

A Practical Introduction to Clinical Data Acquisition Standards Harmonization (CDASH)

Jennifer Price, Phoenix Data Systems, Inc, King of Prussia, PA

ABSTRACT

On October 1st, 2008, CDASH released the first 16 common CRF streams (or domains) for use by the Pharmaceutical Industry.

The goal when putting together the initial domain streams was to find out which data fields are essential to the analysis of clinical data, and collect only that data in a standard way. In order to do this, CDASH has provided clear definitions, SDTM mapping (if relevant), CRF completion instructions and additional sponsor information for each data element.

The focus of CDASH is on data collection, not data reporting. In some instances the optimal data collection method conflicts with the Study Data Tabulation Model (SDTM) for reporting data. In these cases additional transformations and derivations may be needed to create the final SDTM compliant datasets.

One way to think of it, is that the SDTM standard is for presenting study data using Clinical Data Interchange Standards Consortium (CDISC) compliant specifications, and the CDASH standard is for the collection of this same data.

BENEFITS OF IMPLEMENTING CDASH

Implementing a standard data collection methodology has many benefits. The main benefit is standardizing the definitions for the data that is collected over multiple studies. CDASH defines data that can be used in the cleaning of data and for the conformation of missing data. CDASH is valuable for reducing the production time for CRF design, reducing the training time for sites.

WHAT DOES CDASH LOOK LIKE?

Each domain lists all of the common variables for that domain along with a definition and implementation instructions. Here is a sample:

CDASH CRF Label/Question	Clinical Database Variable Name <i>CDASH variables shaded</i>	Definition	Instruction to Clinical Site	Implementation / Rationale to Sponsors	CDASH Core
Date of collection	LBDTC <i>(expected)</i>	Date of sample collection	Record the date collection occurred.	This is intended to be used as a data management tool to verify that lab data is provided in the electronic data for each line that no data was collected. The date of collection may be derived from the date of visit and if so, a separate assessment date field is not required.	Highly Recommended
Lab Status	LBSTAT <i>(permissible)</i>	Status of whether or not lab was done.	Indicate whether or not lab was done.	This may be implemented for an entire panel, or on a test-by-test basis. This is intended to be used as a data management tool to verify results provided.	Highly Recommended
Panel Name	LBCAT <i>(expected)</i> LBSCAT <i>(permissible)</i>	Type of draw / category / panel name. Used to define a category of related records.	Record the lab test category, if not pre-printed on the CRF	Examples: such as HEMATOLOGY, URINALYSIS, CHEMISTRY	Optional

The common variables: STUDYID, SITEID or SITENO, SUBJID, USUBJID, and INVID that are all SDTM variables with the exception of SITEID which can be used to collect a Site ID for a particular study, then mapped to SITEID for SDTM.

Common timing variables are VISIT, VISITNUM, VISDAT and VISTIM where VISDAT and VISTIM are mapped to the SDTM –DTM variable.

Each variable is defined as:

- Highly Recommended: A data collection field that should be on the CRF (e.g., a regulatory requirement).
- Recommended/Conditional: A data collection field that should be collected on the CRF for specific cases or to address TA requirements (may be recorded elsewhere in the CRF or from other data collection sources).
- Optional: A data collection field that is available for use if needed.

CHALLENGES OF IMPLEMENTING CDASH

There are many challenges to implementing a CDASH compliant CRF. By nature, clinicians are curious. It has been my observation that clinicians sometimes want to collect data that isn't necessarily required for analysis, but may be interesting to them. Statisticians may want to group data into 'categories' that don't exactly fit the data. Data Managers are interested in standardizing the way data is collected in order to link data from different systems such as lab data and EDC data.

While implementing standards is a noble goal, keeping up with the standards and modifying standard documentation is a challenge, especially in long term studies that are ongoing. Mapping legacy data to conform to a new standard is also a challenge.

DETAILS OF IMPLEMENTING CDASH

The CDASH standards are the same for paper-based and EDC systems. Each stream or dataset usually represents a single CRF. For each stream or dataset, data collection fields are presented as 'Highly Recommended', 'Recommended/Conditional' and 'Optional'. Each company will need to decide if the fields identified collect all data required for their specific study protocol. The common fields that are Highly Recommended on every stream are: Protocol, Site and Subject Identifier. The only timing variable that is Highly Recommended is Visit Date. Visit Time is conditional and Visit ID can be derived.

Here is an example of the CDASH information specified for the Adverse Event Verbatim Term:

Variable Name:	AETERM
Definition:	Verbatim (i.e., investigator reported term) description of the adverse event.
Case Report Form Completion Instructions:	Record only one diagnosis, sign or symptom per line (e.g., nausea and vomiting should not be recorded in the same entry, but as 2 separate entries). Using accepted medical terminology, enter the diagnosis (if known); otherwise enter a sign or symptom. If a diagnosis subsequently becomes available, then this diagnosis should be entered on the AE form, replacing the original entries, where appropriate. Death should not be recorded as an event but should be recorded as the outcome of the event. The condition that resulted in the death should be recorded as the event. Do not use abbreviations

It is interesting to note what is not included in the CDASH specifications. Most notably, the order of the fields collected is not specified. It is up to each sponsor to determine logical placement for each form. The field lengths are not specified, and the contents required within a field are not specified in this document, although there is an Appendix with commonly used terminology such as dose units and routes.

Many of the codelists or formats are referred to, but not defined in this document.

In the case of the Adverse Event Severity (AESEV) field, which CDASH defines as:

'The reporting physician/healthcare professional will assess the severity of the adverse drug/biologic event using the sponsor defined categories. This assessment is subjective and the reporting physician/healthcare professional should use medical judgment to compare the reported Adverse Event to similar type events observed in clinical practice. Severity is not equivalent to seriousness.'

The SDTM terminology guide defines the categories as:

- MILD
- MODERATE
- SEVERE
- Missing value is not allowed

This illustrates that CRF form design must take into consideration multiple standards documents.

CHALLENGES OF IMPLEMENTING CDASH

GENERIC CHALLENGES:

The collection of dates is challenging because CDASH does not define when a whole date and when a partial date should be used. The collection of free text and comment fields is discouraged. It is recommended that data collection methods maximize the use of pre-defined lists of responses.

CDASH recommends the use of YES/NO questions whenever possible, there are two exceptions:

1. assessments where the majority of options would be answered 'NO' can have a single 'YES' option,
2. the question 'Check if Ongoing' typically has an option of YES.

CDASH doesn't typically define codelists but refers to the SDTM Terminology documentation, but in some cases where a codelist isn't listed in the SDTM document, CDASH will recommend terms.

SPECIFIC CHALLENGES:

Each domain along with some of the specific challenges is listed below.

DOMAIN: ADVERSE EVENTS - AE

Highly Recommended: AETERM with complete instructions and examples, AESTDAT and AEENDAT, AESEV, AESER, AEREL, AEACN, AEOUT

- CDASH defines an optional prompt field (AEYN) that can be used. The text is: 'Indicate if the subject experienced any adverse events. If yes, include the appropriate details where indicated on the CRF'. This field is only used to help with data cleaning and monitoring and is not included in the SDTM.
- AEONGO is collected, but is not directly mapped to the SDTM variable AEENRF.
- AESTTIM and AEENTIM are Recommended/Conditional, and if collected are combined with AESTDAT and AEENDAT to create the SDTM variables AESTDTC and AEENDTC.

DOMAIN: COMMENTS - CO

It is recommended that unsolicited comments not be collected. There are no field requirements for this domain.

DOMAIN: PRIOR AND CONCOMITANT MEDICATIONS - CM

Highly Recommended: CMTRT with complete instructions and examples, CMSTDAT and CMENDAT

- CDASH defines an optional prompt field (CMYN) that can be used. The text is: 'Indicate if the subject took any medications. If yes, include the appropriate details where indicated'. This field is only used to help with data cleaning and monitoring and is not included in the SDTM.
- An optional Ingredient field CMINGRD can be collected to be used to help with coding. This variable gets dropped and is not included in the SDTM dataset.
- Con Med details can be linked to the relevant medication history and adverse event data with CMMHNO and CMAENO respectively.
- Dose (CMDSTXT) is a CDASH field that is mapped to SDTM CMDSTXT or CMDDOSE.
- Total Daily Dose (CMDOSTOT) is not the recommended method for collecting dose, it is recommended to collect dose, unit and frequency and derive a Total Daily Dose.
- CMSTTIM and CMENTIM are Recommended/Conditional, and if collected are combined with CMSTDAT and CMENDAT to create the SDTM variables CMSTDTC and CMENDTC.
- Ongoing can be collected (CMONGO) This is not a direct mapping to CMENRF, but is used along with the date of data collection in conjunction with end date and the ongoing check box would determine how CMENRF will be populated.

DOMAIN: DEMOGRAPHICS - DM

Highly Recommended: BRTHDAT (or BRTHYR, BRTHMO, BRTHDY and/or BRHTM), SEX.

- CDASH recommends collecting the complete date of birth, but recognizes that in some cases only BIRTHYR and BIRTHMO are feasible.
- The CDASH document references the CDISC SDTM terminology document. This document lists four options for the collection of Sex: Male, Female, Unknown and Undifferentiated (M|F|U|UN). CDASH allows for a subset of these codelists to be used, and it is typical to only add the options for Male or Female.

DOMAIN: DISPOSITION - DS

Highly Recommended: DSDECOD or DSTERM

- This domain collects any type of disposition event or protocol milestone event. Samples are 'Date of Informed Consent', or 'Date of Randomization' or 'Reason for Discontinuation'.
- The standardization of collecting of this type of data are challenging since these types of questions tend to appear with other data on the CRF. These epochs tend to be pre-printed on the CRF. The date related to these epochs is stored in DSSTDAT.

- CDASH recommends using the controlled SDTM terminology or a subset of the defined terms for DSDECOD and recommends the reasons you choose should be pre-printed on the CRF. The wording for this question is *'Document the subject's status at <insert text corresponding to the select trial epoch>. If the subject discontinued prematurely, record the primary reason for discontinuation'*
- The date of collection for Demography can be collected in DMDAT and is mapped to DMDTC.

DOMAIN: DRUG ACCOUNTABILITY - DA

Highly Recommended: DATEST, DAORRES, DAORRESU

This is an optional CRF and recommended for multiple dose studies.

DOMAIN: ECG TEST RESULTS - EG

There are three scenarios presented for ECGs:

1. Central reading for results that are captured electronically by a central vendor.
Highly Recommended: EGPERF
2. Local reading for results that are captured on the CRF.
Highly Recommended: EGPERF, EGDAT, EGTEST, EGORRES
3. Central reading for results that are captured electronically by a central vendor and evaluated for clinical significance.
Highly Recommended: EGPERF, EGDAT, EGTEST, EGCLSIG

The implementation of this standard is fairly straightforward. It is interesting to note that when collecting clinical significance, this data in this form is not collected in SDTM. Typically the data in this field is used to cross-check that an adverse event of this type is entered.

DOMAIN: EXPOSURE - EX

Highly Recommended: EXSTDAT

There are several fields defined to collect data about missing or adjusted dose amounts. This information is used to clean the data and does not get mapped to SDTM.

DOMAIN: INCLUSION AND EXCLUSION CRITERIA - IE

Highly Recommended: IEYN, IETESTCD

- The CDASH domain mimics the SDTM domain in which only the criteria NOT met are collected.
- IEYN is a CDASH field that is used to assure data managers that all inclusion/exclusion criteria was considered. The recommended wording of this field is *"Record 'Yes' if all eligibility criteria were met for the study. Record 'No' if the subject did not meet all criteria at the time the subject was enrolled."* If NO is recorded, the criteria NOT met are collected in IETESTCD.

DOMAIN: LABORATORY TEST RESULTS - LB

There are three scenarios presented for Laboratory Test Results:

1. Central processing for results that are captured electronically by a central vendor.
Highly Recommended: LBPERF, LBDAT
2. Local processing for results that are captured on the CRF.
Highly Recommended: LBPERF, LBDAT, LBTEST, LBORRES, LBORRESU
3. Central processing for results that are captured electronically by a central vendor and evaluated for clinical significance.
Highly Recommended: LBPERF, LBDAT, LBTEST, LBCLSIG

The implementation of this standard is fairly straightforward. It is interesting to note that when collecting clinical significance, this data in this form is not collected in SDTM. Typically the data in this field is used to cross-check that an adverse event of this type is entered.

DOMAIN: MEDICAL HISTORY - MH

Highly Recommended: MHTERM

- The question *"Has the subject experienced any past and/or concomitant diseases or past surgeries"* can be asked and results are stored in MHYN. This is used to aid in the monitoring and data cleaning of the medical history. Terms may be collected on a form next to their associated category, condition or event where the category, condition or event is pre-printed on the CRF. If this is the case, the data pre-printed on the CRF must be included.
- The collection of the Start and End date of the event is considered optional.

DOMAIN: PHYSICAL EXAMINATION - PE

Highly Recommended: <none>

- CDASH is recommending that clinical sites are asked to record baseline abnormalities on a Medical History, Targeted Medical History or Baseline Conditions CRF. Post baseline abnormalities or baseline conditions that worsened during the clinical study are to be recorded on the AE CRF, therefore the only

required field (PEYN) on the PE domain is to record whether an exam was done, and this is only required for data cleaning purposes.

- A traditional approach is also offered.
- Highly Recommended: PETEST, PERES, PEDESC
- In the traditional approach, each body system is listed on the CRF (PETEST) and the instructions state that Normal, Abnormal and Not Done options be collected in PERES.

DOMAIN: PROTOCOL DEVIATIONS - PV

Highly Recommended: DVTERM (This field is only highly recommended if the DV CRF is created)
Protocol Deviation data is derived from other domains and should not be collected as a specific CRF.

DOMAIN: SUBJECT CHARACTERISTICS - SC

Highly Recommended: SCTEST, SCORRES

- This domain is for subject related characteristics that are collected only once per subject. Samples of this data are marital status, educational level, other race and gestational age at birth.
- This type of data tends to be collected along with other data and typically does not have its own CRF.

DOMAIN: SUBSTANCE USE - SU

Highly Recommended: SUTRT, SUNCF

The question “*Has the subject ever used/consumed tobacco / alcohol / caffeine, currently consumes tobacco / alcohol / caffeine, or formerly used/consumed tobacco / alcohol / caffeine.*” can be asked and results are stored in SUNCF. The recommended responses are ‘Never’, ‘Current’, and ‘Former’. This is used to aid in the monitoring and data cleaning. Duration and Units are collected separately as SUCDUR and SUCDURU respectively. CDASH recommends the duration options (weeks, months, and years) be printed on the CRF. These two fields will be presented as SUDUR in the SDTM.

DOMAIN: VITAL SIGNS - VS

Highly Recommended: VSTEST, VSORRES, VSORRESU

The Vital Signs domain is fairly straightforward collecting the VSTEST, VSORRES and VSORRESU. CDASH also allows for VSCLSIG to collect whether a vital sign was clinically significant. This indicator can be used in data cleaning.

CONCLUSION

CDASH standards are not a simple end all that determines once and for all which data is collected for each and every protocol, but it is a starting block that can be used to develop standard forms with standard data collection instructions. By developing these standards, sites will benefit since they will be able to report data for all studies using the same standard directions and pharmaceutical companies will benefit by not having to reinvent case report forms and instructions for each and every protocol. Data collected using CDASH standard can be easily mapped to SDTM for submissions.

The opinions in this paper are purely my own and not necessarily endorsed by CDISC or CDASH.

REFERENCES

CDISC CDASH Version 1, September 2008, for more information on the topic follow link:

http://www.cdisc.org/standards/cdash/downloads/CDASH_STD-1_0_2008-10-02.pdf

CDISC SDTM Implementation Guide Version 3.1.1, September 2005, for more information on the topic follow link:

<http://www.cdisc.org/models/sdtm/v1.1/index.html>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jennifer Price

Senior Director, Clinical Solutions

Phoenix Data Systems, Inc. a division of Bio-Imaging Technologies (NASDAQ=BITI)

email: jprice@pdsedc.com website: www.phoenixdatasystems.net

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.