# *Study Data Specifications*

Revision History

| Date | Version | Summary of Changes |
|---|---|---|
| | | |
| | | |

# STUDY DATA SPECIFICATIONS

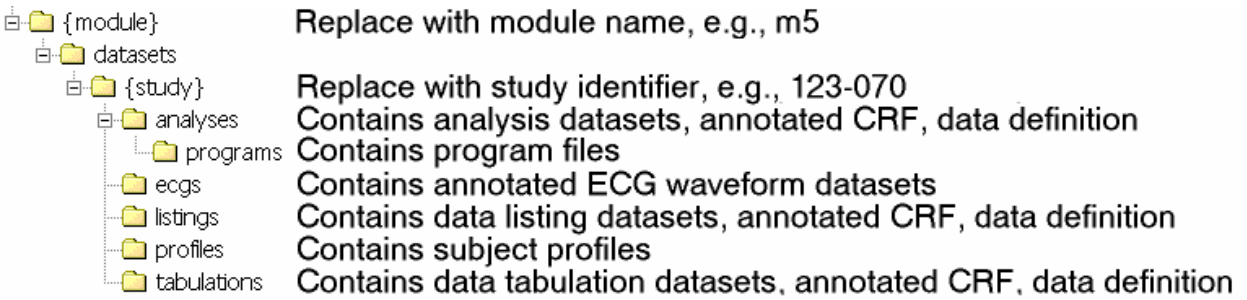
These specifications are for submitting animal and human study data in electronic format. Study data includes information from trials submitted to the agency for evaluation and information to understand the data (data definition). This includes both raw and derived data.

FILE FORMATS

See *Guidance to Industry: Providing Regulatory Submissions in Electronic Format – General Considerations* for details on creating SAS Transport and PDF files.

**Folder specifications**

The specification for organizing datasets and their associated files in folders is summarized in the following figure.




**Specifications for specific datasets and documentation**

**Data tabulation datasets**

**Definition**

Data tabulations are datasets in which each record is a single observation for a subject.

**Specifications**

Specifications for the Data Tabulation datasets of human drug product clinical studies[1], are located in the Study Data Tabulation Model (SDTM) developed by the Submission Data Standard working group of the Clinical Data Interchange Standard Consortium (CDISC)[2]. The latest release of the SDTM and implementation guides for using the model in clinical trials are available from the CDISC web site at www.cdisc.org/models/sds/v3.1/index.html.

[1] Drug products also includes biologic products reviewed in CDER

[2] CDISC, www.cdisc.org, is an open, multidisciplinary, not-for-profit organization committed to the development of worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development.

This SDTM is currently being tested for clinical studies involving biologic products and for animal toxicity studies. The implementation guide for using the model for animal toxicology data is being developed by the Standard for Exchange of Nonclinical Data (SEND) consortium. This implementation guide will also be available on the CDISC web site. Updates to the SDTM and implementation guides as a result of this testing will be available through the CDISC web site.

Each dataset is provided as a SAS Transport (XPORT) file.

**Data Listings**

**Definition**

Data listings are datasets in which each record is a series of observations collected for each subject during a study or for each subject for each visit during the study organized by domain.

**Specifications**

Each dataset is provided as a SAS Transport (XPORT) file. Currently, there are no further specifications for organizing data listing datasets.

**Subject profiles**

**Definition**

Subject profiles are displays of study data of various modalities collected for an individual subject and organized by time.

**Specifications**

Each individual patient's complete patient profile is in a single PDF file. Including the patient ID in the file name will help identify the file. Alternatively, all patient profiles for an entire study may be in one file if the size of each individual patient profile is small and there are not a large number of patient profiles needed for the study. If you do the latter, bookmark the PDF file using the subject's ID. Including the study number in the file name will help identify the file.

**Analyses datasets**

**Definition**

Analysis datasets are datasets created to support specific analyses. Programs are scripts used with selected software to produce reported analyses based on these datasets.

**Specifications**

Each dataset is provided as a SAS Transport (XPORT) file. Programs should be provided as both ASCII text and PDF files and should include sufficient documentation to allow a reviewer to understand the submitted programs. It is not necessary to provide analysis datasets and programs that will enable the reviewer to directly reproduce reported results using agency hardware and software. Currently, there are no other additional specifications for creating analysis datasets.

### Annotated ECG waveform data

#### Definition

These are raw voltage-versus-time data comprising the electrocardiogram recording, to which have been attached the identification of various intervals or other features.

#### Specifications

See the HL7 normative standard for creating the annotated ECG waveform data files. This information may be found on the HL7 web site www.hl7.org.

### Data definition file

#### Definition

The data definition file describes the format and content of the submitted datasets.

#### Specifications

For the specification for providing the data definition files for other datasets, see *Providing Regulatory Submissions in Electronic Format –NDA* for details.

## Annotated case report form

#### Definition

This is a blank case report form annotations that document the location of the data with the corresponding names of the datasets and the names of those variables included in the submitted datasets.

#### Specifications

The annotated CRF is a blank CRF that includes treatment assignment forms and maps each item on the CRF to the corresponding variables in the database. The annotated CRF should provide the variable names and coding for each CRF item included in the data tabulation datasets. All of the pages and each item in the CRF should be included. The sponsor should write *not entered in database* in all items where this applies. The annotated CRF should be provided as a PDF file. Name the file *blankcrf.pdf*.