

First Time Creating a Submission Package?

Don't Worry, We Got You Covered!

Lyma Faroz, Seagen Inc., Bothell WA

ABSTRACT

Creating a clinical trial data package for electronic submission to a regulatory agency is a daunting task. There are many steps that must be executed with precision and efficiency to create a good quality submission package. If you are working on a submission study for the first time, then this paper is for you.

Each electronic submission contains 5 modules; however, this paper will focus on steps involved in creating data set components for Module 5. This includes validating SDTM and ADaM data sets using Pinnacle 21 Enterprise software, generating reviewer's guides (cSDRG and ADRG), creating the define.xml for SDTM and ADaM, and much more! We will also look at important points from FDA's Study Data Technical Conformance Guide and sprinkle in various lessons learned from earlier submissions in pulling all this together to create a high-quality submission package.

INTRODUCTION

There are many components that contribute towards an entire electronic submission package for FDA that is prepared hands-on by statistical programming teams and included in the electronic Common Technical Document (eCTD) Module 5. This module consists of everything related to human clinical trial data such as data sets, reviewer's guides, define.xml, additional definition documents (if needed), and data sets/TLF/macro programs (as agreed with FDA in a Type C meeting or correspondence). In this paper we will discuss and walk through various good practices about creating these items.

Note: This paper assumes that a submission package is created in adherence with CDISC clinical data structures and standards, thus those aspects will not be discussed in this paper.

MODULE 5 FOLDER STRUCTURE

There are several documents that FDA has provided at <https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources> which provide FDA guidance to help create submission-ready data packages. The Module 5 folder structure is taken from one such standard guidelines document, the [Study Data Technical Conformance Guide \(TCG\)](#).

The folder structure for study data sets in Module 5 looks as shown in Figure 1. An example folder structure for an actual study is also shared for reference in Figure 2.

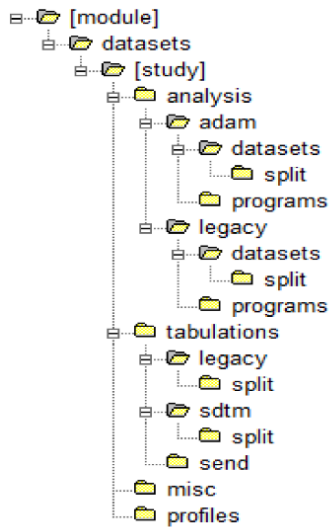


Figure 1. M5 Folder Structure from the FDA TCG

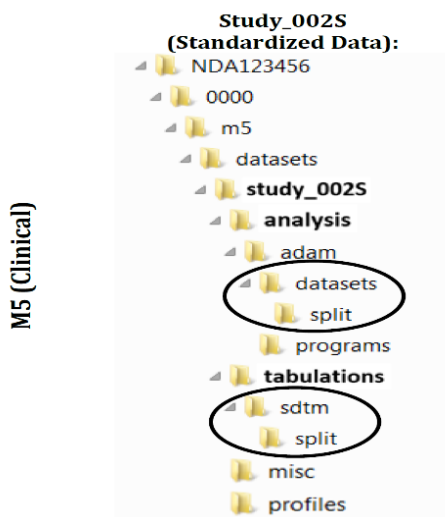


Figure 2. Example M5 Folder Structure for a Study

TABULATIONS FOLDER

This folder stores all components related to the SDTM eSub data package, such as SDTM data sets (as SAS® V5 transport files), SDTM annotated CRF (aCRF), SDTM reviewer's guide (cSDRG), define.xml (and define style sheet), printable define PDF (optional), and any additional definitions document or study-specific supplemental files/documents. In our experience, companies do not submit SAS programs for SDTM data sets, because these data sets are sufficiently standardized to be understood without programming code to reference; such programs also operate on raw data which are typically not provided to FDA, thus further reducing the degree to which the code may be meaningful to reviewers.

ANALYSIS FOLDER

This contains all information regarding ADaMs such as the ADaM reviewer's guide (ADRG), define.xml and style sheet, printable define PDF (optional), ADaM data sets (as SAS® V5 transport files), and

ADaM/TLF/macro SAS or R programs in ASCII format (as agreed with FDA). Note that a common misperception is that when the TCG mentions "ASCII format," it means renaming SAS or R programs to a *.txt extension. SAS and R programs are actually already in ASCII format by default regardless of their file extension, and they can be submitted directly as *.sas or *.r without being renamed to *.txt.

ANNOTATED CRF

Annotated CRFs are blank casebooks mapped to SDTM data set and variable information. While some may consider an annotated CRF as something to be developed right before a submission, several companies, including ours, actually see an upside to making this the very first step towards developing SDTM data sets. Doing so helps expedite and bring efficiencies to the process of building SDTM data sets. During the lifecycle of study, a working version of this blank CRF can be maintained that has annotations for both raw database extract variables as well as SDTM variables, as it helps with internal traceability and fast development of SDTM data sets. Note that for submission purposes the requirements are more strict: a blank CRF without raw data variable annotations should be used to annotate SDTM data sets and variables. This is the aCRF file that will be submitted to the regulatory agency.

SOME IMPORTANT POINTS TO CONSIDER:

1. Annotate only the unique CRF pages used in the study. Repeat pages that follow later on are annotated only with a reference back to the original page, along with annotations for fields that are different than the first page that was fully annotated.
2. Multiple domains can share a single CRF form, so domains should be annotated in different domain-specific colors. Beware of using red vs green as a differentiator, since these will appear the same in the most common form of human color-blindness. Also avoid using bright font colors on bright backgrounds, like bright red font on bright blue background.
3. Information that is not submitted in SDTM data sets should be marked as [NOT SUBMITTED].
4. The annotated CRF should be named 'acrf.pdf' in order for the hyperlink in define.xml to work and to be submitted to the FDA.
5. Once annotations are finalized, the PDF page number should be accurately referenced in SDTM define.xml for relevant datasets and variables, and the file should be flattened prior to submission, so that annotations are not editable. To not lose bookmarks after flattening the file, make sure to use the "Optimize PDF" tool in Adobe® Acrobat PRO, select the preflight option, and then flatten the file.

DM=Demographics

SC=Subject Characteristics

Enrollment/Screening [ENROLL]

SGNXX-XXX Screening ID SCREENID (num, 8) [NOT SUBMITTED]	<system-generated, see note below> {SCREENID} ^a
Do not enter unless patient has signed informed consent.	
Patient Initials SUBJINIT (char, 3)	SCORRES when SCTESTCD = SUBJINIT
<STUDY SPECIFIC/OPTIONAL> Patient initials not provided due to Protected Health Information (PHI) regulations. INITNA (num, 8) [NOT SUBMITTED]	Checkbox: [1] <input checked="" type="checkbox"/>

Display 1. Annotated CRF with raw data and SDTM variable annotations

DM=Demographics

SC=Subject Characteristics

5.001 CW 17SEP2019: _Unique Forms

Project Name: [REDACTED]

Form: Enrollment/Screening

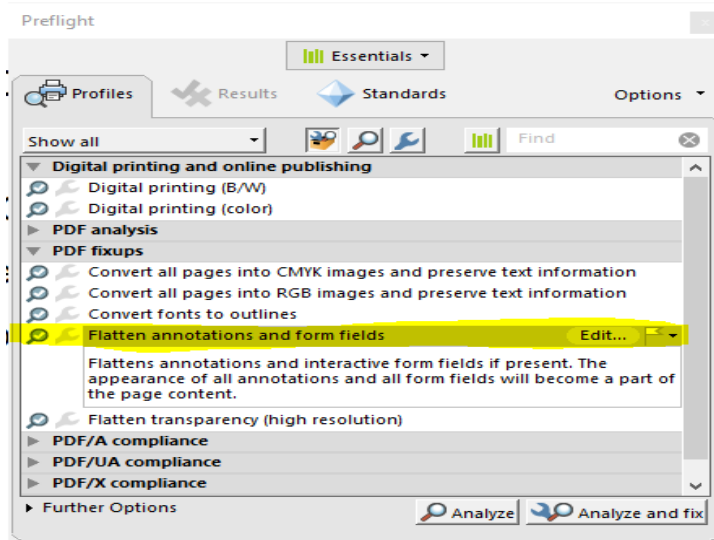
Generated On [REDACTED]

[REDACTED] Screening ID	[NOT SUBMITTED]
Do not enter unless patient has signed informed consent.	
Patient Initials	SCORRES when SCTESTCD=SUBJINIT

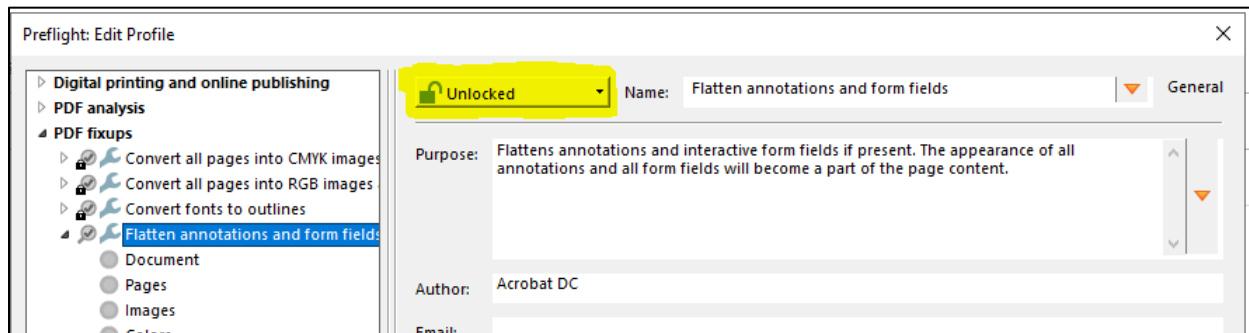
Display 2. Annotated CRF with SDTM variable annotations only, prepared for submission

STEPS TO FLATTEN ANNOTATED CRF: USING ADOBE® ACROBAT PRO VERSION 2017

Options to flatten the file: Tools > Optimize PDF > Check for option "Flatten annotations and form fields" > Edit to its right > Click on Edit.



Display 3



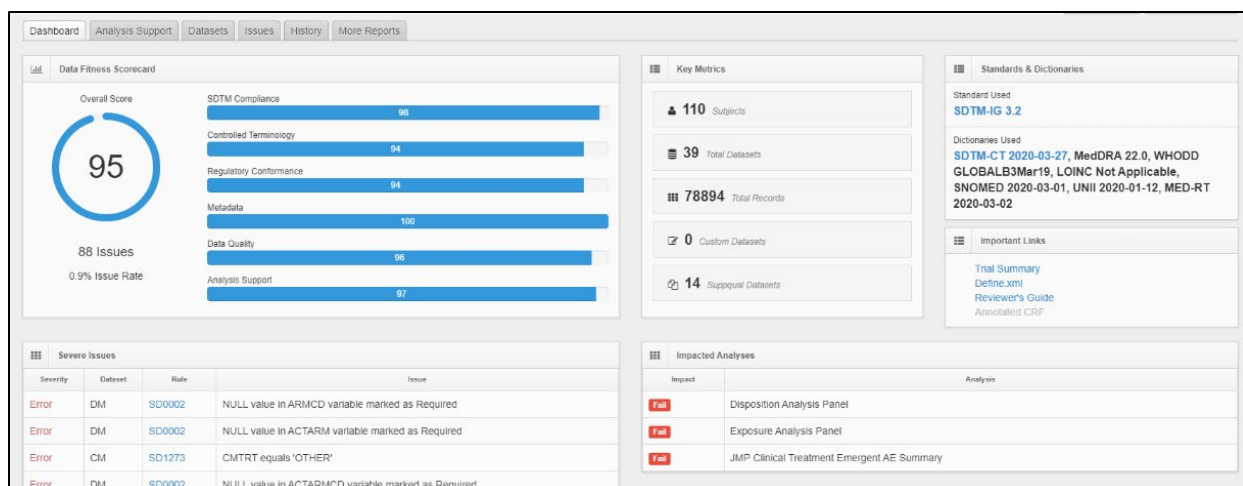
Display 4. Check if flattening options are unlocked > Save > Analyse and Fix > Save as PDF > Rename to acrf.pdf and save.

PINNACLE 21 ENTERPRISE

This section is applicable only for subscribers to Pinnacle 21 Enterprise software, installed on the user's computer. It is expected that respective project, study, dictionaries and standards information and versions are accurately entered to the Pinnacle 21 Enterprise software prior to using it for package validation and generating define.xml and reviewer's guides.

Once the study CRFs are annotated with SDTM mapping, the next step is to develop SDTM and ADaM data sets based on data specifications, the SDTM and ADaM IG, study SAP, TLF mock shells, etc. Once a considerable number of datasets are programmed, it is recommended to validate them using Pinnacle 21 software. Currently, there are two version of Pinnacle 21 available to the users. The Community version is freely available to download from the website and Pinnacle 21 Enterprise is the paid version, that is built over the community version and performs additional checks, provides data fitness scorecards, and facilitates generation of define.xml and reviewer's guides.

This section will cover some tips and tricks in using Pinnacle 21 Enterprise to leverage additional features of the enhanced software.



Display 5. Pinnacle 21 Enterprise Dashboard

Use Pinnacle 21 Enterprise instructions to upload and validate the data. Once the data is uploaded to Pinnacle 21 Enterprise, based on the quality and completeness of the data, the dashboard will look something like the image shown in Display 5.

While there is no specific target score that makes or breaks your submission, in general it is fair to say that the higher the data fitness score, likely the better the quality and structural integrity of the data. Users can improve the data fitness score by reviewing the list of issues in the “Issues” tab and updating the data package accordingly. Once the data is updated, a reupload is needed to see the change in score.

Pinnacle 21 Enterprise also facilitates issue explanation and tracking, to help build the “Issues Summary” section in the reviewer’s guides.

GENERATING DEFINE FILES USING PINNACLE 21 ENTERPRISE

Generating define files in XML and PDF format is not a straightforward process when attempted from scratch. Pinnacle 21 provides an efficient way to generate these files. It uses a source template that should be updated with a data set specification for a study, often developed by a programming team in Excel which is then uploaded into Pinnacle 21 Enterprise in a specific pre-specified load file structure.

To make the loading process easier, it will be beneficial when developing the Excel data set specifications during analysis development to think ahead with define.xml generation in mind. For example, your Excel spec...

- Had best be structured similarly from one study and even product to the next.
- Should use the more granular XML types for numeric variables rather than only the SAS "numeric" type which is not acceptable for define.xml.
- Prepare Excel file that contains resized variable length.

Once you have used your Excel spec to develop several data sets on your study, you can create the Pinnacle 21 load file based upon it. This will usually involve a degree of transformation of your internal spec to the load file structure. At Seagen we capitalized on our department-wide standard specification structure by developing a central macro that creates the load files automatically based on any of our internal study specs; one of the many benefits of a centralized and structurally consistent representation of data specs. Once your load file is ready, you can bring it into Pinnacle 21 as follows:

1. Click on the 'Define.xml' link at the right-hand corner of the dashboard.
2. A new window will open which will give tab options such as Import, Export, Preview, etc.
3. Use the **'Import'** tab and upload the data set specification using the **'Excel Spec'** option.

- Once the specification is uploaded, use the 'Export' tab, and download the 'Define.xml', 'Define.pdf' and 'Stylesheet', all in the same folder.



Display 6. Pinnacle 21 Enterprise: Link to generate define.xml on the dashboard



Display 7. Pinnacle 21 Enterprise: Upload Excel Specs



Display 8. Pinnacle 21 Enterprise: Download Define.xml, Define.pdf, Stylesheet

Supplemental Documents

- Datasets
- Controlled Terminology
- Methods

Expand all VLM

Collapse all VLM

Standard: ADaM-IG 1.1

Study Name: [REDACTED]

Study Description: [REDACTED]

Protocol Name: [REDACTED]

Metadata Name: Study [REDACTED] Data Definitions

Date/Time of Define-XML document generation: 2020-07-04T02:58:53Z
 Define-XML version: 2.0.0
 Stylesheet version: 2018-11-21

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
ADPRIOR	Subject Level Therapies Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	STUDYID, USUBJID		adprior.xpt
ADSL	Subject Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	STUDYID, USUBJID		adsl.xpt
ADAE	Adverse Events Analysis Dataset	BASIC DATA STRUCTURE	One record per adverse event per subject	Analysis	STUDYID, USUBJID, AESEQ		adae.xpt

Display 9. Example define.xml created using Pinnacle 21 Enterprise. Sections are accurately hyperlinked. SDTM versions also have correct link between aCRF pages and corresponding define.xml data specifications

SOME IMPORTANT POINTS TO CONSIDER:

1. The define.xml file and corresponding style sheet should be in the same location for the define.xml to open in a browser in the expected display format.
2. Check for any truncations in the 'Comment' or 'Method' column of define.xml, especially for variables that have a long derivation logic.
3. Check if all the hyperlinks work and open as desired.
4. A define.pdf can also be generated using Pinnacle 21 Enterprise, if desired. This is optional if define.xml v2.0 or higher is submitted, yet we recommend producing it regardless, so reviewers have a more easily printable version of these critical metadata.
5. CRF page numbers can be automatically generated by Pinnacle 21 if both aCRF and the excel load file are provided. Hence, users do not need to manually maintain the CRF pointers or build tools for it if they have access to Pinnacle 21 Enterprise.
6. Additional documents or datasets can be hyperlinked in define.xml as well. For links to work, they should be mentioned in the 'Documents' tab of the Pinnacle 21 spec template and stored in the same folder as the specs. If any documents are a part of the package and mentioned in the 'Documents' tab, it should be carried over to 'Documents' column in 'Methods' or 'Comments' tab of the Pinnacle 21 spec template.

A	B	C
ID	Title	Href
adrg	Reviewers Guide	adrg.pdf
lu-meddra-smq-v22-0	lu-meddra-smq-v22-0	lu-meddra-smq-v22-0.pdf
lu-ctcae-v5-0	lu-ctcae-v5-0	lu-ctcae-v5-0.pdf
CODE.001	adsl.sas	../programs/adsl.sas
CODE.002	adae.sas	../programs/adae.sas
CODE.003	adcm.sas	../programs/adcm.sas

Display 10. Create hyperlinks in define.xml as mentioned in point #5

REVIEWER'S GUIDE

Pinnacle 21 can be used to generate reviewer's guides as well for both SDTM and ADaM, using the link at the right-hand corner of the dashboard. An efficient feature of Pinnacle 21 Enterprise is that all unresolved issues can dynamically be extracted into the 'Issue Summary' section of the reviewer's guide.



Display 11. Pinnacle 21 Enterprise: Link to generate Reviewer's Guide on the dashboard

If you do not have access to Pinnacle 21 Enterprise, no worries! PHUSE provides stepwise instructions along with a template which can help to generate reviewer's guides as well.

<https://advance.phuse.global/display/WEL/Deliverables>

SOME IMPORTANT POINTS TO CONSIDER:

1. The reviewer's guide is a good place to document important information regarding the study as well as data set-related structural or logical decisions taken throughout the course of the study, to be shared with the reviewers of a submission.
2. Check that all hyperlinks work accurately throughout the document in its final location – not just in the location where you first wrote the guide.
3. The 'Issue Summary' section should be completed with proper explanation for any unresolved errors, warnings, or notices, such that reviewers can easily understand why a specific item was not resolved before submission.
4. The reviewer's guide for SDTM should be named 'csdrg.pdf' and for ADaM 'adrg.pdf,' to follow the correct eCTD naming conventions as expected by FDA.

TIPS FOR OVERALL SUBMISSION PACKAGE

1. Per the FDA document '[Technical Rejection Criteria for Study data](#)', a TS data set should be included in the tabulations folder whether or not the study data package is in CDISC format, for the Agency's automated eCTD validation process, which checks the study start date. Absence of this data set will trigger an automatic technical rejection.
2. If PDF documents are submitted, then they should be checked for the following eCTD properties.
 - a. PDF version should be compatible **with v1.4 to v1.7**.
 - b. Check File > Properties > Description to verify the PDF is optimized for fast web view, i.e., **Fast Web View: Yes**.
 - c. Check File > Properties > Security, verify no security setting is present, i.e., **Security Method: No Security**.
 - d. Check File > Properties > Initial View, verify the Navigation tab shows "**Bookmarks Panel and Page**."
3. Use hyphens instead of underscores in filename for all files included in the submission package. The Agency understands that if companies submit SAS programs that were finalized with underscores in their name, or SAS macros that were automatically included in the programs via the SAS global options **mautosource** and **sasautos** rather than with explicit **%include** statements, such programs may not exactly align to the file names they reference in headers or macro names in all cases. This is acceptable, especially if briefly explained in the ADRG.
4. Programs submitted as part of the package should be cleaned for any "dead code" or unwanted comments. Programs should be clean, and comments should be provided wherever necessary.

CONCLUSION

As daunting a task as it may seem, creating an acceptable electronic submission package for FDA benefits greatly from early and detailed planning. It is very important to gain familiarity and knowledge of guidelines provided by FDA. With the help of efficient planning, knowledge, and various tools and software provided to Industry like Pinnacle 21, sponsors can create a high-quality package for regulatory agencies.

REFERENCES

Study Data Technical Conformance Guide: <https://www.fda.gov/media/143550/download>

Technical Rejection Criteria for Study Data: <https://www.fda.gov/media/100743/download>

Providing Regulatory Submissions in Electronic Format - Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications:
<https://www.fda.gov/media/135373/download>

PHUSE Reviewer's Guide Templates (ADRG and (cSDRG):
<https://advance.phuse.global/display/WEL/Deliverables>

Ready, Set, Go: Planning and Preparing a CDISC Submission: <https://www.lexjansen.com/phuse-us/2018/sp/SP03.pdf>

ACKNOWLEDGMENTS

I would like to thank Shefalica Chand, John Shaik and Michiel Hagendoorn for their valuable feedback and constant support and guidance.

CONTACT INFORMATION

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

Lyma Faroz
Seagen Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
lfaroz@seagen.com