

# Selecting a CRO for Creating and Integrating SEND Datasets from Multiple Organizations

## Interorganizational SEND

### Introduction

Creating and integrating SEND datasets from multiple CROs and LIMS can be a complex process requiring good partnerships and upfront understandings between Sponsors and CROs. The Interorganizational SEND (I-SEND) Project Team within the FDA/PhUSE Nonclinical Working Group has developed a Points to Consider list to help guide and establish CRO capabilities, logistics, and Sponsor expectations. The list is organized into multiple categories, including process, compliance, domain content, RELREC, define, data transfer, cost, and study-specific questions. Some highlights include: 1) SEND and CT versions; 2) identification of domains and variables supported; 3) file format for data extracts and file exchange; 4) process for combining files from different sources; 5) need for manual data entry; 6) number of different LIMS and whether customization is required; 7) role of QA; 8) validation. Our poster presents this list in detail along with key examples where establishing communication upfront made a significant difference.

### Types of Partnerships

The following are some typical cases where multiple organizations might collaborate to create a SEND package. Regardless of the partnership type, for individual studies, the *Study-level Points to Consider* should be utilized.

**Sponsor and CRO - CRO creates full package**  
Production and packaging work is done exclusively by the CRO; the sponsor may provide inputs. For this case, the Sponsor would first engage the CRO using the Partnership-level Points to Consider (Initial Survey), followed by the Partnership-level Points to Consider (Detailed) list.

**Sponsor and CRO - CRO creates most datasets**  
A majority of work is done by the CRO, but the Sponsor creates some portion of the datasets (e.g., PK, pathology), with the CRO being responsible for compiling the package. For this case, the Sponsor should first engage the CRO as in the previous example, but the CRO may benefit from understanding the format of the datasets it will receive and thus in turn ask the same questions of the Sponsor.

**Sponsor, CRO, and Subcontractor - CRO creates most datasets, subcontractor creates remaining datasets**  
In this case, most of the work is done by the CRO, but a subcontractor is utilized to create some of the dataset contents (e.g., subcontractor doing some lab work or PK), with the CRO being responsible for compiling the package. For this case, the Sponsor would first engage the CRO using the Partnership-level (Initial Survey) followed by the Partnership-level (Detailed) list. The CRO may then also do the same with the subcontractor.

### Points to Consider

Several considerations must be established before initiating a partnership for the assembly of SEND datasets, including determining capabilities, costs and specific needs. The following Points to Consider lists are intended to help facilitate that process.

- Note:**
- To avoid affecting the critical path of a study or submission timeline, anything that can be determined early should be.
  - The answers to these questions should be established early in the process to avoid a negative impact on study timelines, as well as periodically re-assessed as needs and capabilities change.
  - For partnership types not discussed here, the same Points to Consider list should be employed as applicable.

#### Partnership-level Points to Consider

The following are lists of questions and discussion items to raise with a potential partner for the production of SEND or other datasets at the partnership-level. These questions are especially valuable to establish early to avoid affecting study timelines, as well as periodically re-assessing as needs and capabilities change.

#### Partnership-level Points to Consider (Initial Survey)

The following *Initial Survey* list of questions can be used to assess basic capabilities when first exploring a potential partnership. As a partnership evolves, the "Partnership-level Points to Consider (Detailed)" list has more detailed considerations.

- Do you have SEND capabilities? (what endpoints are and are not covered? What study types are covered?)
- Do you have sample datasets for review?
- What portion of your capabilities are automated and/or validated?
- Do you outsource any of these capabilities?
- How many studies have you created SEND datasets for in the last year?
- Do you have alternate means of providing data other than XPT files (e.g., Excel, CSV)?

### Points to Consider (Cont.)

#### Partnership-level Points to Consider (Detailed)

An in-depth list of questions and discussion points at the partnership level. This typically pertains to a Sponsor with a CRO which will be providing SEND datasets, but could apply to a CRO and subcontractor partnership or other types of partnerships.

#### Process Questions

- Can you provide a sample study data package (including sample define file), to review capabilities?
- Will production of SEND datasets be included in master service agreement for all applicable study types or contracted study by study?
- Is there an impact to the delivery timelines for completed studies, if SEND datasets are requested as part of the process to finalize a study?
- Can you provide interim/draft SEND datasets? If yes, what are the costs associated with doing so? (Note that specifics may be defined on a study-by-study basis)
- If any data will be commonly collected by organizations other than the CRO (e.g., if the Sponsor will always do some of the lab work), who will merge the datasets (e.g., LB, CO, RELREC, etc)? (Note that this may also be defined on a study-by-study basis)
- When integrating sponsor and/or subcontractor generated data in SEND format...
  - Who is responsible for each record in SEND for domains which have sponsor and/or subcontractor data?
  - How will the define file be created/merged?
  - How will the Reviewer's Guide be created? (This is expected to be required for FDA submissions)

#### Cost Questions

- What are standard costs for producing SEND datasets for current studies?
- What parts of the assembly incur additional cost?
- What are the costs associated with production of SEND datasets for legacy studies, if this capability is available? (e.g., for special requests for sponsor warehousing)
- What are the costs associated with production of specific versions of SEND or CT, if this capability is available? (this is less common, e.g., for special requests where the receiving organization can only handle specific versions)

#### Compliance Questions

- What is the validation status for 21 CFR Part 11 compliance of the system(s) producing SEND datasets?
- Do you have processes in place to ensure the quality and regulatory adherence?
- How will QA be involved?
- Do you use a tool to validate SEND 3.0 datasets before delivering to sponsors? If yes, what tool / ruleset is used?

#### Fundamental Content Questions

- Which SEND version(s) are supported?
- Which CT version(s) are supported?
- Which domains are generally included (e.g., which are "stock" and which are more manually created by request)?
- Which variables beyond the CRO's sample SEND package can be provided (e.g., Permissible variables not provided by default from CRO and/or Expected variables which are left blank, and of those, which can be reasonably populated)?
- What output formats do you support, including XPT (e.g., \*.xpt, \*.xml, \*.xls, \*.csv, etc.)?

#### Specific Content Questions

- What naming convention is used for USUBJID (e.g., Study + underscore + subject, a numeric ID, etc.)?
- What naming convention is used for STUDYID (e.g., study number, a numeric ID, etc.)?
- How is EX populated - 1 record per dose or 1 record per constant dosing interval (or can both be accommodated)?
- How do vehicle-only doses appear in EX records?
- What conventions are used for creating/naming trial design components (e.g., ARM, ELEMENT, etc.)?
- Which variables are provided in SUPPQUAL datasets? (e.g., RESMOD in MI or a custom variable in SUPPLB for the instrument name)
- Which parameters are included in the Trial Summary, Trial Sets, and Subject Characteristics domains?
- Can you provide lists of the mappings used to Controlled Terminology?
- Can you accommodate extensions or sponsor-specific mappings for SEND CT (e.g., to Sponsor's required terminology) (less common only if desired by the sponsor)

#### RELRECs

- Describe if and how you create RELRECs between PC, PP and CL, MA, MI (i.e., specify the IDVAR used)
- Provide sample cases

#### Define File

- What format and version is used for define file (.xml or PDF)?
- If XML, will a style sheet be provided to view the define.xml, will you also provide a PDF version of the file?
- Will the define file include full controlled terminology lists used during data collection?

### Points to Consider (Cont.)

#### Transfer Logistics

- What methods/tools do you use for transferring files? (e.g., sFTP or other secure file exchange)
- Are there any stipulations on either side for maximum file size?
- Are there any requirements on either side for encryption needed?
- When are the datasets made available versus when can they be available? What should trigger their creation (e.g., X days after data archival; send with draft report; send with final report; etc.)?

#### Study-level Points to Consider

At the study-level, some additional considerations may be necessary to define or refine for the study, points which might differ from the general decisions made at the partnership level.

#### Process Questions

- What data exchange intervals are desired (Interim? Draft? Final?)?
- If any data are collected by multiple parties (e.g., if the Sponsor does some of the LB work), who will merge the datasets (other examples include LB, CO, RELREC, etc.)?

Process for integrating sponsor and/or subcontractor generated data in SEND format

- How will the define file be created/merged?
- What are costs for the study?

#### Compliance Questions

- Review validation needs for 21 CFR Part 11 compliance of the system(s) producing the SEND datasets for the study
- How will QA be involved for the study?

#### Fundamental Content Questions

- Whether any domains should be specifically included or excluded for the study
- Will any domains not be transmitted in SEND format? If so, how will the data be converted into SEND?
- Determine whether any variables beyond the CRO's sample SEND package are desired
- File format for exchange (samples maybe very helpful in \*.xml, \*.xls, \*.csv, etc.)

### Use Case Examples

#### Use Case Example 1

An adaptive study design was used in an exploratory toxicology study, which involved 1 protocol and 1 report, with each divided into 2 parts. Discussions were held between Sponsor and a vendor at the outset of the project to decide how to best represent the trials domains and study data sets in a SEND compliant manner. A joint decision was made to represent the study as 2 independent and complete SEND data sets, each with its own define file. The SEND data sets were linked in the Trial Summary domain with the trial summary parameter, asocstudy (associated study) for 1 study, and altstid (alternate study ID) for both studies. Both SEND data sets were successfully loaded into FDA's SEND system.

#### Use Case Example 2

A Sponsor requested SEND datasets for all of their studies. The Sponsor's initial intent was to use the SEND datasets to populate an internally developed data warehouse. Piloting, including discussions on the contents and logistics as well as producing sample data, proved instrumental in resolving issues ahead of time. A key piece of this piloting work involved the CRO providing sample SEND datasets to the Sponsor which were used to assure that the SEND datasets loaded into the Sponsor's data warehouse as expected. With this test submission, the Sponsor and CRO were able to work through technical issues and expectations in a testing environment. As a result of this piloting activity, the production SEND datasets were submitted and loaded into the Sponsor warehouse successfully with none of the timeline delays that could have occurred if the technical requirements were worked out in production rather than during testing.

### Discussion

Working with multiple external partners and internal systems to create SEND datasets for submission can be a complex process. There are many possible partnership arrangements and many functional details that should be considered prior to conducting the work. Regardless of the partnership type, this Points to Consider list should be used as a tool to; facilitate communication between partners, plan service agreements and plan studies. Addressing these points early in the process can reduce the overall risk for all parties involved in the generation of SEND datasets.

If you would like more information about SEND, selecting a vendor or to access the Points to Consider list please visit; [www.phusewiki.org](http://www.phusewiki.org) and click on *CSS Working Groups*, then *Non-Clinical Road-Map*, then *Impacts on Implementation*.

### Acknowledgements:

Gerard Randolph, Roche TCRC, Inc., New York, NY; Troy Smyrnios, Zoetis, Kalamazoo, MI; William Houser, Bristol Myers Squibb, Mt Vernon, IN; Laura Kaufman, Preclinical Data Systems; Debra Oetzman, Covance Laboratories, Inc, Madison, WI; Lynda Sands, GlaxoSmithKline, King of Prussia, PA; Peggy Zorn, INDS, Inc, Ann Arbor, MI; Steven Denham, MPI Research, Mattawan, MI; Jennifer Feldmann, Instem, Conshohocken, PA; Christy Kubin, MPI Research, Mattawan, MI; Maureen Rossi, Roche TCRC, Inc., New York, NY; Kathryn Brown, Sanofi US, Inc., Bridgewater, NJ;

Note: The opinions expressed in this poster are those of the authors and do not necessarily represent the opinions of their respective companies