www.cdisc.org required), verified April 7, 2016.

[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 7, 2016.

[16] US Department of Health and Human Services, U.S. Food and Drug Administration. "Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications", Revision 3, issued May 15, 2015. Available at http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm333969.pdf , verified April 7, 2016.

[17] Pharmaceutical Users Software Exchange (PhUSE), Study Data Standardization Plan template, a draft of which can be found at http://www.phusewiki.org/wiki/images/e/ea/SDSP_Template.pdf, verified April 7, 2016.

[18] eData Room response to question about what guidance to follow when preparing legacy data for submission to the FDA:

Dear David,

Thank you for your questions regarding preparation of electronic submissions to the FDA.

Your question:

The high level question is, if a Sponsor needs to submit data based on a non-standardized format (examples will follow), what file organization and publishing standards should they follow now that the Study Data

Specifications, which used to delineate the requirements for organizing and publishing data in non-standardized format, is no longer in effect?

Specifics:

Legacy Data Submission Scenarios

- *Scenario #1 – A submission includes studies run well before submission standards were in place. Legacy format source data came out of a clinical database management system, legacy format analysis datasets were created, TLFs were generated, a CSR was produced. This study will now be included in a regulatory submission and SDTM domains will be created to facilitate integrating the data to support ISS/ISE activities. In this case I would see the sponsor having to submit both the legacy source data and the legacy analysis data.*
- *Scenario #2 – SDTM & ADaM standards were used to organize the CRF and analysis data, respectively, but the clinical database management system was not developed using CDASH standards, therefore a significant programming effort was required to transform the "extracted" data to SDTM. In the case I would see the Sponsor submitting the "extracted" data along with the SDTM & ADaM data.*

Guidance from Study Data Specifications Document for Legacy Data

The following sections of the Study Data Specifications guidance document, using Version 2.0 in this example, provide the specific guidance for how to organize and publish legacy format data:

- *Section 2.5 – General Considerations for All Datasets*
- *Section 2.7 – General Considerations for Analysis Datasets*
- *Section 3.1.2.2 – Data Definition / Metadata (non-standard/legacy)*
- *Section 5.4 – Annotated Case Report Form*

Key Points

- *The Study Data Technical Conformance Guide only talks about submitting via the standards in the Data Standards Catalog. Legacy data formatting requirements, which used to appear in Sections 2.5 & 2.7 of the Study Data Specifications document, are no longer a part of guidance. QUESTION: What are the data formatting requirements for the legacy data?*
- *The Study Data Technical Conformance Guide only talks about providing a define.xml in version 2.0 for SEND, SDTM & ADaM and provide a companion define.pdf derived from your define.xml if your define.xml is in version 1.0. It says nothing about the requirements for the define.pdf that accompanies legacy format data that used to appear in Section 3.1.2.2 of the Study Data Specifications. QUESTION: What are the define.pdf requirements for a define.pdf that would accompany legacy data?*
- *The Study Data Technical Conformance Guide, in Section 4.1.4.6, states requirements that should be followed regardless if the clinical database is in legacy format or if it is SDTM compliant, but further states that you should use the SDTM Metadata Submission Guidelines for additional information. QUESTION: Can you confirm that Section 4.1.4.6 completely and adequately supersedes SDS v2.0 Section 5.4 regardless of the type of data being submitted (legacy vs. SDTM)? Should a submission of legacy data name the blank CRF annotated to the legacy data "acrf.pdf" or keep the name "blankcrf.pdf"?*

The Agency prefers sponsors follow the *Study Data Technical Conformance Guide (2.1)* to submit all scientifically relevant data collected in a study. Please note the latest version.

The *Study Data Technical Conformance Guide (2.1)* defines, clarifies, and supersedes certain content of the previous guidance containing information/specifications on standardized data. (Study Data Specifications documents (Versions 1.0 - 2.0) and CDER Study Data Common Issues Documents (Versions 1.0 -1.1).

Where the Study Data Technical Conformance Guide does not explicitly re-define, restate, clarify, or otherwise alter the information provided in previous guidance, sponsors should adhere to the previous guidance. This primarily relates to “legacy” data information, as the Study Data Technical Conformance Guide focuses on “standardized” data.

The Study Data Technical Conformance Guide does include a section on Data Validation and Traceability, which describes the Agency’s current thinking on validation of standardized study data and the traceability of study data within and across the components of the submission (8. Data Validation and Traceability). This section also includes a specific sub-section on the conversion of legacy study data to standardized study data (8.3.2 Legacy Study Data Conversion to Standardized Study Data). The Agency prefers standardized study data, but also recognizes the need to transition to standardized data submissions and to support various methodologies in the transition; to include support for the concurrent submission of legacy and standardized study data within the same submission.

Again, while the Study Data Technical Conformance Guide is specifically written to implement the electronic submission requirements of section 745A(a) of the FD&C Act with respect to standardized study data contained in certain investigational new drug applications (INDs), new drug applications (NDAs); abbreviated new drug applications (ANDAs); and certain biologics license applications (BLAs) that are submitted to the Center for Drug Evaluation and Research (CDER) or the Center for Biologics Evaluation and Research (CBER), the Agency does expect sponsors to adhere to previous applicable guidance where the Study Data Technical Conformance Guide does not explicitly supersede, and/or provide specific information and/or specification related to study data submission. For example, the Study Data Technical Conformance Guide does not provide information and/or specification for legacy study data submission, except to the extent described in the paragraph above.

The following provides additional clarification on specific topics that existed in the previous Study Data Specifications guidance:

- Section 2.5 – General Considerations for All Datasets
- Section 2.7 – General Considerations for Analysis Datasets
- Section 3.1.2.2 – Data Definition / Metadata (non-standard/legacy)
- Section 5.4 – Annotated Case Report Form

Legacy data formatting should follow Sections 2.5 & 2.7 of the Study Data Specifications document if the Study Data Technical Conformance Guide does not explicitly provide information and/or specification for legacy study data submission.

The data definition for legacy data can be defined in either a pdf file or in an xml file which adhere to the define.xml standard. The previous guidance provided should be used for data definitions related to legacy data, if standardized data definitions are not used.

The annotated case report form submitted should adhere to the Study Data Technical Conformance Guide (Section 4.1.4.6), unless the Study Data Technical Conformance Guide does not explicitly re-define, restate, clarify or otherwise alter the information provided in previous guidance. When information is not superseded, sponsors should adhere to the previous guidance.

For your specific questions:

QUESTION: What are the data formatting requirements for the legacy data?

Please refer to the information provided above concerning guidance for legacy data and consider the guidance for communication with the review division (provided below).

QUESTION: What are the define.pdf requirements for a define.pdf that would accompany legacy data?

The Agency does not provide specific guidance for the define.pdf content, only the preferred location and Portable Document Format per eCTD guidance.

QUESTION: Can you confirm that Section 4.1.4.6 completely and adequately supersedes SDS v2.0 Section 5.4 regardless of the type of data being submitted (legacy vs. SDTM)? Should a submission of legacy data name the blank CRF annotated to the legacy data "acrf.pdf" or keep the name "blankcrf.pdf"?

The Agency does prefer sponsors transition from "blankcrf.pdf" to the "acrf.pdf" naming for the annotated case report form. Per the *Study Tagging File Specification and Related Files*, the sponsors would continue to indicate the ectd study-values file-tag valid-value as "annotated-crf".

Please consider the following additional general guidance in preparing your submission. These topics are also covered in the latest guidance (*Study Data Technical Conformance Guide (2.1) and related guidance - Study Data Standards Resources).*

Research study designs should define the protocol for data collection. The Agency's methodology and submission structure supports research study design, as indicated in the *Guidance to Industry, Providing Regulatory Submissions in Electronic Format - Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications* and the *Study Data Technical Conformance Guide (2.1).* The Agency's methodology and submission structure also supports integrating study data collection for Safety and Efficacy study submission. The Agency prefers implementation of analyses datasets to tabulations datasets traceability. In addition, the Agency prefers each study submitted to be complete and evaluated on its own merits.

The Agency expects sponsors to evaluate the risk involved converting study data collected to standardized data, if applicable. The Agency prefers sponsors to submit study data conversion explanation and rationale. The study data conversion rationale and explanation should address either scenario; decision rationale for not converting or decision rationale for converting. The Agency expects the sponsor's evaluation and rationale includes study data scientifically relevant to the application's safety and efficacy representation. As such, the evaluation and explanation may include rationale based on the pooling/integrating of data from multiple studies.

Finally, please note the review team assigned to the application determines the acceptability. If there are other expectations already communicated with the review division, please follow these best practices noted in the *Study Data Technical Conformance Guide (2.1)*, "The Guide is intended to complement and promote interactions between sponsors and FDA review divisions. However, it is not intended to replace the need for sponsors to communicate directly with review divisions regarding implementation approaches or issues relating to data standards." (p. 1) and "If there is a question regarding a specific submission or a particular data standard implementation, the sponsor should contact the review division for specific submission questions or the appropriate contact for data standards issues (cder-edata@fda.hhs.gov or cber.cdisc@fda.hhs.gov)." (p.2).

For additional questions, please feel free to send an email to edata@fda.hhs.gov.

Additional Links:

Electronic Regulatory Submissions and Review Helpful Links

Study Data Standards Resources

Electronic Common Technical Document (eCTD)

Case Report Tabulation Data Definition Specification (CRT-DSS)

ACKNOWLEDGMENTS

The author would like to thank the many clients and colleagues from organizations large and small that have given me the opportunity to work through the challenges of preparing and submitting legacy data to the FDA as well as the formal and information guidance from the FDA that has made this paper possible.

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 participation in the 2015 FDA/PhUSE Computational Sciences Symposium meeting, held March 16th & 17th, 2015, are available here: http://www.phuse.eu/CSS-Presentations2015.aspx . Additional presentations from the 2016 FDA/PhUSE Computational Sciences Symposium meeting, held March 14th & 15, 2016, are available here: http://www.phuse.eu/CSS-Presentations2016.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) 467-8790
E-mail:	david.c.izard@accenture.com