

SEND Implementation FAQ

Frequently Asked Questions on SEND Implementation

Table of Contents

Basics	3
Timing/Regulatory.....	3
When Was SEND Released?	3
When Is SEND Mandatory?	3
Submission Questions.....	7
What is the status of FDA pilots (CDER, CVM, CBER)?	9
Technical Rejection Criteria	10
GLP vs Non-GLP.....	11
Protocol, Report, QA, QC Impact.....	11
nSDRG	12
What is an nSDRG? Must one be sent with each SEND package?	12
Working with SEND Files	14
Are there publicly available sample SEND datasets?	15
What to Include in a SEND Dataset or SEND Package?	15
Miscellaneous	17
Controlled Terminology (CT).....	18
Modeling Questions	20
Define File.....	24
How can I view a Define File? Where can I find a stylesheet/style sheet?	25
Which variable(s) should be described in value-level metadata?.....	28
What makes up a good dataset key?	29
Domain Specifics	31
When is it ok to include a simplified vs a full TS?	31
SEND Future.....	33

Is ____ Included?.....	34
<i>SEND Issues</i>	<i>36</i>

SEND Implementation FAQ

Welcome to the knowledge base of Frequently Asked Questions regarding SEND.

This page contains answers to a wealth of questions, primarily geared toward those beginning or going through their initial implementation.

If this is your first time visiting the page and you are brand new to SEND, consider checking out the set of pages, which provides some basics on concepts and how they work together.

If you already know the basics and are looking for a question, check out the table of contents on the right, or a "find on page" (Ctrl+F) for a keyword (e.g., "ts.xpt" or "derived" or "ORRESU")

If you can't find the answer you are looking for, ask on the

Check out and watch the page for updates regarding SEND guide versions, FDA news, etc.

Note: The content of this page was prepared by the PHUSE working group and SEND team members and should not be considered as official FDA responses. This content exists to concisely summarise answers that are usually available across/within other documents or pages, to provide implementers with quick, unofficial, and useful answers to their questions.

Table of Contents

Basics

What is SEND?

SEND, or the Standard for Exchange of Nonclinical Data, is an implementation of the CDISC Standard Data Tabulation Model (SDTM), which specifies a way to present nonclinical data in a consistent format.

Timing/Regulatory

When Was SEND Released?

NOTE: Please see the When Is SEND Mandatory? Question below for information regarding when any of the above are required, with the being the official resource and guidance.

When Is SEND Mandatory?

The FDA submission requirement depends on when the study starts, the type of submission, and where in the eCTD it would go.

See the following sections below:

Submission Types and Study Start Timing - this section covers timing with regard to submission types and study start dates

eCTD Sections - this section covers which sections of the eCTD apply

Documents - for supporting documentation for the above

Part 1: Submission Types and Study Start Timing

Each version of the SENDIG also has its own specific window of when it may first be used and when it may last be used (aka sunset). The definitive list of supported/mandated versions is maintained by the FDA in the FDA Data Standards Catalogue (which provides links to relevant guidance, etc.):

This is your first stop for understanding the required dates for published standards, including CDER vs CBER. Submission types, start and ends of versions, and so on.

Since SEND has already been required since at least 2017-12-17 (INDs) and 2016-12-17 (NDAs), as to the question of "when do I first need to be SEND-ready", the answer is "now".

Note: dates for other versions, updates, etc. follow a different implementation timeline (notably, removing the distinction between IND and NDA/BLA after SENDIG 3.1); please see the "When new study types or versions of the SEND Implementation Guide are brought online, when will they be required?" question below.

Note:

Since studies included in an IND are nearly always included in the subsequent NDA, many organisations prepare to have SEND for all studies intended for an NDA or IND submission that started on or after the 2016-12-17 date.

These milestones apply to each study individually. Some submissions may span many years; for these, only studies that start after the dates above are mandated to be in SEND.

That said, the preference is that, where feasible, all repeat dose, single dose and carcinogenicity studies would be submitted with SEND datasets even if not technically required, since it can improve the review process.

Any questions can always be sent to or.

Examples:

NDA submitted to CDER before 2016-12-17 - does not need to be submitted in SEND format (although it is preferred)

NDA submitted to CDER after 2016-12-17 - needs any studies which started after 2016-12-17 to be submitted in SEND. The other studies in the submission do not need to be submitted in SEND format (although it is preferred)

IND submitted to CDER before 2017-12-17 - does not need to be submitted in SEND format (although it is preferred)

IND submitted to CDER after 2017-12-17 - needs any studies which started after 2017-12-17 to be submitted in SEND. The other studies in the submission do not need to be submitted in SEND format (although it is preferred)

Part 2: eCTD Sections

The [FDA website](#) (which can be reached from the [FDA website](#) as well) has several resources for submission, including the Technical Rejection Criteria for Study Data.

This is the definitive source for which sections of the eCTD are applicable for submission of data (including SEND datasets). Currently, this is specified as WILL APPLY and WILL NOT APPLY.

The Electronic Common Technical Document Specification (eCTD), including the full list of possible sections in which a study may be submitted, can be found in the [eCTD Specification](#).

Part 3: Supporting Documents

These requirements to produce SEND datasets for the FDA hinge around the following documents:

The effective date is the 1st of October, 2012, including the fifth authorisation of the [FDASIA](#). (Note that the electronic format for submissions is addressed in FDASIA Sec 1136, which amended)

The “Guidance for Industry - Providing Regulatory Submissions in Electronic Format - Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act” (aka "final guidance") was finalised on December 18, 2014.

The “Guidance for Industry - Providing Regulatory Submissions in Electronic Format - Standardised Study Data” (aka "eStudy Data guidance") was finalised on December 18, 2014.

For more information, see the following FDA pages:

The Japanese Pharmaceuticals and Medical Devices Agency (PMDA) is working to adopt CDISC standards, including SEND. At the April 2014 CDISC Europe Interchange, plans were presented for adoption at some point after FY2017. For more information, see the following PMDA page:

Timing for New Versions / Study Types / CT

When new study types or versions of the SEND Implementation Guide are brought online, when will they be required?

When will Safety Pharm be required?

When will Respiratory and Cardiovascular be required?

When will Repro be required?

When is SENDIG 3.1 required?

How long is SENDIG 3.0 valid?

(Note: this pertains to updates made after the initial SEND requirement laid out in the previous question, e.g., IG updates, new study types, etc.)

After new standards or updates are published, pending an evaluation by CDER, CDER will add the standard to the page with a timeframe for the requirement. The timeframe for these will be at least 12 months after the standard/version is added to the page, and will apply only to new studies. It is expected that larger-scale additions (such as completely new subject areas) will have a longer timeframe for Sponsors to implement and ramp up before they become required.

Note:

To get updates (highly recommended), click the "Sign up for email updates" link at the top of the page.

At any given time, within a class of submission (e.g., NDA vs IND), only one version of a document will be officially required. For instance, as soon as SENDIG 3.1 becomes required for NDAs, SENDIG 3.0 is obsolete for NDAs.

For studies which started when an older version was required (compared to what is required at the time of submission), sponsors have the option to submit with the newer version. For example, say a chronic starts before the 3.1 requirement date, but by the time the study finalises, 3.1 is now the required version for studies that start. In this example, the Sponsor may choose to submit the study in 3.1 (instead of 3.0 as technically required).

As of 2018-01-23 (per), any updates will follow a schedule as follows after the first Transition Date, after inclusion on the Study Data Standards page (with Transition Date next March 15th after Study Data Standards inclusion date):

Updates: 12 months after Transition Date

New Standards: 24 months after Transition Date

Example - below gives an example of a standards update and the resulting requirement date...

2018-08-10: a revision to a standard is published by CDISC in 2018-08-10

2018-09-04: It completes review by CDER and is added to the page

2019-03-15: Transition Date, since this is the first March 15th after the Study Data Standards inclusion

2020-03-15: requirement starts

Then...

Up to 2018-08-10 (CDISC publish): Sponsors have the opportunity to be a part of the development and public comment periods on the new standard/version leading up to the publication from CDISC, and can anticipate CDER adding it to the Study Data Standards Resources page soon.

From 2018-09-04 (Study Data Standards inclusion) to the requirement date (2020-03-15), sponsors have this time to gear up implementation for the updated version.

For a new submission submitted to CDER before 2020-03-15, studies within do not need to adhere to the new standard, although it is encouraged/preferred

For a new submission submitted to CDER after 2020-03-15, studies within the submission which started after 2020-03-15 need to adhere to the new standard. The other studies in the submission which started before 2020-03-15 are not required to be submitted according to the new standard, although it is encouraged/preferred.

When will new Controlled Terminology be required?

New CT will be required for studies "within a reasonable timeframe" from the new CT's release date. What is "reasonable" is open to interpretation, but we recommend keeping within a year of the CT's release date when packaging a study.

Note that this stipulation applies to CT active at the time of the creation of the SEND package for the study. For instance, if a SEND package is created for a study in 2013 and not submitted until 2017, the CT to which it must adhere is the CT active at the time of the packaging (e.g., 2013 or shortly before it). There is no requirement to retroactively update past studies with CT that comes out after finalisation.

Visit the page to get the most recent CT.

When can SEND replace TUMOR?XPT in FDA submissions?

The intent is to phase out tumour.xpt in the future, so that it can be generated from a SEND package. However, it is currently still active. As always, consult the Technical Conformance Guide (as referenced in the) to see what is or isn't required.

Submission Questions

What applies to SEND

How do I know whether SEND is mandatory for any given endpoint?

Do I need to include data for unmodeled endpoints?

Is endpoint ____ required for a study type that isn't required?

For an endpoint to be required, it must meet two criteria to be considered mandatory for FDA submissions:

The study type is one explicitly called out in the FDA Study Data Standards

The endpoint is one that is currently covered/modelled by the domains in the SEND Implementation Guide

In other words:

If the endpoints are not currently modelled, it doesn't matter if it happens to be under the right study type – it doesn't have a hard requirement.

If the study type is not one asked for, then it doesn't matter if it happens to be a modelled endpoint – it doesn't have a hard requirement.

However, this is purely speaking to what is mandatory. The FDA has stated on numerous occasions that they still would prefer to receive data for both of the above exceptions (e.g., via custom domains or to have packages for non-required study types if they have fitting domains). Please contact us for additional advice about such a submission.

Is SEND mandatory for study types that were not piloted?

For example, if there is a safety pharm study with body weights, will the body weights need to be submitted?

The Study Data Standards page provides current expectations/recommendations. Note that while typical study types will be piloted, a study type need not be piloted to be specified by the Study Data Standards page to be mandatory.

However, electronic submissions of data are encouraged, even when the study type is not yet mandatory. Please contact us for additional advice about such a submission.

For studies I submit with SEND datasets, what is the FDA's recommendation for including non-SEND datasets? (e.g., custom domains)

Is it required to submit data not modelled in a domain yet?

Generally speaking, from the industry side, it is not considered valuable to provide custom domains, given the issues with nonstandardized data (nonstandard format; can't use across organisations, etc.). Additional domains not part of SENDIG would still be present in individual tabulations (e.g., PDF) for submission coverage purposes.

However, in general, this is encouraged by the FDA. Please contact us for additional advice about such a submission.

What exactly needs to be included in a complete SEND package that is ready to submit to the FDA?

See the for more information on getting started.

A typical package includes:

The basic minimums specified in the SENDIG (usually TS, TX, TA, TE, SE, DM, EX)

Whatever endpoints you reported which have a domain in SEND

The define file (usually define.xml)

An nSDRG (Nonclinical Study Data Reviewer's Guide)

In addition, for the submission, the following are generally needed:

A cover letter (e.g., summarising what's in the submission, reiterating some of the information provided when initiating the submission process, etc.; see FDA site/contact FDA for more information)

Other FDA Questions

What is the status of FDA pilots (CDER, CVM, CBER)?

Please see the page for pilot status.

How will the FDA use SEND files?

The FDA uses the files for the review process via the Nonclinical Information Management System (NIMS) suite. This suite provides tools that are built to use SEND datasets' information, so that they are able to review a submission more efficiently than when they receive only PDF or printed submissions that contain the individual animal data. Before loading the files, their gateway will first perform validation checks against the data to make sure they are SEND-compliant. Note that validation in this sense is not computer validation but rather a series of nuts and bolts checks, such as whether required variables are missing or have an unsupported value in a field that requires Controlled Terminology.

Is SEND only a U.S. requirement?

Will the EMA (EU) require SEND?

Will the PMDA (Japan) require SEND?

SEND will only be a requirement in the United States for certain FDA submissions. However, it has operational use, such as transfer between organisations, sponsor warehousing, etc., such that it is a good idea to produce SEND datasets, even if not technically required for submission.

As far as the European Medicines Agency (EMA) goes, the Clinical Trial Advisory Group on clinical trial data formats (CTAG2) is working on advising the EMA on clinical data formats, where it is leaning toward CDISC standards (although if it accepts, it would likely follow a similar progression as the FDA, with a 2-3 year pilot. Here is a link to the recommendations CTAG2 provided the EMA: .

As of 2016, PMDA has put forward a schedule for requiring SDTM on the clinical side and plans to explore the nonclinical side as well (with a possible pilot).

What software is used by reviewers to visualise and review SEND data?

There are several tools. FDA does not endorse any particular vendor or tool.

There are SEND solutions available from nonclinical data acquisition software vendors (e.g., Instem, PDS, PointCross, Xybion) which provide the ability to produce SEND datasets, with varying levels of analysis/review options available.

Does the FDA mandate or endorse the use of a specific validator like the one from Pinnacle 21?

It has been stated in open forums that the FDA CDER does not, and does not intend to, have a required or preferred validator. However, the validation rules developed jointly by the FDA and industry will be published in the future. A variant of these rules is currently used by the Pinnacle21 Validator tool and can be viewed through its configuration. Organisations are free to build their own validator tools to build on the rules, though, such as to validate against organisation-specific data cases, provide additional checks for incoming data that are to be consumed, etc. Validation rules aside, the SENDIG provides the official rules for what comprises a SEND-compliant package, so the implementation guide takes precedence over any discrepancies between the implementation guide and validation rules.

What data mining opportunities will SEND enable?

Data standardisation is the first step in the chain to realising cross-study querying and data mining. SEND is expected to open the door for such data mining; although the benefits that can be realised will not be discovered until either a sufficient time has passed to create a significant enough database of historical data or studies are converted and loaded into repository systems to facilitate such queries. One of WG6's foci has been to identify these key areas and facilitate and drive progress.

Technical Rejection Criteria

As a reminder, the Technical Rejection Criteria are available on the.

When do the Technical Rejection Criteria go into effect?

As of writing this (2019-06-17), it is not yet officially in effect. In theory, it will be 30 days after publication on the (or potentially, the).

Which TSPARMCDs are required by the FDA?

Just TSPARMCD=STSTDTC is explicitly required (for legacy studies under covered study types, e.g., single dose, carc, etc.). All other studies should have a full TS.

There are new variables added in the base SDTM after the SENDIG was published. Will the FDA accept submissions with these not included?

There were several variables added in SDTM v1.3, e.g., TSVALNF, TSVALCD, TSVCDREF, and TSVCDVER. Will the FDA accept submissions with these not included?

If it is for legacy data submission, you don't need to include those variables for TS. For data submissions, please create a TS according to the IG you use.

Is the TS.xpt file for studies that start before 2016-12-17 required beginning on 2016-12-17? Data responded 2018-01 to individual organisations that the short version TS is not required for nonclinical studies unless XPT files are provided. However, the Technical Rejection Criteria document has not yet been updated to reflect this change.

As to officially posted policy, if you submit SEND after 2016-12-17, TS is required (see "When is it ok to include a simplified vs full TS?" question for more details). For legacy cases (e.g., only the report), then only the short version TS is required (single row TS; no DM or Define). All other studies should have a full TS.

The Technical Rejection Criteria can be found on the. On this page is a TRC Self-Check Worksheet you can follow to check what you need.

GLP vs Non-GLP

Does SEND cover non-GLP studies?

The short answer is yes. Whether SEND is required is independent of whether the study is GLP or non-GLP. So, if this type of study were SEND-applicable if it were GLP, then it's also applicable for it as non-GLP.

Does 21 CFR Part 11 apply?

The short answer is yes. It is a critical component of GLP Validation, and there are plenty of cases where GLP Validation applies to SEND datasets.

Minimum requirements-wise, it is per study and case that would dictate whether GLP Validation (and 21 CFR Part 11) apply (e.g., if used for pre-final purposes to affect study decisions, etc.).

Many choose to validate the software producing all datasets (with 21 CFR Part 11 coverage inherent) even if used for a mixture of cases, as the intent to use them may vary, and it's easier to validate the single process.

Protocol, Report, QA, QC Impact

How will my study reports from a CRO change when SEND files are used?

For the near future, they will not likely change. However, it is a longer-term goal for the SEND datasets to eventually replace the individual tabulated datasets. The main body of the study report, the summary tables and other appendices would remain in their current format, though.

Should SEND be listed in the Protocol or treated only as a contract deliverable?

We recommend that SEND be listed in the Protocol to ensure that the endeavour is resourced properly and that expectations are set and met.

Please see the page for more details and discussion around this question.

Should SEND datasets be sent through QA review?

Depends on whether they are bound by GLP.

Please see the page for more details and discussion around this question.

What QC should be applied to SEND datasets?

Depends on whether they are bound by GLP.

Please see the page for more details and discussion around this question.

Will the study director's signature on the report also indicate accountability for the SEND datasets?

Depends on whether they are bound by GLP.

Please see the page for more details and discussion around this question.

What documentation is required for SEND datasets created retrospectively (after report finalisation)?

Depends on whether they are bound by GLP.

Please see the page for more details and discussion around this question.

nSDRG

What is an nSDRG? Must one be sent with each SEND package?

An nSDRG is a document which is meant to aid the reader in understanding the SEND dataset in the context of the study report. The Technical Conformance Guide published by the FDA states that it "...is recommended as an integral part of a standards-compliant study data submission."

What is the proper format of an nSDRG? Does it need to contain anything specific?

While the content and format of an nSDRG are not mandated, the FDA has provided some expectations in the Technical Conformance Guide.

nSDRG: Describing Conformance Issues - Section 5

What kinds of information do Reviewers find useful in this section?

The information in this section has the most value to the Office of Computational Science's (OCS) Data Loading and KickStart teams. Also, Business Rule errors are very important to the statistical reviewers in the Office of Biostatistics (OB) for the submission of carcinogenicity studies. Nonclinical reviewers will also find value in Business Rule errors related to carcinogenicity studies (e.g., BR FDAB072).

Do you expect that explanations of issues in section 5, when impacting data content, are referenced or further explained in the dataset explanations (section 4.2)?

Yes, any issues referenced in Section 4.2 should be referenced in greater detail in Section 5. Please also state that there is additional information in Section 5 when listing the issue in Section 4.2.

Are there some examples of what has been seen in NSDrg and considered “too technical”? (in messages or explanations of validation issues.)

There is no information that is considered too technical for use by OCS. OCS staff will use the information provided to help with some aspect of the review. It is best to provide as much specific information as possible rather than boilerplate language.

Should all validation issues be described in this section, including those determined to be “false”? (false = errors or warnings that result from bugs in Pinnacle 21 Community or other validators that may be used. Some examples of common rule breaks that are false in nature: Errors: DD0059, DD0028, DD0024, DD0064. Warnings: FDAN037, FDAN169, FDAN212, FDAN218, FDAN341, DD0029, PCO497.)

All issues identified by the tool used, regardless of whether it is “false”, should be identified by the sponsor in Section 5.

Are there particular domains which need special consideration regarding the handling of conformance issues (such as MI and Tumour data, where there are a lot (14) of business rules)?

Yes, there are times when cross-domain checks cannot be automated, or specific Business Rules do not have automated checks. The Business Rules should not be ignored for these non-automated checks. For carcinogenicity studies, these non-automated checks against 14 Business Rules should be performed manually, with any errors noted in Section 5. For other types of studies, if a sponsor is confident that the process used to create the SEND datasets aligns with these non-automated rules by establishing that multiple datasets have no errors, manual checking of each dataset is not required.

Do people who do “data loading” use the nSDRG?

Yes, if a Data Management and/or Data Loading teams have trouble loading the data, the nSDRG is one of the first places the teams will look for additional information on conformance issues.

Is the FDA tracking validation rule violations by quantity and/or rule ID?

OCS staff is tracking validator rule violations by quantity, but is not tracking the specific rule ID associated with them.

Are trends emerging to suggest certain rules could be either A) impractical for companies to adhere to, or B) too vague in meaning? For example, a high variety of explanations could indicate a wide misunderstanding of the rule or its application.

The FDA noted several common error trends in datasets submitted by sponsors. The FDA clarifies how to correct the errors in industry meetings and other public forums, as well as updates the Technical Conformance Guide (TCG) to ensure the FDA’s needs are met.

Working with SEND Files

Using Files, File Formats

What's in a SEND Package? What is a SEND File? What is a SEND dataset?

A SEND package consists of a number of dataset files (in XPT format, a.k.a., SAS v5 Transport format) and a define.xml file (which provides information about what's in the datasets).

If I receive a SEND file (i.e. from a CRO), how do I open it/view my data?

How do I open XPT files?

To open XPT files, you have a few options:

Download and use the free. This was previously known as the SAS Viewer, so if you have the SAS Viewer tool, that works in the same way. Note that this tool is very limited in features.

Open in SAS using the XPORT libname option, e.g.:

```
libname MyLib XPORT "C:\test.xpt"
```

```
data test;
```

```
Set MyLib.test;
```

```
...
```

Open in R

When opened, they appear similar to Excel workbooks.

Various other vendors have products that will allow you to view SEND datasets as well. These tend to be more robust in visualisation/analysis capabilities, and enabling review.

How can I create a ts.xpt file with the Study Start Date?

The nonclinical script group has prepared an R script to help do this. It is available from. To run this script, you will first need to obtain R and confirm you can use it to read an XPT file. Once you are successful with that, do the following.

Open a command prompt and enter the command: `java -version`

If Java is installed, the version number should be displayed. If it isn't displayed, install Java. If you have a 64-bit installation of R, you also need a 64-bit installation of Java.

Copy the files from the aforementioned GitHub folder to "c:\temp\r testing".

Create an empty sub-folder named "C:\Temp\r testing\xpt output".

Launch RGui

Run the script by selecting the menu options "File" -> "Source R code..." and selecting the file "c:\temp\r testing\CreateTsXPT.R"

Confirm you have XPT files in the appropriate sub-folders of the "XPT output" folder.

Modify the Excel file and/or script as needed to get the results you need.

What should the catalogue (dataset) name be for the SAS Transport files?

The catalogue (dataset) name should be the same as the name of the XPT file. For example, a SAS transport file named BG.XPT should have a catalogue name BG.

Will the XPT files be replaced with something easier to work with, like XML?

Yes. Please see the "When will the XPT files be replaced with XML?" question further below under the SEND Future section.

Are there publicly available sample SEND datasets?

Here is a list of places to obtain sample datasets:

The PHUSE GitHub has a number of example datasets available:

Contact [with a subject line of "Send me SEND"](#) to get an FDA-validated SEND dataset for an example 28-day toxicity study

Go to [download an FDA-validated multi-organisational SEND dataset](#).

Many CROs have sample datasets they can provide to evaluate their capabilities or prepare for your own implementation

There are also many examples in the SENDIG of domain records.

What to Include in a SEND Dataset or SEND Package?

Should I include all variables shown in the implementation tables in my dataset?

The "Core" column in the SEND implementation guide defines for each variable whether you must include the variable in your dataset. You must include all variables that are listed as Req (a.k.a., required, meaning not nullable) and Exp (a.k.a. expected, meaning with or without data, but should have a good reason if not populated). Include variables listed as Perm (a.k.a., permissible, meaning they can be excluded from the dataset if you do not collect them) if you have data to report in them in the domain dataset. Note that a few variables listed as permissible are either/or (e.g., AGE and AGETXT in DM); in these cases, you are expected to provide at least one of two, although they are defined as permissible because either is acceptable. Such a case will be noted in the CDISC notes for the variable(s) and/or the assumptions for the domain.

Should I include pretest/stock animals in a SEND dataset?

The animals included in a SEND dataset should be consistent with the animals included in your study data report. Generally, this means including pretest/stock animals if they have been assigned to a test treatment group and have study data collected.

What study types can/should I use SEND for or include in my SEND package?

As far as possible, the domains modelled in SEND can be applied to any study that has them

(SEND covers many standard endpoints), although study types not yet piloted may have some endpoints not yet covered.

As far as should, per the SENDIG, section 1.1: "This version of the SENDIG is designed to support single-dose general toxicology, repeat-dose general toxicology, and carcinogenicity studies."

These are the types of studies covered in the CDER pilot. A pilot is open with the CVM group, and pilot plans are being discussed for CBER and other FDA divisions. Additional study types are currently being modelled (repro, safety pharm), during which pilots will be conducted, and updates to the SENDIG to include those study types will be completed. However, note that the domains modelled in SEND can be applied to any study that has them, as SEND covers many standard endpoints, although study types not yet piloted may have some endpoints not yet covered.

Working with CROs

What Should I Ask My CRO?

A number of considerations arise when initiating a conversation with another organisation around SEND file production. The page has an extensive "Points to Consider" question list to help smooth this process.

Working with Multiple Files/Studies/Versions

Should the SENDIG and/or CT version used be the same for all studies in a submission?

Not necessarily. The only requirements are that the same SENDIG/CT version be used within a study, and that the version is what is (reasonably) up-to-date as of the creation of the package. Especially with submissions spanning many years, it is very possible for there to be different SENDIG versions across studies, as well as different versions of CT. The only expectation per the draft guidance is that the versions used at the time of creation are current within a reasonable timeframe (what is "reasonable" is being formulated and is expected to be stipulated in the final guidance).

What do I need to do when collating datasets for a domain?

For instance, when two independent labs contribute lab test data toward LB...

In some cases, pieces of a domain may come from different sources, such as different labs or different systems. When bringing the data together into a single dataset for the domain, here are some things to keep in mind:

--SEQ may need to be re-sequenced so that no two rows have the same value within an animal

And if you do, RELREC records based on --SEQ values may also need to be updated in the same way

--NAM may need to be populated to distinguish between labs performing the testing

How is versioning handled for different versions of a SEND package for a study?

When providing interim datasets, do you provide deltas or full loads?

First, SEND packages are technically only considered valid or complete when they contain all data; there are no specifications on providing partial data (e.g., deltas between versions). Subsequent versions of a package (version 2, 3, etc.) would be cumulative from past versions, including the existing data and any data collected since the last package.

In a number of operational cases, it is necessary or useful to be able to tie together records between a current version and the version prior, such as for the detection of deltas. There are a couple of options for this:

To definitively identify a record as being the same record across versions of a package, use the RECID variable (invariant record ID), new in SENDIG 3.1. Values might typically be sourced from internal database IDs for the source records, which do not change over time. If this is feasible to provide, this is preferable, since it then gives the consuming organisation a definitive way to detect which records are new, updated, or removed.

If the consuming body(-ies) is not interested in figuring out deltas, you could just submit the new version of the package in its entirety.

Miscellaneous

Are calculated results reported in SEND?

Depends on the meaning of "calculated results".

Derived endpoints, like body weight gains, etc., are reported in SEND if you would have included those results in your individual tabulations in the submission.

Statistics, like descriptive statistics (mean, standard deviation/standard error, N, etc.), do not apply even if they may be present on those tables.

Why are there entries in SEND files such as body weight gains and organ weight ratios, when that information can be derived from information in other SEND files?

BG and the relative organ weights in OM were added because they are used in most submissions as individually reported endpoints (and SEND is focused on the individual animal data). Including them as separate endpoints removes any ambiguity or duplicated calculation on the reviewers' part. In the future, these endpoints may be removed as analysis modelling matures for nonclinical data (e.g., through ADaM).

How do I manage unscheduled data (i.e., data collected in an unscheduled interval)?

For unscheduled findings, leave the VISITDY (the planned study day) blank. TPT and TPTNUM values would also be null.

How do I indicate derived records?

When should I use --DRVFL?

What is --DRVFL used for?

Can I use --DRVFL to indicate any derived records?

There are two flavours of derived to consider.

When the record in the dataset is derived from other records in the dataset, then (and only then) you can use the --DRVFL variable to delineate between the derived record and the ones going into the derivation (one example could include blood pressure readings, where the result record is picked or averaged from 3 constituent readings, and so the flag serves to indicate which is the derived record (and in this case, the one to keep). It cannot be used if the record was derived from information outside the dataset; this constraint means it has limited use.

When the value is calculated through a calculation, algorithm, etc., from data often outside the dataset, then this status can be indicated via the define file (i.e., whether the test was collected, derived, etc.).

SENDIG 3.1 and beyond describes this in more detail.

Controlled Terminology (CT)

For more information on Controlled Terminology, check out the page.

Controlled Terminology Versioning

How frequently are controlled terminology files updated?

CDISC/NCI releases controlled terminology in "packages." New packages are released as needed throughout the year; generally, 2-4 new packages are released a year.

Where can I find the most recent controlled terminology?

SEND terminology is available for direct download from the NCI File Transfer Protocol (FTP) site in Excel, text, odm.xml, PDF and HTML formats.

Where can I find old controlled terminology versions?

All published versions of SEND controlled terminology are on the NCI File Transfer Protocol (FTP) site. The date included in the file name is the date of publication.

Should the controlled terminology version used be the same across a single study?

Are multiple controlled terminology versions okay within a study?

It is required that within a study, only one version of controlled terminology is used.

Additionally, it is expected that data from multiple contributors (e.g., different CROs contributing data to a study) is aligned for the study.

Should the controlled terminology version used be the same for all studies in a submission?

Not necessarily. The only requirements are that the same controlled terminology version be used within a study, and that the version is what is (reasonably) up-to-date as of the creation of the package. Especially with submissions spanning many years, it is very possible for there to be different CT versions across studies and even different versions of the SENDIG. The only

expectation per the draft guidance is that the versions used at the time of creation are current within a reasonable time frame (what is "reasonable" is being formulated and is expected to be stipulated in the final guidance). The same considerations apply to versions of the SENDIG as well.

How do I know what version of controlled terminology was used with a dataset?

In the TS (Trial Summary) domain, when TSPARMCD=SENDCTVER, the TSVAL variable contains the SEND controlled terminology version.

How do I get the Pinnacle 21 Validator to validate against a particular version of CT?

The Pinnacle 21 Validator runs off the same terminology files that you can download (e.g., "SEND Terminology.txt") and comes automatically packaged with whatever version is active the last time the Validator was published (and so it can be out of date). To update the CT against which it validates, or to swap in a particular version of the CT, please see the page under the "Updating the Pinnacle21 Validator Configuration" section.

Controlled Terminology Mapping

Does the case of the controlled terms matter?

Yes. When using controlled terminology in a SEND dataset, you must use the submission terms exactly as they are listed in the controlled terminology file.

Do I need to change my units when mapping to ORRESU?

Do I need to convert units to map to ORRESU?

When mapping to ORRESU, the key is to map to the same unit concept (but the label might be different). The specific unit label used by an organisation may differ from the SEND CT preferred label; however, the same concept is there (just as a synonym). For example, for the gram unit, a sponsor might use a label of "grams" or "G" internally, but the SEND preferred term is "g" - this is still the same conceptual unit, just with a different preferred label. Another example is the unit label of "ng/mL", whose submission value is "ug/L" - same unit, just a different label ("ng/mL" is in the synonyms list for "ug/L"). If you are mapping to SEND and do not see your unit represented in the CDISC Submission Value column, check the CDISC Synonym(s) column as well. As this is just a label change, at no point should a value conversion be performed (or needed) on your original result value. For example, if you collected a value of "30 ng/mL", your value is still 30, just with a different label for the unit.

If you have a unit that is simply not on the Unit codelist in any shape or form, then you have a case for a new term request, which can be submitted to the

Are all controlled terminology for tissues in the singular form (KIDNEY)?

In general, yes. The --SPEC variable to which tissues are mapped represents the material type of the specimen, which is typically inherently singular (there are some exceptions for cases where the tissue is generally considered as a unit, e.g., MENINGES of the brain). Plurality for most tissues is defined through the qualifying variables of --LAT (e.g., LEFT, RIGHT, BILATERAL, UNILATERAL) and, less commonly, --PORTOT (e.g., "MULTIPLE", "SEVERAL").

What do I do if I have terms that are used in our organisation that do not map to controlled terminology?

If the variable that you will report the term in uses controlled terminology, and the associated controlled term list is not extensible, then you must find a controlled term to which to map your term (if you do not, you will get validator errors). If the controlled term list is extensible, then you can report the term that you use, so long as it meets the basic requirements of the variable field for length and characters used, and you should also suggest your term for inclusion in Controlled Terminology. The web page handles suggestions for both new terminology and changes to existing terminology.

The implementation guide mentions "ISO 8601 format" and "ISO 8601 character format". Is there a difference?

No. ISO 8601 format, as it is referenced in SDTM, refers to a standardised way to specify a date or datetime in character format. All date/datetime fields in SEND adhere to this format. The SENDIG provides some guidance on how to do this in the "FORMATS FOR DATE/TIME VARIABLES" section.

How are --TEST and --TESTCD related?

How do I link a --TEST term to a --TESTCD term?

Certain variables in the SENDIG are paired, in which case, their values strongly interrelate. In the case of --TEST and --TESTCD, it is actually IG-enforced that this is the case, and it is a 1:1 relationship. However, especially in domains whose paired variables have hundreds of terms, it can be unclear as to how to link terms that are intended to be paired with one another. In these cases, the synonym is set to the same value between the two terms, and so this can be used to determine which corresponding term should be used. So, typically speaking, you would start by finding the term of interest in the --TEST list (as the long text is more human-readable). From there, you would record what it has in the synonym field. And then finally, within the --TESTCD list, you find the --TESTCD term that has that same value in its synonym field.

What do I do when I have two different --STRESU values for the same --TEST?

Currently, there is a warning/error in some validator applications if you have differing --STRESU values within a --TEST, which is unfortunately, a very real possibility, such as when the standard unit differs based on --CAT or --METHOD, as is the case with the LB domain. This is a known issue; for now, populate with the values you intend and be prepared to explain the valid reason why this rule triggers.

Modeling Questions

Modelling: SEND-wide

Does the order of rows in a domain file matter?

No

Does the order of columns in a domain file matter?

Columns should be included in a SEND domain file in the order they are listed in the SEND Implementation Guide's domain tables.

Does the value of --SEQ matter?

The only requirements on --SEQ are that it is (1) an integer and (2) unique for each record within an animal. This means that each animal could have *a* record with a --SEQ value of 1, but no two rows for that animal could have the same --SEQ. One option for populating --SEQ is to make it unique for the entire dataset. This method covers the requirements of --SEQ and also provides a unique key for the dataset.

What is --GRPID used for?

--GRPID is used to link together records in a single domain for a single subject. The meaning behind a --GRPID value is entirely up to the sponsor. While specific to a domain, --GRPID can also be used in conjunction with a RELREC relationship to link those records to records in other domains. For example, a --GRPID could be set in the CL domain to link together three clinical observations, and then a RELREC relationship based on that --GRPID and a record from another domain, to establish the relationship with a single record, instead of individually making a relationship between each record.

I have to make a SUPP domain entry because the entry in my domain field exceeds 200 characters. What do I use for QNAM if the domain variable is already eight characters (like LBREASEX)?

In cases where the standard domain variable name is already eight characters in length, sponsors should replace the last character with a digit when creating values for QNAM. As an example, for Reason for exclusion in LB (LBREASEX), values for QNAM for the SUPPLB records would have the values LBREASE1, LBREASE2 and so on.

How are records which are scored (like some clinpath, dermal, or neurotox tests) supposed to be presented?

Include the score as the result (--ORRES, --STRESC) as you would on your tables. Typically, these scores do not represent a true number, so typically, --STRESN would not be populated. The meaning behind the scores (e.g., that 1 means "Quivers of limbs, ears, head or skin") would be provided in the define.xml file as a CodeList with their coded and decoded values.

Why is QEVAL expected in SUPP-- domains, but --EVAL is permissible in all other domains where it is present?

This allows all SUPP-- domain files to have a consistent structure: all variables will be present in all SUPP-- files since all of the variables are either required or expected.

If animals are scheduled to have findings in the future (e.g., at terminal sacrifice) and the animals are removed from the study early (e.g., unscheduled sacrifice), how should findings be represented (e.g., with VISITDY, --STAT, and --REASND)?

Only in the instance where you carry out a planned assessment according to the plan (within the

grace days that you have defined for that VISITDY), can you populate a record with VISITDY information. All other instances must be considered unscheduled activities and cannot have a populated VISITDY.

Findings that were originally planned for animals removed early from the study are not required to be artificially inserted into the SEND datasets (e.g., with a-- STAT of NOT DONE).

A common example of this is with terminal body weights (TERMBW), which can often fall into the grace periods of long-running studies (in which case, VISITDY could still apply).

When the protocol is amended during the study to change planned activities, when/how should VISITDY be populated?

As soon as the schedule changes by amendment, that is now the plan, and so VISITDY should be populated with the amended planned date(s). For example, if a group is decided to be sacrificed early by protocol amendment, then the corresponding disposition records would have VISITDY populated with the early sacrifice day(s).

How should the --DTC/--DY variables be populated in cases where --STAT=NOT DONE?

When --ORRES cannot be populated, --STAT and --REASND take the place of --ORRES and thereby represent the outcome of the --TEST. In that instance, --DTC and --DY describe the outcome indicated by --STAT and --REASND. If you are collecting timing information about missing results for planned tests, then you should indeed populate the --DTC and --DY with this information (e.g., the date when the planned test was marked missing).

For outsourced studies, should the STUDYID be the study identifier used by the CRO test facility, or should it be the sponsor's study identifier, if they are different?

The following statement is from the CDER Data Standards Questions Team: "It is our position that the STUDYID variable should be populated with the identifier used during the course of the conduct of the study (in this example, the CRO study ID) and that the TSVAL when TSPARMCD=SSPONSOR should be the sponsor's identifier."

Numeric Data (e.g., Measurement Data)

I collect Body Weight in kilograms for males and in grams for females. How do I export the data in SEND?

SEND has different variables for the original versus standardised results:

--ORRES reflects your original result in whatever units under which it was collected

--STRESC/--STRESN reflect your result in the units of your choosing (e.g., if units are standardised for reporting)

So, in the example given in the question, --ORRES would be the result as collected (in kg for males and g for females), and if your reported tables would all be in terms of kg (in other words, you choose to standardize to kg), then your standardized results would be the results in kg form (e.g., males as is and females converted).

Units (--ORRESU and --STRESU) are whatever units apply to the results above, but mapped to their scientifically synonymous Controlled Terminology preferred term. For instance, the original unit of "kilograms" for males would be mapped to "kg" in --ORRESU, and the original unit of "grams" for females would use the CT term of "g" in --ORRESU (note - no unit change; just a label change). Likewise, all results would have "kg" in the --STRESU instead of the "kilograms" label used in reporting.

Do I have to convert my results to different units for --ORRESU and --STRESU?

No. There are no stipulations that you must use particular units, only that you use controlled labels for the units that you do use. The Controlled Terminology that applies to --ORRESU and --STRESU only enforces the preferred label for the same base concept, not a preferred unit you should be using. You should use the units under which you collected your result (--ORRESU) and the units you reported (--STRESU), as mapped to Controlled Terminology preferred labels. For instance, if you collected with a unit of mg/mL, you would use the preferred label of "g/L" (scientifically equivalent to "mg/mL"), and your result (--ORRES) will not change. You should not perform any unit conversions when including your original result/unit, only re-labelling. Consequently, for the standardised results, you would only convert units if you converted units for reporting/submission.

What do I do when the precision differs between collection and reporting?

My original result was collected with X sig figs, but we reported with Y sig figs. What significant figures should I use in SEND?

--ORRES is meant to store the original result, and --STRESN/--STRESC represent the standardised, aka reported, value. In theory, the two should be a unit change away from each other. However, therein lies an issue in that it is possible for the original result to have been collected with a different precision than that used for the reported result.

What to do in this situation for --ORRES is currently not clear; however, --STRESN/STRESC should definitely be populated with the reported result, character/numeric, respectively. For --ORRES, there are several approaches being used, but the predominant one follows the description of the --ORRES variable, which is to populate with the value as collected (in raw precision). As to the counterargument of "but they should be the same value, save a unit change", the counterargument is that such rounding is usually only done to levels where greater precision is considered meaningless scientifically.

In the SEND implementation guide, in the CDISC notes for the --STRESN variable for some domains, it mentions "continuous or numeric results". What is a continuous result in this context?

The SENDIG discusses this under the ORIGINAL AND STANDARDIZED RESULTS section. In summary, though, --STRESN is meant to contain the numeric representation of what is in --STRESC, provided that what is in --STRESC is actually to be considered a number. If the --STRESC values do not actually represent a number but instead a code for something, such as is the case with scores or graded scales, then it is likely inappropriate to populate --STRESN, as these values should not be considered numeric for the purposes of calculations.

How do I handle clinpath results which are above or below the limits of quantification (BLQ), such as <1.0? The SENDIG discusses this under the ORIGINAL AND STANDARDIZED RESULTS section. Briefly, this value is not actually a numeric value, so --STRESN should be left null. However, a value of 1 may actually be used in calculations; the guide directs how to populate a -CALCN variable (usually as a SUPP-- variable) to contain this information.

(SENDIG 3.0 only) What should you do when LBORNRL0 and LBORNRLHI are text (since the variables are Num)?

You can set up those variables as text and then explain them in the study data review guide.

Note: this issue has been addressed in future versions

Categorical Data (Findings Data)

I have food consumption observations on my study, such as "reduced food intake". How do I report these data in SEND?

Currently, the FW domain does not permit observational data. Report these observations in the CL domain (there is an example of this in the SENDIG 3.1 onward).

What do I do with clinical observations modifiers that do not have a specific variable for them (like colour)?

This information will generally be embedded in the CLORRES variable as part of the text string of the original finding. If there isn't a "bucket" for it already, then it is generally not considered important and thus optional (since this information is represented in CLORRES if desired).

However, if it is useful for operational needs to include it as a separate variable, this is exactly the purpose for which SUPP-- was devised. Through SUPPCL, you can add any additional variables you want (although be sure to describe them in your define.xml file).

Note: SENDIG 3.0 says a supplemental qualifier --RESMOD isn't "expected to be used"; this was removed in later versions.

If I have a comment for the FW domain and my data are pooled, how do I set the USUBJID in the SEND CO dataset?

In SENDIG 3.1 and later, the CO domain supports the POOLID variable, so you would leave USUBJID null and populate POOLID like any other domain with pools.

In SENDIG 3.0, the SEND CO domain was not able to use the POOLID. If you are submitting a SENDIG 3.0 submission, you can leave the USUBJID empty (USUBJID is an Expected field, not required). Or include POOLID anyway and explain the error message in the nSDRG.

Define File

Define File: Basics

For an introduction to key concepts around the Define File, see.

What is a define file? Must one be sent with each SEND package?

Each define file is specific to a package for one study and is the roadmap to the overall content for that study (e.g., which datasets and columns are present). It allows you to explain individual anomalies in your data and connect them to a specific field in a specific domain, or it can direct the reviewer to apply the concept to your entire dataset. User-defined controlled terminology is also contained here. The define file must be submitted with each SEND package.

A "define file" is usually an XML file, which is co-located with a stylesheet file that gives instructions to a browser on how to represent the XML file in a nice way.

Can the define file be submitted as PDF or does it need to be XML?

Define.xml must be submitted no matter what.

As for the PDF:

For studies that use define 2.0 and beyond, only define.xml must be submitted.

For studies that started before 2018-03-15 (define 2.0 mandate) and used SENDIG 3.0, a PDF must also be submitted, as well as the define.xml

How can I view a Define File? Where can I find a stylesheet/style sheet?

Opening a define.xml by itself is not really human-readable.

You can either:

(Recommended) Use a tool like the one provided to provide an easy way to view/edit a define.xml.

For quickly viewing (not editing), you can use a style sheet to make it more readable (gives instructions to a browser on how to represent the XML file in a nice way)

From time to time, the CDISC XML Technology team publishes a define.xml style sheet for public use.

They can be found here:

Note that any functioning style sheet can be used - it is a matter of preference. The above is just a widely used example.

Define File: Codelists

Must the define file contain the mapping from source field(s) to Controlled Terminology field(s)?

It is not required for a submission to submit the mapping between raw source values and their mapped CT equivalents. Some mappings can include multiple columns, such as a single tissue (e.g., "adrenals") being mapped to multiple terms (e.g., SPEC="GLAND, ADRENAL"; LAT="BILATERAL").

What variable should be associated with a codelist in define.xml?

A variable should be associated with a codelist in define.xml if:

Terms from a published CDISC codelist have been used for the variable

The variable has some other list of allowable values associated with SENDIG (QORIG, for example)

The variable includes abbreviations or codes, so that the brief description of each code can be provided as the decode.

The variable contains a result with a scoring scale or a discrete range of values, so that the full context of an individual result in the dataset can be understood. Note that this may require that the codelist be defined at the value level for a subset of records for a test or category of tests.

The variable includes a discrete list of terms defined in the study protocol

Codelists are not generally needed for:

Variables containing continuous numeric data (not a finite set of allowable values)

Variables with values collected as free text (COVAL, for example)

Identifiers not from a predefined list and with no cross-animal meaning (Examples: SEQ, REFID, GRPID, COREF, XFN, IDVARVAL, TAETORD)

DTC variables

Other variables should be considered on a case-by-case basis and included if a codelist could provide value to the user of the SEND data.

What terms should be included in a codelist?

If all terms available for a study are required in order to correctly interpret one term used in the study, then all terms available for the study should be included in a codelist. This would be the case for scoring scales and severity scales. Another example is when "ALL TISSUES" is used in MASPEC or MISPEC; the associated codelist would need to include all protocol-required tissues to interpret what "ALL TISSUES" means. Otherwise, include only the terms actually used in the study.

Can a define.xml codelist contain all terms from the CDISC Published Controlled Terminology Codelist?

A define.xml codelist should only contain terms relevant to your study. If the CDISC Published controlled terminology codelist contains only the terms relevant to your study, then the full codelist can be used. MITESTCD and MITEST are examples of when this is possible. However, most CDISC Published Controlled Terminology Codelists contain many more terms than would apply to a single study.

How should sponsor-specific extensions to official CT be handled?

Best practice is to list only the items relevant to the study (not many CDISC codelists are

applicable in their entirety to a study). Use the "Alias Name" element for published terms (using the term's C code as the value) and the "ExtendedValue" flag set to "Yes" for extended terms used on the study.

Must the define file contain definitions for sponsor-specific controlled terminology (where no official CT exists)?

It is a good idea to list sponsor-specific terms in the define file. Specifically, if there are coded terms, such as 1=minor, 2=moderate, then the decoding should be provided. Refer to the information below about the use of the Decode value for a term.

In a define.xml CodeList, when should a Display Value (Decode) be included with a term?

Define.xml 2.0 codelists can include either EnumeratedItem or CodeListItem. Use of CodeListItem allows both a term and its display value (decode) to be included. All terms in a CodeList must use either EnumeratedItem or CodeListItem.

Use EnumeratedItem elements in a CodeList when the terms themselves are sufficient for data interpretation.

Use CodeListItem elements with a Decode in a CodeList when a decode facilitates data interpretation - when the code value is an abbreviation, acronym or short code that represents a word or phrase.

In a define.xml Codelist, what should be in the Display Value (Decode) entry for a term?

Decode for a CodeListItem element should contain the following: When the coded value has a definition (decode) in a paired variable in the data (whether or not the variables used CDISC Controlled Terminology), use the value of the paired variable in the decode. For example, the value of LBTESTCD has its decode in LBTEST; ETCD has its decode in ELEMENT; etc.

When the coded value does not have a decode in a paired variable in the data:

If the coded value is CDISC Controlled Terminology, use the value in the "NCI Preferred Term" column of the published Controlled Terminology version in the associated datasets (this column in the published CT file always has a single term and is always filled in).

If the coded value is NOT CDISC Controlled Terminology and the coded value is in an associated study report table, there is likely a key, footnote or other explanatory text in the study report tables that includes the code value. The decode should match the information explaining the value in the study report table. If this information is not in the study report, discuss this with the report author.

If the coded value is NOT CDISC Controlled Terminology and the coded value is NOT in an associated study report table (for example, if it is metadata collected but not reported, or if it is from SOP information associated with the study), the decode should contain a short unambiguous word or phrase that explains the coded value.

See SENDIG 3.1 Section 4.3.4 regarding use of coded result values in SEND datasets.

Note that some sponsors have received a comment from the FDA that the decode should not contain the full CDISC definition of a term.

When and how can you use one define.xml Codelist with multiple variables?

A good practice is the following:

In order for a codelist to be used by more than one variable, each codelist variable must have all the same terms. Additionally, the same term in each codelist must have the same meaning.

Note there is no requirement to share codelists across variables. But if you want to share codelists, follow the above methodology.

For example, if "g" is the only value available on the study for BWORRESU and BWSTRESU, you can use a single codelist containing "g" for those two variables in your define.xml file. On the other hand, in EX, the EXDOSU and EXVAMTU each have possible values, so they cannot share one codelist.

While this can be done across domains (e.g., if BWORRESU and FWORRESU only had "g" in the list), it is advised to keep the lists for different domains distinct, as they serve different contexts.

Define File: Value-level Metadata

What is value-level metadata?

Value-level metadata describes a subset of entries in a dataset variable.

When should value-level metadata be included in define.xml?

When all values in a variable in a dataset cannot be described using the same DataType, SignificantDigits (number of decimal places), Origin Type, Derivation Method, Codelist, or Comments, value-level metadata should be used to describe subsets of the entries in a dataset. Different lengths only are generally not sufficient reason to use value-level metadata (use the largest length at the variable level).

Which variable(s) should be described in value-level metadata?

For TS and TX, the subsets of the --VAL variable should be described.

For findings datasets with numeric results, scores, short-text entries, --ORRES and --STRESC entries contain the result, so those are the best candidates for value-level metadata. Since STRESC is generally used in analysis, that is likely the best choice.

For findings datasets with observational data, --STRESC would be described in value-level metadata. In addition, some modifier variables might also need value-level metadata; for example, if multiple severity scales are used for observations in CL, each would be described separately in value-level metadata.

Define File: Other Questions

Must the definition file contain the formulae for calculated measurements?

Any variable with an origin type=Derived must have a documented derivation method in define.xml.

What makes up a good dataset key?

It must describe a unique record in the dataset

It must contain only variables in the dataset.

It must not contain the --SEQ variable since a natural key should not contain a surrogate key.

Note: If it is truly not possible to create a natural key without --SEQ, such as the case in MA with clinical signs follow-up entries, include a dataset comment to indicate that --SEQ is used intentionally, and describe why it is required.

It should not include a variable if its entries do not contribute to uniqueness. Some examples are:

A variable that has no entry for any row in the dataset

A variable that has a one-to-one relationship with another variable in the key (--TESTCD and --TEST, for example)

A variable that has only one value for the subset of data described by other key variables (for example, if --CAT is in the key, and no category has more than one subcategory, do not include --SCAT)

What does the SENDIG 3.0 say about define.xml comments?

The SENDIG 3.0 has a number of statements about required or recommended contents for define.xml files for SEND, and the published SENDIG 3.0 Conformance Rules reflect these statements. Here are some statements that cover common cases, followed by their conformance rule number when available:

Section 4.1.3 - SDTM Core Variable: when no data exist for an expected variable, a null column should still be included in the dataset, and a comment should be included in the data definition file to state that data was not collected. (Rule 14)

Section 5.1.1 - DM domain, RFSTDTC CDISC Notes: The sponsor must define what collected date is used to populate RFSTDTC in the data definition file. (Rule 100)

Section 5.1.1 - DM domain, RFENDTC CDISC Notes: The sponsor must define what collected date is used to populate RFENDTC in the data definition file. (Rule 101)

Section 5.1.1.1 – DM Domain, assumption 4e: Sponsors should indicate how AGE was populated in the define file comments. (Rule 112)

Section 5.3.1.1 – SE Domain, assumption 3: If the start date/time of an Element was not collected directly, the method used to infer the Element start date/time should be explained in the Comments column of the data definition file. (Rule 129)

Section 6.1.1.1 – EX Domain, assumption 4: EXDOSE: The sponsor's data definition file should indicate whether the values in EXDOSE represent intended or actual dose levels. (Rule 137)

Section 6.3.6.2, LB domain, Example 5: The following example shows cases of categorical data that cannot be considered as numeric, even though in some cases, it appears that the data are a number. The allowed values in these ranges should be defined in the data definition file.

Some less common cases are also discussed:

Section 6.3.6, LB domain, LBNRIND CDISC Notes: Sponsors should specify in the study metadata (Comments column in the data definition file) whether LBNRIND refers to the original or standard reference ranges and results.

Section 6.3.6, LB domain, LBTOX, LBTOXGR CDISC Notes: The sponsor is expected to provide the name of the scale and version used to map the terms, utilising the data definition file external codelist attributes (Rule 173)

Section 6.3.12.1, PP Domain, Assumption 2: Information pertaining to all parameters (e.g., number of exponents, model weighting) should be submitted in either the SUPPPP dataset or the define file.

Section 6.3.16, EG Domain, EGDRVFL CDISC Notes: Used to indicate a derived record. The value should be Y or null. The derived flag should only be used where there is, within a dataset, a test code that contains both collected and derived data. In this circumstance, the data definition file will contain information that describes the reasons and approach. (Rule 209)

Some other, even less frequent cases are as follows:

Section 4.2.4 CASE USE OF TEXT IN SUBMITTED DATA - The sponsor's data definition file may indicate whether case sensitivity applies to text data for any or all variables in the dataset. (note: Removed for SENDIG 3.1)

Section 4.3.2 CONTROLLED TERMINOLOGY TEXT CASE - It is recommended that controlled terminology be submitted in upper-case text for all cases other than those described as exceptions below. When extending a controlled terminology list, the case-sensitivity convention of that list should be followed. Deviations to these rules should be described in the data definition file.

Section 4.4.5 REPRESENTING ADDITIONAL STUDY DAYS - The SDTM allows for DTC values to be represented as study days (-DY) relative to the RFSTDTC reference start date variable in the DM dataset, as described above in Section 4.4.4. The calculation of additional study days within

subdivisions of time in a study may be based on one or more defined reference dates not represented by RFSTDTC. In such cases, the sponsor may define Supplemental Qualifier variables to store these study days, and the data definition file should reflect the reference dates used to calculate such study days.

Section 4.5.3.1 TEST NAME (--TEST) GREATER THAN 40 CHARACTERS - To address this issue, sponsors may include the full description for these variables in one of two ways: • In the data definition file Origin column for --TEST, provide a link to the source containing the full test description. • Alternately, create a PDF document to store full-text descriptions. In the data definition file, the Comments column for --TEST insert a link to the full test description in the PDF.

DM Example 3 – Multiple Sites: The in-life part of the study in this example is conducted at two different sites: SITEID = LAB1 or LAB2, which would be further described in the data definition file.

Domain Specifics

Trial Summary

When is it ok to include a simplified vs a full TS?

For standard study types (e.g., single/repeat dose carcinogenicity, etc.), i.e., the eCTD sections as outlined in (which can be reached from the):

If you're submitting a report and SEND is required, a full TS is required

If you're submitting a report and SEND is not required, a simplified TS is required (see below for more info)

If you're not submitting a report (data-only), consult with or

How do I generate a simplified TS?

For nonclinical studies that were started before the SEND requirement, the FDA is asking for a simplified TS file. How can I learn more about what formats are acceptable?

This page contains some information and has a link to a Simplified TS creation guide:

In addition, several tests were supplied to the FDA in July 2021. The results of these tests are here:

The summary learning points are:

All tools used to create the ts.xpt created acceptable datasets

Full YYYY-MM-DD STSTDTC required for acceptance (Study C)

When full STSTDTC was present, neither the absence of TSVALNF nor what it was populated with affected acceptance (Studies A, D, D1)

When the study start date was after the date requiring a full SEND data package and only a simplified ts.xpt was provided, the dataset was rejected. (Study X)

Datasets without a STSTDTC were acceptable as long as only NA was supplied in TSVALNF (Studies F, O, S, T)

Incorrect variable labels did not affect acceptance (Study H)

Additional variables and parameters, as specified by SENDIG v3.0 and v3.1, did not affect acceptance (Studies G, L, S, T)

The Study ID in the Study Tagging File can be successfully matched by one of multiple SPREFIDs when the ts.xpt STUDYID does not match (Studies E, U)

Trial Design Domains

Is the Trial Design (TE, TA, TX, TS) meant to be planned or actual?

The Trial Design domains are meant to describe the planned design (i.e., that which the protocol and amendments prescribe). The actual progression of elements can reside in other domains, such as Subject Elements (SE) and Exposure (EX).

Is it acceptable to use "Last day of element" as the end rule in TE?

The start and end rules should not self-reference the trial design (i.e., referring to epochs, arms, elements, etc.); instead, they should be based on study concepts or events. If you desire a more generic end rule, then consider anchoring the end of dosing, such as "Last day of dosing with X", so that it is based on something tangible and readable.

Exposure

Do I have to submit each dose?

The SENDIG allows you to choose. You can submit one record for each dose, or you can submit one record for each period of consistent exposure for an animal (e.g., same lot, test material, route of administration, dose frequency, etc.). For example, if all animals received one dose a day with the same lot for the entire study, then you could submit 1 row per animal to describe the exposure details.

Note:

If you submit one record per animal for an entire period (versus submitting one record per animal per dose), it can complicate correlation with dose-related data (e.g., PC/PP), so it is usually recommended to submit one record per animal per dose.

For studies with special dosing periods, like inhalation/infusion, it is usually recommended to include the intended/planned dose (as in many cases, the calculation for the actual amount is

complicated and not simply represented), with a note in the define file that the EXDOSE represents the intended dose levels.

PK Domains (PC and PP)

Can PK data be put into the SEND format? Is there a specific template or type of file to upload for PK data?

The PC (concentrations) and PP (parameters) domains have been constructed to handle PK/TK data.

How do you populate PCTESTCD/PCTEST vs. PPCAT?

Below are some principles from the IG to keep in mind:

PCTESTCD and PCTEST are in a 1:1 relationship and define the analyte/specimen (in many cases, this is effectively the analyte code and name, respectively).

PCTESTCD is limited to 8 characters, can have no special characters, and cannot begin with a number. Often, preparers will populate this with a truncated version of the analyte's identifier if it is unlikely that two analytes with the same final digits will be included on one study, e.g, if compound ABC-1234567 is the test article, this is likely to also be one of the analytes and might be shortened to "ABC12345".

PCTEST is limited to 40 characters. This is often just the analyte's name or identifier. In some longer treatment names, a raw truncation would not be unique enough, such as cases where the treatment names are verbose. In these cases, the preparer might try a more educated abbreviation or an internal identifier for the analyte. If the analyte's identity isn't clearly understood from these fields, an explanation in the study data reviewer's guide may be necessary.

PPCAT must equal a PCTEST value - these being equal drives the natural link between PC and PP for all but the most complicated PK cases.

The above points tend toward the most common cases. The Technical Conformance Guide has more information on populating these domains.

How Do I Handle PK Data or Analysis with CROs/External Labs?

A number of considerations arise when initiating a conversation with another organisation around SEND file production. Page has some tips and an extensive "Points to Consider" question list to help smooth this process.

The following is a Template to Facilitate Creating Pharmacokinetic SEND Datasets:

SEND Future

SEND Roadmap

See this link for the Nonclinical Standardisation Roadmap group's roadmap, a general (but not binding) indication of the relative priority of domains:

Please note that this page is old and missing images.

Older document - Flag for review

Is ____ Included?

Is Safety Pharm included in SEND?

3.0: As a study type, no.

3.1: Partial. 3.1 introduces modelling for a subset of Safety Pharm endpoints (cardiovascular in CV, respiratory in RE); others are still in review (such as CNS: FOB and neuro/neurotox endpoints).

Is DART / Repro Tox included in SEND?

DART has its own guide published (SENDIG DART 1.1) for Embryo-fetal developmental (EFD) studies (see).

Note that at the time of writing, it will be required by CDER for EFD (and EFD supporting cases, e.g., nonpregnant female pilot study) starting in 2023 (but for the current answer, please consult the updated Data Standards Catalogue and Technical Conformance Guide).

Other study types beyond EFD are expected to be included in later releases.

Will SEND be developed to cover neurotoxicity studies?

See the "Is Safety Pharm included in SEND?" question above.

Will SEND be developed to cover phototox studies?

Some parts of many phototoxicity studies can be represented in SEND 3.0; however, they are not within the scope of studies for the current SENDIG version.

Will SEND be developed to cover genetic tox studies?

There is a subteam formed to model certain types of genetic toxicity studies; however, draft domains are not yet available. See the "Are other endpoints being developed?" question.

Is Anti-drug Antibody (ADA) expected?

If ADA data exist as part of a single-dose, repeat-dose, or carcinogenicity study, is it required to be included in a SEND submission?

ADA data are not considered officially in scope at present.

However, it is generally possible to model in LB, possibly supplemented with some additional SUPPQUAL variables. A PHUSE working group exists for how to model ADA data, should a sponsor desire to do so. Further information and information to participate if desired can be found at the following: Modelling Endpoints: How to Model Anti-Drug Antibody Data in Nonclinical Studies (BROKEN LINK)

It is important to note that a new SEND domain is planned to be developed to house these data in the future (i.e., LB is only a temporary proposed solution).

Is Flow Cytometry expected?

Flow Cytometry is out of scope. A team is working on modelling for it, but it is a way out.

How to model Cytokine and Immune Response Data?

This issue is an open issue across the industry and has been submitted to the SEND team for resolution. There is a lot of confusion in that there does exist a lot of overlap between immunology and currently in-scope study types, and, due to a lack of any distinct guides or domain implementation for it, many companies end up feeling it out study by study. There doesn't seem to have been much feedback or pushback on what people have ended up doing.

In the meantime, the CT exists for some of these endpoints for two cases:

...when these endpoints are legitimately collected as a part of in-scope study types

...when people are putting the data in LB due to the absence of a true domain implementation for cytokine, immune response, etc. In many cases, these are still lab tests and therefore not inappropriate to put in LB. However, there may be cases where additional variables would be needed to fully characterise the data.

If you are looking to model the data, PHUSE subteams have covered some recommendations for Biomarkers and ADA in white papers:

Biomarkers:

ADA: Modelling Endpoints: How to Model Anti-Drug Antibody Data in Nonclinical Studies ([BROKEN LINK](#))

Are other endpoints being developed?

For the time being, the SEND team is focused on Repro and Safety Pharm. Other endpoints, such as dermal/ocular, gene tox, and neuro, are under consideration for inclusion in SEND.

The page (note: last updated 2013, broken links) shows a prioritisation of endpoints captured from the industry in 2013 and compiled by the Nonclinical Standards Roadmap group and still gives a good idea of relative priority today.

As with any currently unmodeled endpoints, the FDA encourages the submission of data, using a SEND domain (if it makes sense) or creating a custom domain.

Will SEND be developed to cover veterinary records (treatment recommendations and observations from the exams)?

Vet records, among other endpoints, are on the roadmap; however, it is not determined when they will be modelled. See the team for more details.

SEND Future: Other Questions

When will the XPT files be replaced with XML?

The FDA is currently working on this, discussing and determining the pros and cons of some of the available solutions (HL7-based implementation, ODM, etc.).

SEND Issues

Known Issues

See the page.

Suggesting Changes

If you see an error or have a suggested change for improving the CDISC SEND standard, enter your comments using JIRA as described on the page. If you don't already have an account, select the option to sign up for an account.

Open Items (Pending Revisit)

The following are open questions, without an answer at present. Most involve the Change Control Tracker (CCT) for an answer from the larger SEND team or others.

Flag for review

My study data does not have all of the matching elements for SEND; what do I do?

Example: not all of the scheduling information that SEND wants (e.g., ELTM) is available.

Example: legacy processes combined scheduled and unscheduled clinical observations; how to distinguish which should have VISITDY populated?

If it was previously distinguished for reporting or other means through some form of algorithm, then you will need to replicate that logic in the population of the SEND variables. For the clinical observation example, if the detection of which observations were unscheduled was through text-matching on the clinical observation type "Unscheduled", then that same logic would need to feed the processes creating the SEND data.

However, if this information is simply not available or parseable from the data, then CCT will determine

What should be done when the only variable distinguishing records is the unit (e.g., --ORRESU)?

For some lab tests, such as protein, there can be multiple results for the same TEST and METHOD but with different calculations, resulting in multiple results with different units. CCT item submitted (#83).

Submitting New Issue(s)

If you have exhausted the above (including asking in the) and think there is an issue with the SENDIG, new issues may be submitted in the.

Steps:

Navigate to

Log in with your username and password (if you do not have an account, click on the "Sign up" link)

Once in, click the "Create" button along the top

For Project, select "SEND (SEND)" from the drop-down. Note: This is critical to select SEND. Using another project means your issue may be lost.

Fill out the fields as best as you can. In particular, with the description field, please note at a minimum:

The SENDIG version you are referencing that has the issue

Section and page (and domain name if relevant)

Quote the relevant text

Provide some description of what is wrong with the text

Provide some suggestions on what solution you think would remediate the issue

Example of a good description:

"In SENDIG 3.1, section 6.3.1.1, page 65, BW domain, Assumption #2 states "Body weight gains are submitted in the Body Weight Gain domain", but this is not descriptive enough. Please change to "Body weight change (gains and losses) is submitted in the Body Weight Gain (BG) domain."