Sorting Out the Paperwork – Define.xml versus Reviewer's Guide and other Submission Documents

Sandra VanPelt Nguyen, Pfizer

Ellen Asam, Merck Wei Dong, Cytokinetics Annette Travalent, Eli Lilly

Abstract: Although most regulatory submissions are now electronic and we're not literally dealing with paperwork, there are still numerous required or recommended documents which support and describe the study datasets. To facilitate regulatory review, these documents need to be provided in a submission with sufficient information to support assessment of the data and ultimately the study results, however it can be time-consuming and frustrating to sort through the requirements, understand the intent of each document, and appropriately populate each document. This poster will help to "sort out the paperwork" so that anyone submitting datasets in support of a regulatory submission can more easily determine which documents are needed, will outline distinctions between the documents, and will identify standard templates, guidelines, or examples when available. Although this poster focuses on submissions to the U.S. Food and Drug Administration, similar documents are required or recommended for submissions to other regulatory agencies.

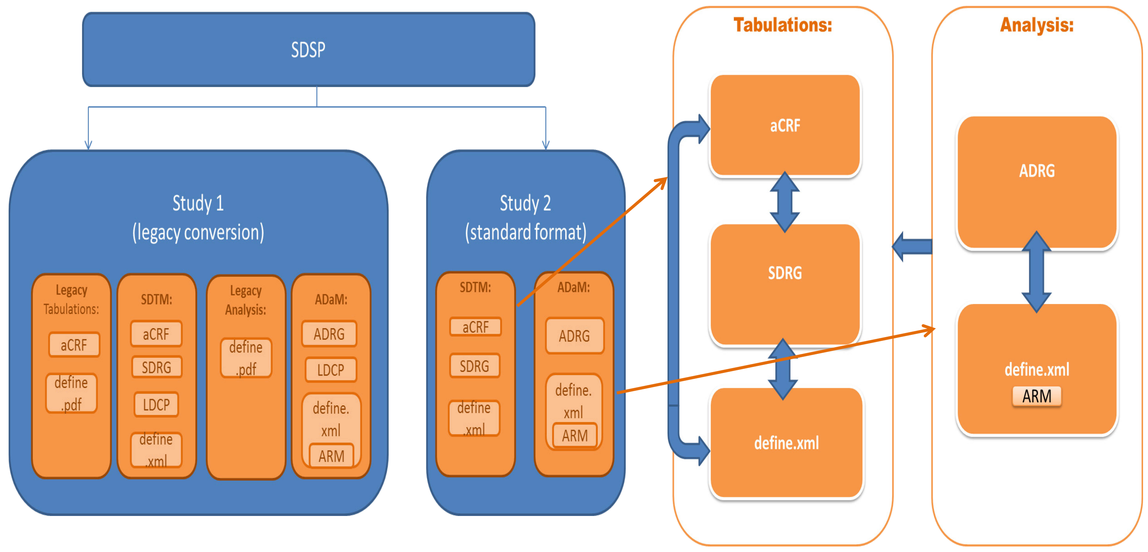
Several documents are referenced in the FDA's Technical Conformance Guide which are required or recommended to support and describe the clinical trial data included in a regulatory submission. Each document has a distinct purpose, however there is some intentional overlap between the documents as well. We recommend reviewing the FDA's Study Data Standards Resources page[1] to become familiar with the applicable guides, rules, and accepted standards. Additionally, industry organizations such as CDISC [2]and PhUSE[3] are actively working to create and maintain data and documentation standards, templates, examples, and completion guidelines in support of initiatives to streamline regulatory reviews. There is a lot of information to gather and absorb, which can be overwhelming. Trying to sort out and create the necessary documentation can be even more challenging and demanding. The table below is provided as a high-level introduction and comparison of the various submission documents which support the study datasets:

Document	What is it?	Required?	Important Considerations	Key Distinctions	Related Documents
Study Data Standardization Plan (SDSP)	Summarizes the completed, ongoing, and planned clinical and non-clinical studies, noting type/phase, study design, and the planned data standards, formats, and terminologies to be used for each, along with their corresponding versions.	Recommended	This is an important tool for discussions with the FDA which should be initiated and utilized early in development. Additionally, this document identifies any studies in legacy data structures as well as any which were originally tabulated and/or analyzed in legacy structures but were subsequently converted to a standard format.	This is a program-level document. It is also a living document which is expected to be updated from early development through to submission.	Data standards and version information will also be included in study-level documents (SDRG, ADRG, and define.xml). Studies which have had legacy data conversion done should have a LCDP created.
Legacy Data Conversion Plan and Report (LDCP)	The plan describes the legacy data and the process for converting to standard format. The report presents the results of the conversion, and notes any issues identified and their resolution or explanation.	Conditional: Should be provided when data has been converted from a legacy format to a standard format.	It is an important tool for documenting traceability, identifying data points or values which did not fit appropriately into the standards, and explaining differences between the original data/results and standardized data/results.	This document is only to describe the conversion process and outcome. Details on the converted data structures and contents are described elsewhere.	The SDSP should indicate where legacy data conversions have been performed.
Annotated Case Report Form (aCRF)	This is a blank case report form with annotations alongside each data collection field identifying the location of the data (dataset and variable name) in the SDTM or legacy tabulation datasets.	Required	The annotated CRF provides a regulatory reviewer with an understanding of the relationship between the data collected on the CRF and the submitted tabulation datasets. Each field should be annotated with the corresponding dataset and variable in the submitted datasets or should be annotated as "NOT SUBMITTED" if not included in the submitted datasets.	The annotated CRF provides a visual representation for how the data collected in the CRF is represented in the SDTM or legacy datasets.	The SDRG can be used to provide additional information regarding the annotated CRFs such as explanations of content organization, general annotation conventions, and explanations for variables which were collected but not submitted. The Origin column in define.xml should provide a hyperlink to the page in the annotated CRF where the variable in a submission dataset is located.
Data Definition File (define.xml)	Describes the structures and contents (datasets, variables, and controlled terminologies) of the submitted datasets. There is one file for each set of datasets (legacy, tabulation, analysis).	Required	The file should be as detailed as possible, particularly for derived variables, and should be written such that someone unfamiliar with the study can locate data values of interest and understand the source of the value or how it was computed or assigned. Define.xml v2.0 is the preferred version. A corresponding style sheet should be provided. A PDF version should be provided with the XML if the XML is not printable.	This is a machine-readable file which can be used by tools to help translate or verify the data. The structure is strict and must be followed to be compliant and compatible with tools. It is the key reference for documenting the source of the values within each variable.	Some information is intentionally duplicated between define.xml and the reviewer's guide. Additionally, the define.xml may link to the reviewer's guide to provide additional explanation or information that does not fit into the define.xml structure. The tabulations define.xml also contains hyperlinks to the annotated CRFs for data which was collected in the CRFs.
Analysis Results Metadata (ARM)	This is an extension to the define.xml file which can be used to link the analysis results to documentation and datasets and provide additional details on the study analyses.	Recommended	It provides traceability for the study results by linking the results to the source analysis datasets (and metadata) along with documentation of the analyses via text or programming code. The sponsor determines which key analyses to include.	This is a machine readable file and the only document which specifically links the Clinical Study Report (CSR), SAP, analysis datasets, and programming logic.	ARM is provided as part of the analysis define.xml and may contain links to the ADRG as well as to SAP, programs, and/or CSR.
Reviewer's Guides: **Study Data Reviewer's Guide (SDRG),** **Analysis Data Reviewer's Guide (ADRG)**	These guides provide additional context beyond the data definitions file, identifying the standards and versions implemented, describing the trial design, and noting any special considerations for the datasets and how they relate to or impact the study results.	Recommended	The reviewer's guide complements the define.xml and can be used to provide additional information about the datasets as well as convey information that does not have a place in define.xml or is not easily expressed in that format. The SDRG is focused on the tabulation datasets and data collection/mapping decisions or points of interest while the ADRG is focused around the analysis datasets, providing an overview of the analysis of the study and describing how the Statistical Analysis Plan (SAP) was implemented within the datasets.	It provides additional information to regulatory reviewers in a human readable format. The document is flexible in terms of its structure and the manner in which information is presented, allowing for visuals such as diagrams, tables, and figures and lengthier text which can help to illustrate or describe more complex data flows or computational algorithms. It also provides a summary of SDTM or ADaM conformance findings.	The SDTM/SEND define.xml (or define.pdf for legacy data) may provide a hyperlink or reference to additional details or information in the SDRG. The SDRG also describes annotation conventions or notes for the annotated CRFs. The ADaM define.xml (or legacy define.pdf) may provide a hyperlink or reference to additional details or information in the ADRG. Additionally, there is some intentional duplication of information found in the protocol and Statistical Analysis Plan (SAP).

Each study has its own set of documents, with exception of the SDSP, which is a program-level document (Figure 1). Within a study, the documents are highly related and each document will often hyperlink to or reference some of the other documents (Figure 2). Additionally, full traceability will allow a reviewer to go from the study results to the analysis data to the tabulation data to the data collection sources, thus the analysis documents should relate back to the tabulations documents.

Figure 1: Overview of Study Document Organization

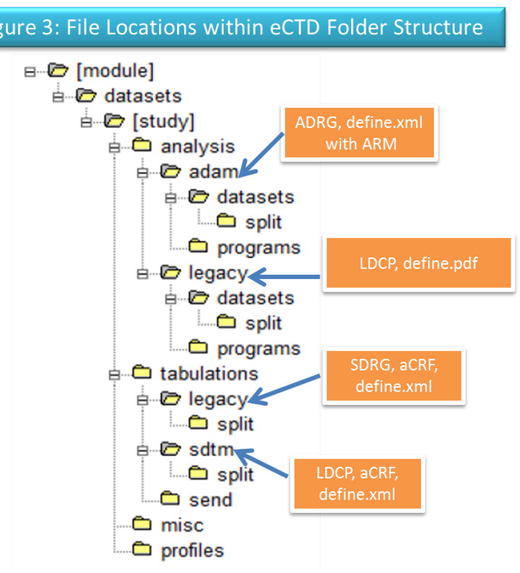
Figure 2: Relationships Between Documents



Although these documents are designed to describe data from individual studies, they can generally be adapted to describe integrated/pooled datasets as well.

Figure 3 shows where the documents should be stored within the eCTD folder structure. Note: the SDSP should be located within the general investigational plan for Investigational New Drug Applications (INDs). Otherwise, it is a stand-alone document. It can be included as an attachment in regulatory documents (e.g. Type C Briefing Document), but should be maintained separately.

Figure 3: File Locations within eCTD Folder Structure



Standards, Templates, Examples, and Completion Guidelines:*

SDSP: http://www.phusewiki.org/wiki/index.php?title=Study_Data_Standardization_Plan_(SDSP)
LDCP (note: template under development):
http://www.phusewiki.org/wiki/index.php?title=Legacy_Data_Conversion_Plan_%26_Report
aCRF: https://www.cdisc.org/standards/foundational/study-data-tabulation-model-implementation-guide-sdtmig/metadata-submission
define.xml: https://www.cdisc.org/standards/foundational/define-xml
ARM: https://www.cdisc.org/standards/foundational/adam
SDRG: http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide
ADRG: http://www.phusewiki.org/wiki/index.php?title=Analysis_Data_Reviewer%27s_Guide

Additional Resources and References:*

[1] FDA/CDER Study Data Standards Resources:
http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/
[2] CDISC: https://www.cdisc.org/
[3] PhUSE Working Groups: http://www.phusewiki.org/wiki/index.php?title=FDA_Working_Groups

*These sites and documents are updated regularly, so be sure to reference the latest versions when preparing for a submission.