



DH12-Reviewing External Vendor Data Transfer Specifications from Programming Standpoint

Mahesh Minnakanti
PHUSE US Connect 2023



Introduction & Background:

- ❑ Data transfer specifications (DTS), drafted by the data management team, carries information on how the non-CRF related raw data should be transferred for integration with the clinical study database.
- ❑ The specifications are finalized after due feedback from the biostatistics, vendors, and the programming team.
- ❑ The data transfer specifications hold a potential to not only to define a dataset but also a potential to delay the go live of a clinical trial if the trial transfers are not carefully tested in time.

Current Challenges:

- The submission database is presented as clean package of datasets with associated define documentation. Creation of the neat submission package is often complicated by data collection inconsistencies and data reconciliation issues that cause inefficiencies to programmers.
- With the addition of vendors comes the challenge of determining how data external to the clinical database will be collected, stored, transferred and integrated.
- DTS/DTA formats vary across vendors and gives rise to multiple mapping efforts for the same terminology (e.g., raw variable names, visit formatting values). Decentralized trials only make the matter more complex.
- Finalizing the DTA before the database go live or First patient is in (FPI).
- Harmonizing data from local laboratories, including mapping to common test names and converting to common units is very time consuming and potentially error-prone.

Objectives:

With the advent of decentralized trials and increase in device data, it is imperative that the data transfer specifications are carefully reviewed and approved. We will discuss

- ❖ The general contents of DTS and what aspects of DTS are to be reviewed from programming standpoint with few examples.
- ❖ Certain approaches on pre-programming activities to map data not only as per CDISC SDTM standards but also from analysis standpoint.

General DTA Contents:

- Import Scope (e.g., Incremental, Cumulative)
- Schedule of Delivery (e.g., every first Monday of the month, ad-hoc)
- Means of Transfer (e.g., sFTP, clients/vendors server)
- File format (e.g., .csv, .sas7bdat)
- File Naming Convention (e.g.,
vendorname_labs_YYYYMMDD_HHMMSS_test_n.SAS7BDAT,
vendorname_labs_YYYYMMDD_HHMMSS_prod_n.SAS7BDAT)
- Blinding and unblinding process (e.g., sections on how blinded data is presented and unblinding section on when and how unblinding will happen.)
- File Structure (e.g., variable names along with length and their possible values and controlled terminology)
- Appendices for value level metadata (e.g., list of possible tests, test codes, along with their units, vendor vs. EDC visit and timepoint mapping tables)

How to Review External Vendor DTA from a Programming Standpoint?

- Prior to the creation of any database, the definition of few key variable values should be considered. Based on the trial design, more variables may be needed for unique identification, but for most trials, subject and visit identifiers are the minimum requirement.
- Ensuring the subject identifier is of the same format in all supporting databases is critical.
- Visit and data time point consistency across the databases.
 - Example 1: Below visit mapping is straight forward across the vendors 1 and 2 with minor naming convention changes.

Study A – Vendor 1		Study A – Vendor 2	
Clinical Database	Vendor Database	Clinical Database	Vendor Database
Screening	Visit 1	Screening	Screening
Week 2	Visit 2	Week 2	Week 2
Week 4	Visit 3	Week 4	Week 4

How to Review External Vendor DTA from a Programming Standpoint?

- Example 2:

- From the below example we understand that the study A has multiple cohorts and each cohort have its own visit schedule.
- Vendor 1 is differentiating between cohort 1 and 2 which makes visit mapping straight forward.
- But Vendor 2 is not differentiating between the cohorts. Which needs to be addressed in the DTA on how the cohort information for a subject is presented.

Study A – Vendor 1		Study A – Vendor 2	
Clinical Database	Vendor Database	Clinical Database	Vendor Database
Cohort 1 Screening Day -28	C1 Visit 1	Screening	Screening
Cohort 2 Screening Day -29	C2 Visit 1	Week 2	Week 2
Cohort 1 Week 2	C1 Visit 2	Week 4	Week 4
Cohort 2 Week 4	C2 Visit 2	Week 6	Week 6

How to Review External Vendor DTA from a Programming Standpoint?

- ✓ In addition, key variables differ from domain to domain and therefore a domain-by-domain review must be made (e.g., laboratory, electronic diary [ediary], electrocardiogram [ECG], magnetic resonance imaging [MRI], pharmacokinetic [PK], Questionnaire etc.).
- ✓ As an example, Safety laboratory testing may utilize the laboratory associated with each site (commonly referred to as "local laboratories"), or a central (core) laboratory.
- ✓ Reviewing and understanding the variable types, description and their formats along with field lengths is important.
- ✓ It is important to review and understand the comments field section in the file structure that contains the examples of data values and all the possible values if no appendix for value level is presented.

Below is the example of a file structure for the falls data:

Col#	Field Name	Item Description	Item Format	Field Length	Comments
1	STUDYID	Study Identifier	CHAR	12	Default: STUDYID
2	SITE	Site number	Numeric	5	The site number is specified by a 5-digit #
3	SUBJID	Subject Identifier	Numeric	3	Subject identifier is specified by a 3-digit #
4	RECTYPE	Type of record	CHAR	5	Code list: • Compliance • Fall
5	DATE	Date of recorded fall and daily compliance of medication and fall	DD-MMM-YYYY	11	e.g., 05-May-2010
6	INPUT	Source of input	CHAR	6	Code List: Sensor : (from sensor device) Diary : (reported by patient)
7	Time	Time	Hh:mm:ss (24-hour clock)	8	e.g. 13:25:02

If RECTYPE==Compliance, time is set to 00:01 (dummy time)

If RECTYPE==Fall and INPUT==Diary, this denotes the time when the user reported the fall during fall diary

If RECTYPE==Fall and INPUT==Sensor, this denotes the time when the fall was detected by the sensor.

Few recommendations and best practices:

- ✓ It is always good to check with the vendor if they can provide data in SAS format.
- ✓ Central lab samples data is processed in the same manner to yield a single data file containing common test names and units thus bringing efficiency to the SDTM programming.
- ✓ Make sure the mapping related variables such as categories, specimen details, alert flag and a flag variable for not done records along with the comments are reported.
- ✓ Seek to request and procure supporting documents (such as conversion factors and testing attributes file) well ahead of time or in time.
- ✓ Requesting project Manager to accommodate enough time for planned test transfers to help identify data mapping issues before hand.

Few recommendations and best practices:

- ✓ It is a good practice that any production data transfer from Vendor are first reconciled by CDM team before any programming activities begins.
- ✓ Any inconsistency between the DTA and production data are to be communicated to the vendor immediately and no work around programming should be done.

Few Considerations for Pre-programming activities for Vendor Data:

Based on the DTA, creating and maintaining variable level format catalogs. (e.g., visit, test, units) for creating submission level data.

Build some checks to

- identify unexpected unblinding issues.
- identify unexpected visits.
- Check for value level formatting issues.

CONCLUSION and TAKE AWAYS:

- ✓ Programming inputs and review of DTA at the development stage brings more efficiency and programming awareness.
- ✓ It is good to have finalized DTA before the database go live along with the CRF. It gives enough time for the programming team to have the submission datasets ready if there is any planned immediate safety data review.
- ✓ In addition, development of proactive reconciliation checks result in cleaner data, and ultimately more efficient statistical programming time.
- ✓ Selection of vendor plays a major role. As now a days some of the vendors maintains robust database systems that are close to the submission standards which makes programming easy.

Thank you!!