

SEND Implementation Wiki - SEND Fundamentals



What is SEND?

SEND, or the **S**tandard for **E**xchange of **N**onclinical **D**ata, is an implementation of the CDISC Standard Data Tabulation Model (SDTM) which specifies a way to present nonclinical data in a consistent format.

By having a common model to which the industry can conform, many benefits can be realized, including more efficient inter- and intra-company exchange of nonclinical data, and the ability for vendors to develop tools that can use this data.

A SEND package consists of several components, but the main focus is on individual study endpoint data. Endpoints typically map to domains (essentially, datasets), with a number of variables (a.k.a., columns or fields).

Where is the Guide?

The SEND Implementation Guide (SENDIG) is a fairly exhaustive resource for implementers, covering both the specifications for modeling data and providing pretext around modeling concepts. It is located here:

- <https://www.cdisc.org/standards/foundational/send>

This page is a general resource page for SEND; look for "SENDIG" to get the most recent version. This page also contains links to controlled terminology (CT), define.xml specs, and so on.

Reading the Guide

When reading the SEND Implementation Guide, there are a few important concepts to understand from the outset.

Each SEND domain has a table that contains a list of the generally-used variables in the domain, and a description of each variable. Some variables are shared between domains (for example: USUBJID, VISITDY) and other variables are named in a way that indicates what domain it belongs to (For example: LBTEST belongs to the LB domain, BWORRES belongs to the BW domain). Those domain-specific variables are also shared across domains: variables that are the the same except for the domain designation represent the same concept. For example BWTEST, BGTEST, OMTTEST, LBTEST all represent the concept of a test name, and the first two letters represent the domain the variable is in. When reading the guide, these shared variables are often represented with "--" replacing the domain abbreviation: --TEST indicates the test name variable in any domain. Therefore, while it may seem like there are a very large number of SEND variables, there is actually a much smaller number of concepts that are represented by those variables, and understanding the variables from one domain will help you understand the variables in another domain of the same type.

It is also best to read the introductory material in Section 1 through 4 of the guide in conjunction with the Domain Model descriptions in Section 5 on, so that the domain descriptions and examples can be used to understand the detail on the concepts described in the beginning of the guide.

Model Basics

The following is a summary of some of the key basics of the model. All of the information below is available in more detail in the SENDIG.

Key SDTM Concepts

Domain

- Specifies the structure for a dataset
- Usually endpoint-centric
- Represented by a two-letter code

A domain is a collection of variables (a.k.a., columns or fields) which describes how to layout a given type of data (usually centering around endpoints) and which, when populated for a study, manifests in a dataset. For instance, the BW (Body Weights) domain model tells you a number of fields to include when modeling body weight data, such the animal, date, result, etc. to describe each body weight record.

Variable

- A field used in a domain
- Corresponds to an SDTM column
- Usually prefixed with the domain code

A variable is a field or column used in part to describe a record for a domain. For instance, the Body Weights domain's BWDTC variable is used to describe the datetime (DTC) at which a body weight was collected. Typically, variables are prefixed with the domain 2-letter code, then followed by the SDTM column name.

Permissibility / Optionality (aka Core)

The permissibility of a variable is expressed through the Core column of a domain and has the following values:

- *Req* (Required) - States that not only is the variable required but also must have a value for each record (no blanks)
- *Exp* (Expected) - States that the variable must be included but may include blanks if the variable is not applicable to the entry or there is no data available for that variable for the entry
- *Perm* (Permissible) - States that the variable should be included (if you have data for it) but may be left out of the dataset if there are no data to populate the variable for any entry in the dataset.

The Anatomy of a SEND Package

A SEND package includes a number of domain datasets (usually endpoints) and a define file (which describes the datasets provided), along with its companion transformation file.

The SENDIG notes that the following domains generally form the foundation of each SEND package: TS, TX, DM, and EX, along with at least 1 endpoint-related domain dataset (such as BW for body weights); however, the list of domains will be specific to each package and what data were collected. Below is an **example** set of files for a simple submission with body weights and clinical observations:

- **define.xml** - to describe what is in the submission, such as the columns, comments about the columns, data types, Controlled Terminology used, and so on. Refer to [SEND Implementation - Define Fundamentals](#) for more information.
- **define-1-0-0.xsl** - a static file which allows a visual presentation of the information in the define.xml file when opened with a browser
- **ts.xpt** - Trial Summary dataset, which provides high-level details about the study
- **ta.xpt** - Trial Arms dataset, which provides a specification of the sequence of planned treatment-related events in the study
- **te.xpt** - Trial Elements dataset, which provides a listing of the treatment-related events in the study used in Trial Arms
- **tx.xpt** - Trial Sets dataset, which provides group information
- **dm.xpt** - Demographics dataset, which provides a listing of the animals on study
- **ex.xpt** - Exposure dataset, which provides dosing information
- **bw.xpt** - Body Weights dataset, which provides body weight results collected on study
- **cl.xpt** - Clinical Observations dataset, which provides clinical observation results collected on study

Note that the files' names are lowercase in anticipation that they would be submitted in accordance with ICH guidelines.

Controlled Terminology

Controlled Terminology (CT) is the standardization of terms to be used in SEND packages. CT ensures that when referring to the same base concept, everyone calls it the same thing. CT may be specified for a variable in a domain to indicate that there are specific terms that may be used for that variable. CT can be extensible (terms outside the list can be used if not present already) or not (only terms from the list allowed).

Please see the [SEND Implementation - CT Fundamentals](#) page for more information on CT, including resources and mapping considerations.

Define File (e.g., define.xml)

The define file (e.g., define.xml) is a companion file included with each SEND package which describes the SEND package, including such information as which datasets are present, which columns are used, which CT is used, types of the variables, comments on the variables, and so on. In other words, it provides the metadata for the package.

Please see the [SEND Implementation - Define Fundamentals](#) page for more information on the define file, including resources and tips.

XPT Files

XPT files (a.k.a., SAS v5 Transport format) are a SAS file format used for output of SEND datasets. When viewed in the free viewer below, they appear similarly to Excel workbooks:

- [Universal Xpt File Viewer \(Open Source\)](#)

Note that this was previously the SAS Viewer; if you already have the SAS Viewer tool, it works the same way and can be used to open the XPT files.

Results in Standardized Units

There are a few variables for expressing the original and standardized results:

- --ORRES reflects your original result as collected, *in whatever units under which it was collected*
- --STRES/-STRESN reflect your result *in the units of your choosing* (e.g., if units are standardized for reporting)

Likewise, there are variables for expressing the units of those results:

- --ORRESU reflects the units of your original result, with their Controlled Terminology preferred label
- --STRESU reflects the units of your choosing for the reported/presented result (e.g., if units are standardized for reporting), with their Controlled Terminology preferred label

It is important to point out the difference between conversion (changing units) and CT mapping (relabeling units). **There are no stipulations that you convert to particular units**, only that you use particular labels for the units that you used. The --ORRESU and --STRESU variables only represent a mapping or relabeling of the unit used for the result and in no way dictate any conversion required. For instance, if you have an original result of 50 mg/mL, then your --ORRES remains 50, while the --ORRESU gets relabeled to g/L (the *scientifically equivalent* CT-preferred label for mg/mL). Along those lines, any conversion applied to the standardized result should not be on account of the CT; it should be because you chose to report and present a different unit from what was originally collected.

These concepts are also discussed in the SENDIG in section 4.5.1.1.

Trial Design

Trial Design is a collection of domains which describe the planned study design.

Trial Summary

At the highest level of the trial design, the Trial Summary (TS) domain contains study-level details, such as the study title, experimental start/end dates, the study director, route of administration, and so on. Typically, these parameters are 1 value for the entire study.

Trial Elements

Trial Elements are planned portions of a study that have the same duration and same treatment (or lack of treatment) for all animals. For example, the predose period may be an element, as it has a specified duration during which animals will not be treated. A dosing period with a specified dose level is also an element. Elements are building blocks of Trial Arms.

Trial Arms and Trial Elements

The Trial Arms domain allows specification of the planned sequence of treatment (or lack of treatment) events in the study. Arms comprise a number of Trial Elements in a defined sequence.

In simple cases, Trial Arms for a study each have 2 elements: a screening element (predose, before animals are randomized) and a dosing element. For instance, the control arm would have screening and treatment with 0 mg/kg via the control, while a dosing arm would have 2 screening and X mg/kg.

As designs become more complicated, such as latin squares or reduction of dose due to toxic effects, so too do the Trial Arms, mirroring the sequence of treatment changes. The SENDIG has a number of examples of these cases.

It is important to note that specification of the study design is *planned* - which includes changes to treatment resulting from protocol amendments. For instance, if a dose is planned to be reduced for the high dose group due to toxic effects and is changed by an amendment, then the trial arms should reflect this (e.g., the arm would consist of a screen element, the high dose element, and now the reduced dose element).

In addition, note that Trial Arms and Trial Elements are defined only by treatment (or lack thereof), duration and sequence. Additional factors which may influence the grouping of animals into study groups, such as TK status, housing status, and so on, are not represented by arms/elements. Because of this, multiple groups may be assigned to the same arm, yet be considered separate groups for the study and its analyses. Factors of this nature are represented in Trial Sets.

Trial Sets

The Trial Sets domain allows specification of the unique sets of animals on study that are considered separately from other sets by various factors. Dose level is generally automatically one of these factors, but housing type, TK status, and so on, can also contribute toward making different sets.

Because Trial Sets include all of the various factors that make sets of animals stand out from other sets of animals, they are usually at the granularity of groups but can sometimes be a level or two down. For instance, if a sponsor has TK animals considered to be in the same study group as the main study animals, then these may have the same group and group label, but these would be two separate sets. For examples, see the SENDIG.

The Relationship Between Study Groups, Trial Arms and Trial Sets

Traditional preclinical studies are organized into dose groups. Due to the differences in how dose groups are used defined by different organizations and in various data collection systems, it is not possible to say definitively that group=arm or group=set. If results for all animals in a group are analyzed together and are not combined with another group for analysis, chances are your group=arm and set. If you divide the animals in a group for analysis (perhaps main animals and TK animals), then your group is likely an arm and each subset of animals is a set. If you combine groups that have the same dosing regimen for analysis, chances are your groups are sets, and the combined collection of animals is an arm.

It is important to understand the concepts behind Trial Arms and Sets when mapping your study dose groups to these concepts.

Dosing (Exposure Domain)

The Exposure domain holds the specification of treatment exposure for animals.

There are 2 methods for populating this domain - 1 record per contiguous exposure or 1 record for each dose - with the method up to the sponsor.

1 Record Per Contiguous Exposure

Under this method, 1 record is included for each span of time where the dosing parameters were constant. For instance, if an animal received once daily dosing with the same lot for the entire study, then you could include 1 record for the animal, comprising all dosing between the start at day 1 and end of dosing.

If any parameters change, such as the dose level or lot, then create a new record for each change, specifying the start and end of that change. This includes cases where the exposure parameters return to their original values after a period of something else, such as if a high dose group receives a reduced dose for a short period of time before returning to the high dose level. In this case, there are still 3 contiguous periods (high, reduced, then high again).

1 Record Per Dose

Under this method, 1 record is included for each individual dose. If dose documentation for individual animal doses has been collected electronically, this may be the easiest way to report this data.

PK Domains (PC and PP Domains)

Pharmacokinetics (PK) data, sometimes also referred to as TK, are represented in SEND through the Pharmacokinetic Concentrations (PC) and Pharmacokinetic Parameters (PP) domains.

The PC domain covers concentrations. PC data are usually collected at a few set days in the study, and within each day, at multiple time points. Each concentration observed is included in this domain.

The PP domain covers the analysis parameter results derived from modeling a curve to the concentrations data. Aspects of the curve, such as Area under the Curve (AUC) or time to max concentration (t_{max}) are modeled in SEND as parameters in this domain, and are usually tied to the days where the concentrations were measured.

PC and PP are linked through the --RFTDTC variable, which is usually the dose. This provides an implicit link between the concentrations data and their corresponding curve results. The SENDIG provides some more complex examples.

Comments (CO)

The Comments domain provides a place for comments tied to records in other domains (most commonly, findings domains).

The most common form of comment is to a specific record in another domain. In this case the IDVAR is typically set to the associated domain's --SEQ variable. For instance, a CO record with RDOMAIN="BW", USUBJID="ABC123_101", IDVAR="BWSEQ", and IDVARVAL="163" would tell us that this comment relates to record in the BW domain for USUBJID ABC123_101 where BWSEQ is 163.

Within a domain, comments may be tied to other variables as well. For instance, a comment tied to a value in the MI domain's MISPEC column (aka tissue) could relate a tissue comment.

Comments may also be tied to the domain itself (general domain comment), or the study at a whole, also these are less typical. Examples of these cases are shown in the SENDIG.

Record Relationships (RELREC)

Relationships between data are specified through the RELREC special purpose dataset, in which related entities are listed as rows, with their associated domains, animals, and IDs

RELREC Basics

RELREC is centered around providing a place to put objective, known relationships between information (e.g., relationships collected in the system, or implicit relationships like PP records' link to the PC records used to create them). These RELREC relationships are most commonly defined within a subject (or pool), linking specific records from one domain to other records for that same subject in another domain. Each relationship is given a RELID (Relationship ID) which identifies the relationship (e.g., all of the rows with RELID=1 are related to one another; all of the rows with RELID=2 are related to each other; etc.).

Within a relationship, contributing records are identified within a domain by way of the IDVAR (Identifying Variable) column, which states which variable in a domain we want to use to relate. For instance, if the IDVAR=BWSEQ, then we know that this relationship contains some related body weight records, which we call out by way of their BWSEQ values. In this way, each contributor to the relationship is listed, until all of the related records are represented in the RELREC.

Note that RELREC is not used for subjective, *post hoc* relationships used more toward analysis or interpretation (such as a Study Director later deciding during the writing of his/her results that some clinical observations might be related to some clinical pathology findings); these types of relationships are not in the scope of SEND and are more a function of analysis modeling (e.g., ADaM).

Common RELREC Population

In its most common incarnation, a RELREC relationship links --SEQ variable values from one domain to another (i.e., the IDVAR will be --SEQ). In this case, there is simply 1 row for each related record, with the related --SEQ value. The example below shows 2 relationships of record-to-record cases. The first two rows establish one relationship (RELID=1), linking CLSEQ=172 from the CL domain to MASEQ=98 from the MA domain for USUBJID ABC-123_101. The last 3 rows establish another relationship (RELID=2), this time for USUBJID ABC-123_102, linking one result in CL (CLSEQ=180) to two records from MA (MASEQ values 103 and 105).

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
ABC-123	CL	ABC-123_101	CLSEQ	172		1
ABC-123	MA	ABC-123_101	MASEQ	98		1
ABC-123	CL	ABC-123_102	CLSEQ	180		2
ABC-123	MA	ABC-123_102	MASEQ	103		2
ABC-123	MA	ABC-123_102	MASEQ	105		2

In other examples, the relationship might not be at the record-to-record level, but rather, from a group of records to another group of records (e.g., using --GRPID). Such is the case with --SPID (the specimen identifier, e.g., mass ID), where the relationship as used between CL, MA, MI, and TF defines how given values of --SPID (which could represent a number of records in the respective domains) correlate between domains. A --SPID case is seen below, where a relationship is defined linking all records from CL with CLSPID="MASS 1" to all records from MA with MASPID="MASS 1":

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
ABC-123	CL	ABC-123_103	CLSPID	MASS 1		3
ABC-123	MA	ABC-123_103	MASPID	MASS 1		3

RELREC Use Cases

The following tables talk to cases of RELREC. Note that each case is defined by the domains related.

Expected and/or Useful RELREC Case

Domains Represented	Case	Needed When	Typical IDVARs	Notes
PC, PP	Linking pharmacokinetics results with the specific concentrations on which they are based	Optional at this time, per the SENDIG, but a good idea when the implicit USUBJID/POOLID link (i.e., all PC records for an animal/pool linking to all PP records for the animal/pool) does not accurately cover the relationship (e.g., when some of the PC should not be linked).	SEQ and/or GRPID (see SENDIG)	Relationships can be made in a number of ways, as specified in the SENDIG.
CL, MA	Linking clinical observations to macroscopic observations	When these relationships are collected for studies with a pathology component	SPID or SEQ	Common
PM, MA	Linking palpable mass observations to macroscopic observations	When these relationships are collected for studies with a pathology component	SPID or SEQ	Sometimes

MA, MI	Linking macroscopic observations to microscopic observations	When these relationships are collected for studies with a pathology component	SPID or SEQ	Common
CL, MA, MI	Linking clinical observations to macroscopic observations and corresponding microscopic observations	When these relationships are collected for studies with a pathology component (and the relationship is defined for all of the domains at once)	SPID or SEQ	Less common - this is usually seen instead as two separate linking activities – the correlation of clinical observations to macroscopic findings is one collection, and the later correlation of macroscopic findings to microscopic findings is another.

Possible but Not Common or Expected RELREC Cases

The following cases are possible RELREC cases, with some information around why they are not generally needed.

Domains Represented	Case	Needed When	Typical IDVARs	Notes
MI, TF	Linking microscopic observations to tumor findings	Not generally needed, since TF is generally derived from MI	SPID or SEQ	This relationship is generally implicit.
TF, CL	Linking a tumor to its “onset” clinical observation	Only when this relationship is collected	SPID or SEQ	Typically not done. This case is only appropriate if the onset designation can be considered a collected event. Cases where the onset is merely a factor of earlier correlation RELRECs and/or is algorithmically based likely would not apply.
TF, PM	Linking a tumor to its “onset” palpable mass observation	Only when this relationship is collected	SPID or SEQ	Typically not done. This case is only appropriate if the onset designation can be considered a collected event. Cases where the onset is merely a factor of earlier correlation RELRECs and/or is algorithmically based likely would not apply.
LB, MA, MI	<i>needs filling in. interpretive analysis?</i>	Only when this relationship is collected		
DS, DD	Linking disposition to its related diagnosis	Only when (1) there are multiple diagnoses for an animal and (2) this relationship is collected	SEQ	Generally implicit, in that typically, only one DD record exists per subject

RELREC Considerations

- Note that RELREC is intended to represent objective, known relationships, not arbitrary or interpretive relationships assigned *post hoc* at the time of the report preparation or summarization. For example, if a clinical observation is explicitly established during collection as correlated to a mass in MA, that is fair game. Likewise, the implicitly known, a priori relationship between the concentrations in PC and the PP results created from them also belongs in RELREC. However, it would not be appropriate to add an after-the-fact, interpretive RELREC relationship for the case where someone report-side decides to link a body weight with a clinical pathology test result to point out a link of interest.
- For further examples, see the SENDIG provides specific examples of RELREC usage.