

SEND Implementation Wiki - Getting SEND-ready

Table of Contents

<i>Training.....</i>	<i>2</i>
<i>Vendor Evaluations</i>	<i>2</i>
Vendor Solutions	2
<i>Deciding on a Vendor.....</i>	<i>2</i>
<i>Sample Datasets</i>	<i>3</i>
<i>Working with CROs.....</i>	<i>3</i>
Initiating the Relationship	3
Sponsor Responsibilities	4
<i>Internal Implementation Considerations</i>	<i>4</i>
Forming Your Implementation Team	4
<i>Mapping Exercises.....</i>	<i>5</i>
Domain Mapping	5
Controlled Terminology (CT) Mapping	6
Maintenance and Support Planning	7
<i>Validators.....</i>	<i>7</i>
<i>Dealing With Validator Errors and Warnings.....</i>	<i>8</i>
<i>Validation</i>	<i>8</i>

This page is a high-level primer on getting your organisation ready for SEND, with such information as where to start, things to consider, and so on, when beginning your SEND implementation and planning.

Training

Training is available from a number of sources:

- CDISC offers training on SEND and other topics; for CDISC-approved training, please see <https://www.cdisc.org/education/course/send-implementation> or contact training@cdisc.org to discover the training options that would most benefit you and your company
- CDISC-registered solution providers often provide high-level and/or special SEND training as part of their SEND offerings; if you need an introductory crash course or are focusing on aspects that are not covered by CDISC's courses, inquire with your vendor about any SEND training they might offer. See <https://www.cdisc.org/resources/rsp>
- Joining the SEND and PHUSE teams is one of the best ways to familiarise with SEND and keep abreast of developments (not to mention contributing to the greater cause). To join (if you're not sure which, join both!):
- For SEND (the data standard and its development), please contact Lou Ann Kramer (lkramer@cdisc.org) to join.
- For PHUSE (surrounding efforts for usability and adoption), please contact Sue DeHaven (Susan.DeHaven@sanofi.com) to join.

Vendor Evaluations

Vendor Solutions

The following vendors offer SEND solutions:

- [Instem](#)
- [Integrated Nonclinical Development Solutions, Inc.](#)
- [PDS](#)
- [PointCross Life Sciences](#)
- [Xybion](#)
- [Zifo RnD Solutions](#)

Deciding on a Vendor

The vendors offer a number of features in their solutions to make it easy to handle the numerous aspects of data management, absorption, and production, and it is a good idea to evaluate multiple tools to see what's available, what best fits into your systems, as well as what kind of new functionality they provide (such as review/analysis).

When on the hunt for a vendor, the following are useful to consider:

What functionality do you need

The larger solutions are built around the need for SEND production or utilisation, with features such as:

- The ability to read outside data in from multiple sources (or SEND files)
- The ability to define and maintain mappings from company-specific lexicons to SEND CT
- The ability to output complete or partial SEND packages, including datasets and define file
- Automation of such activities

Consider which of these are important to you and ensure that your needs are met when evaluating products.

Note: If you only need a way to open the files to view (for example, a Sponsor who contracts all SEND production out), then you may consider using the [Universal Xpt File Viewer](#), which is a free tool which can open the XPT files. *Note that it is extremely feature-limited.* This tool was previously known as SAS Viewer, so if you already have the SAS Viewer, you can use that as well to open the files. The "How do I open XPT files?" on the [FAQ](#) page lists other methods as well, including using base SAS and R to open the files.

What systems do you have

The maker of your collection system likely has a solution to produce SEND files from the data (and usually adapters to work with other data). If your data collection is more or less centralised on a single vendor, then their product may provide some good synergies with your existing data. Most of the providers listed, though, have adapters to work with others' data, so often, the decision is more heavily weighted toward functionality and cost.

Sample Datasets

Accessing sample data sets is useful as a reference while learning the CDISC standard, for exercising tools or processes that use SEND datasets, or for verifying your own study datasets. With final guidance from the FDA on the horizon, seeing a complete SEND data set will help you be prepared.

See the "Are there publicly available sample SEND datasets?" question on the [FAQ](#) for a list of places where you can obtain sample datasets.

Working with CROs

Initiating the Relationship

When working with CROs, there are several details which should be understood and agreed to in advance, from logistical considerations (like adding to the master agreement) to

specific content capability questions (like method of creating the Exposure domain). Many of these considerations are advantageous to determine early on in the process.

The [SEND between Organisations](#) page has questionnaires designed to help in this process of setting up and maintaining a partnership.

Sponsor Responsibilities

Next, even with a full-service CRO, the Sponsor is ultimately responsible for the submission of SEND datasets. As such, there are some key responsibilities of the Sponsor with regard to the specification and handling of the SEND datasets that cannot be performed by the CRO or other external body. At a high level, they are:

- **Specify** - the Sponsor must be ready to define expectations for the datasets' creation (e.g., special requirements)
- **Receive** - the Sponsor must be ready to physically receive data from the CRO or provider and potentially incorporate it into its systems
- **Review** - the Sponsor must be ready to review the data to satisfaction
- **Submit** - the Sponsor must be ready to submit the data, as part of a larger package (e.g., in eCTD structure)
- **Archive** - the Sponsor must be ready to archive past data according to relevant archival policies and procedures

Internal Implementation Considerations

Forming Your Implementation Team

Your implementation team will need data-savvy individuals from various parts of the organisation. The following are suggested roles that can help make an implementation a success. Single individuals may fill multiple roles (especially the case with smaller organisations), and these roles may also be filled by outside help.

- **Data specialist(s)** - data scientist(s), IT programmer/architect(s), and/or statistician(s) with expert knowledge of the various sources of nonclinical data present in your organisation. Someone who is already familiar with data transfer to the FDA (such as with the "99 guidance" datasets) will be of special help. Larger organisations that can do so should consider including a few of these individuals, spanning multiple areas (IT, stats, reports, etc.)
- **Submissions professional** - a regulatory submissions expert who is familiar with preparing submissions to eCTD specification and other such formats
- **Validation / QA professional** - an individual who understands systems validation and QA and can be on board early to understand and inform the impact on validation and QA policy for these data
- **Study Director** - a study director or related professional who can understand and inform the impact on protocol definitions and study activities
- **Subject Area SMEs**: certain subject areas will warrant special knowledge in order to map properly:

- *Pathology SME* - an individual with an expert understanding of pathology glossaries, e.g., a pathologist or advanced path support personnel, to help with the definition of mappings to SEND terminology.
- *Clinical Pathology SME* - an individual with an expert understanding of clinical pathology glossaries, e.g., a clinical pathologist or advanced clinpath support personnel, to help with the definition of mappings to SEND terminology.
- *PK SME* - an individual with an expert understanding of PK/TK data, to help with the proper creation of the related PC and PP domains.
- **Management** - executive sponsorship of the project to ensure it gets proper funding, prioritisation, and the devoting of resources

Mapping Exercises

Domain Mapping

The following are **recommended** high-level steps for mapping to SEND. Regardless of whether you are planning for a vendor or homegrown solution, a mapping assessment is most likely necessary to get a handle on which systems are involved, changes that might need to be effected, terminology mapping, and so on.

1. First, at the domain level, without thinking about the variables, consider the major systems that would source the data. This exercise will give you a non-definitive list of interfacing systems, which will help guide which attendees and SMEs can help with future steps. For instance, if your path data exists in a different system from your in-life collection system, then this can potentially identify different sets of SMEs (both from the user and IT side of things) who can work through the more detailed domain mappings. When making this assessment, consider not only current systems, but also systems that may have been used in the past and may contain data needed. Remember that one domain may be sourced from multiple systems.
2. Next, review each domain *with its variables* and map out, at a high level, where that data would be sourced - e.g., internal systems, manual entry, and so on. This exercise will give you a full list of systems to consider for integration or interfacing, and, more importantly, can highlight problem areas or holes, along with opportunities to change entry/processes in your source systems to shore up those gaps. It will also be helpful at this stage to mark which variables have Controlled Terminology (this will be useful later)
3. Next, consider, also at a high level, how data will get from point A (source systems or entry) to point B (SEND datasets) and whether intermediate steps will be needed, such as an intermediate warehouse. This could be via a vendor-supplied data adapter or ETL, a user interface to allow supplementary entry, and so on.

Once these steps are done, you should have a high-level understanding of:

1. The systems involved
2. What gaps are present
3. How the data is going to flow into the datasets from the source system(s)
4. How data not available in the source system(s) will be added to the datasets

Some example deliverables which can assist and shape the process could be:

- An interfacing systems list and/or flow diagram
- A catalogue of all domain variables with detailed fields such as Source System, CT Applicable, Whether to Include (for permissible fields) and others. This table can start as a high-level survey and evolve over time with such details as Transformations, Process Flow, and whatever other fields you might find helpful.

Controlled Terminology (CT) Mapping

Another critical piece to the mapping exercise is understanding and working through the mapping of your internal lexicons or terms to Controlled Terminology. For CT basics, see the [CT Fundamentals](#) page for more information on what CT is and some general considerations.

The following gives some high-level steps which may assist in taking control of your mapping.

1. First, identify the various variables which will need mapping. If you did this as part of your domain/variable survey above, great; if not, then look through the domain variables for CT. As a rule of thumb, if it has a CT associated with it, it's probably a variable you'll need to populate.
2. Next, for each case, identify the complete list of values that will need to be mapped. For example, this could be an internal dictionary or sometimes a distinct list of collected values over time. This may also identify cases where internal lexicons can be updated. This is a critical step for both vendor and in-house solutions.
3. The next step is the most painstaking, which is to sit down with SMEs and determine how each term from the relevant lists will be mapped. The [CT Fundamentals](#) page gives some guidance on how to figure out the right CT term to which your term should map. In most cases, it is a 1 1 mapping, i.e., given a single input value, you find the resulting mapped term. This can get more complicated, however, such as when multiple fields contribute to the mapping, or the mapping requires more processing, e.g., parsing or pattern matching.

Some key CT cases represent a lion's share of the mapping effort, being highly labour-intensive and/or requiring multiple subject matter experts. Here are some key CT cases to focus on, representing a large amount of effort:

- **Pathology:**
 - Tissues and their lateralities - Mapping to the CT for these fields (--SPEC, --LAT, --DIR, --PORTOT) can be very intensive, involving pathologists and other subject matter experts.
 - Organ weights - have a similar list of CT fields, also with an intensive mapping process.

- Findings text - Mapping to the CT for the findings themselves (--STRESC), in particular for tumours, which have controlled terms, is also a laborious process.
- **Clinical Pathology:**
 - Endpoints/tests - Endpoints are represented in SEND by tests (LBTEST and LBTESTCD) and the sample type that the tests are performed on (LBSPEC), which have controlled terms, as well as the test method (LBMETHOD). Mapping to these official tests is a large effort due to the sheer number of possible clinical pathology endpoints.
 - Units - On a related note, mapping the units for these endpoints (which have a wide range of units) is also a large effort.

With these exercises complete, you should have:

- A survey of the variables which need CT
- A mapping between source column(s) and mapped CT values

This information will then be a vital feed into either the vendor product or your own internal design.

Maintenance and Support Planning

Once implemented, maintaining SEND readiness requires a certain amount of upkeep. Consider the following common post-implementation support needs:

- **Controlled Terminology:** Controlled Terminology gets quarterly updates. Monitoring CT updates and ensuring that internal glossary mappings are up to date is key and requires ongoing effort.
- **Dataset Validator Issue Review:** data can fail the validators due to a number of reasons, such as new, strange data cases. Validator logs and data should be reviewed to ensure compliance.
- **New SENDIG releases:** periodically, the SENDIG receives updates, which necessitate some updates to your systems.

Validators

Validators perform a functional check of the datasets, looking for structural integrity issues, such as missing values in Required fields, missing columns, incorrect CT, and so on. Different tools check for different issues, and most should not be considered authoritative. That said, they are an extremely useful step in the process of creating datasets, checking for the integrity of a SEND package before it is sent out.

Validators are a separate concept from GLP system validation. The use of a validator does not preclude the need for system validation or QC on the systems or processes which produce the datasets. The validator's purpose is more of a nuts-and-bolts check, intended to

catch structural deviations from the specifications, such as missing values, as opposed to logical issues, such as incorrect calculations, which a system validation would cover.

There is no required or preferred validator. Early on, some validation rules were developed jointly by the FDA and industry. A variant of these rules is currently used by the free, open-source [OpenCDISC 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.

In September 2013, the FDA released a validation rule set that is used when they receive SEND datasets in a submission. This rule set is available from <http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm> in the section on Study Data Validation Rules.

Dealing With Validator Errors and Warnings

Errors are generally problems with your datasets that may need to be fixed. Check the implementation guide to confirm whether the reported error is an issue.

For example, you may get an error saying 'SE0056, SEND Required variable not found, Variables described in SEND as Required must be included', then you must check the dataset and implementation guide, make sure the Required variable is present in the dataset and can not be null.

Warnings can be problems with your datasets, but potentially not, and need to be reviewed. Each should be examined to ensure that there is a good reason why the scenario prompting the warning is ok.

For example, you may get a warning in the LB domain: 'Rule ID SD0029- Standard Units (--STRESU) should not be NULL, when Character Result/Finding in Std Units (--STRESC) is provided'. Since some Clinical Pathology parameters may have no units (Specific gravity, A/G ratio, pH, etc.), this warning is innocuous and can be ignored.

The FDA has a test submission process, where you can test the submission of your data through their gateway to find out if there are any critical issues with your SEND packages. FDA encourages sponsors to use this test gateway for at least the first few submissions to mitigate the impact of issues and not cause delays. Details on the test submission process are included in the following link: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm174459.htm>

Validation

From a regulatory systems validation standpoint, there isn't anything necessarily special about SEND datasets. What can help team members understand the validation needs is to think of the datasets as you would individual tabulations - yet another output of the system you are validating.

That said, it will probably be useful to get your systems validation group on board to understand what SEND means and how it will play into the system changes in which it is implemented.

This page goes into some more discussion on how to treat SEND with regard to QA/QC, the thinking and considerations of which are similar: [Handling of SEND in Study Documentation](#).