Listen to a recording of this issue

The Altascientist
SCIENTIFIC JOURNAL
ISSUE NO. 26

CDISC DATA STANDARDS

Simplifying and Expediting Data Management and Analysis

> Adherence to data standards is integral to successful and efficient drug development for our clients. We are experts in applying CDISC standards, from preclinical SEND to clinical CDASH, SDTM, and ADaM. We are proud to be Gold members of the CDISC consortium, and to bring the advantages of standardization to all the data we deliver.
>
> **- Nicole Maciolek**
> Vice President, Research Services

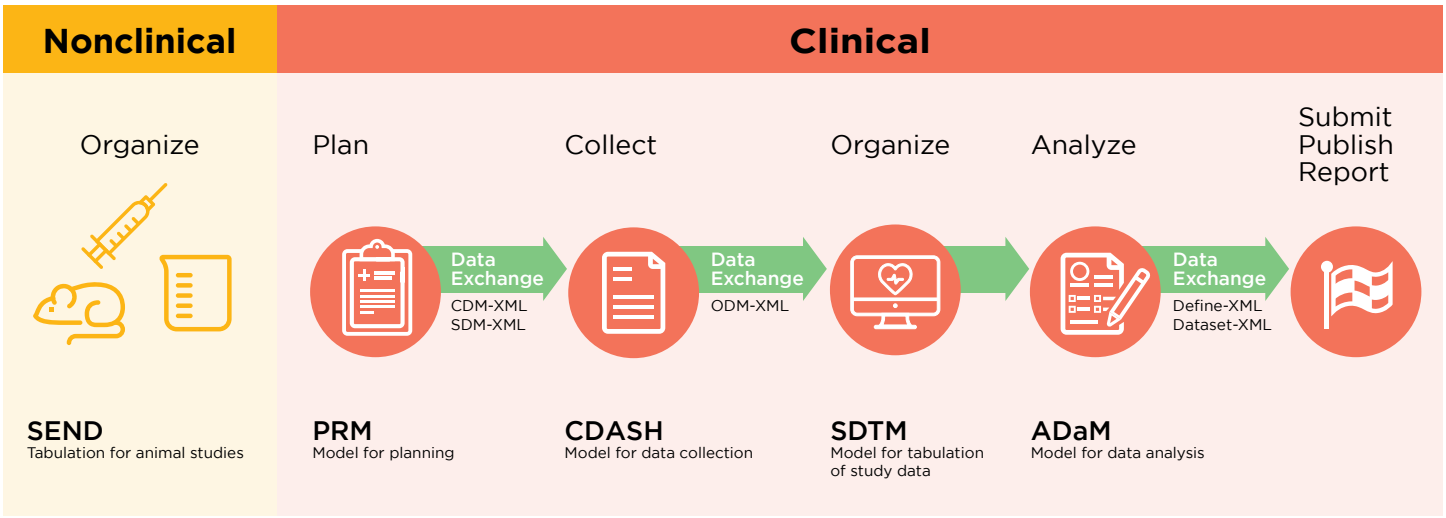
WHAT IS CDISC?

The Clinical Data Interchange Standards Consortium (CDISC) is a worldwide organization for data standardization, ensuring that drug research data delivers the maximum value for sponsors, regulatory agencies, and patients. Data that is accessible, compatible, comparable across regions, and reusable for meta-analysis or reanalysis, serves to improve our understanding of human therapeutics by providing meaningful, efficient research data for the entire global drug research community. Implementing standards to collect, structure, and analyze data makes it easier to aggregate information and take advantage of big data.

CDISC standards apply throughout the drug development journey, and are required for all studies supporting market authorization, as shown in Figure 1.

Figure 1.

CDISC Standards in the Drug Development Process[1,2]



Legend

SEND = Standard for non-clinical data
PRM = Protocol representation model
CDASH = Clinical data acquisition
SDTM = Study data tabulation model
ADaM = Analysis data model
ODM XML – operational data model
Define XML – dataset metadata
Dataset XML – dataset data

THE POWER OF STANDARDIZATION

There is tremendous value in standardizing and sharing data: organizations across the globe that are using CDISC standards produce faster, more efficient research, and pave the way for more breakthroughs that amplify the power of data, in both the short- and long-term.

Data that is presented in a well-organized, templated fashion is easier for regulatory reviewers to understand and interpret, which limits requests to sponsors for clarification or resubmission. As shown in Figure 1, the CDISC standards cover all types of data, from non-clinical to Phase III, across therapeutic areas.

Sponsors benefit from data that is of high quality, easy to interpret, and leads to sound, swift Go/No-Go decisions. Regulatory bodies benefit from a smooth, consistent process that allows them to analyze all the data submitted in the same way, without concerns about the organization or interpretation of data that may be structured differently.

Ease of Review and Impact on Approval Time

CDISC-compliant data greatly facilitates the work of regulatory bodies in approving new drugs for market. In the past, studies submitted would have different structures, with data organized and analyzed in a number of different ways, making it difficult to interpret or compare the quality of the data, the results, and to understand their therapeutic impact.

With the ability to compare datasets easily, across submissions and over time, regulatory agency reviewers can more effectively determine whether a new drug provides significant therapeutic benefit and an additional option for patients. Study parameters, results, and therapeutic claims can be compared with relative ease, without the constant need to remap data points to make a relevant analysis. The ability for regulatory agencies to effortlessly collate and compare submissions can have a positive impact on the overall approval process.

Without any standards, the values under a given variable can be input in a different format, while having the same meaning. In order to avoid confusion, or the need to relabel data to align, controlled terminology sets rules for the values that can be entered in specific variable fields. By consistently applying these rules, common understanding is achieved as to what each code and abbreviation means. This is not to say that without standards it is impossible to interpret the data; explanatory documentation is provided with each dataset. It is to say that controlled terminology allows programs to be put in place without requiring reviewers to constantly refer to the documentation for a clear understanding of the data.

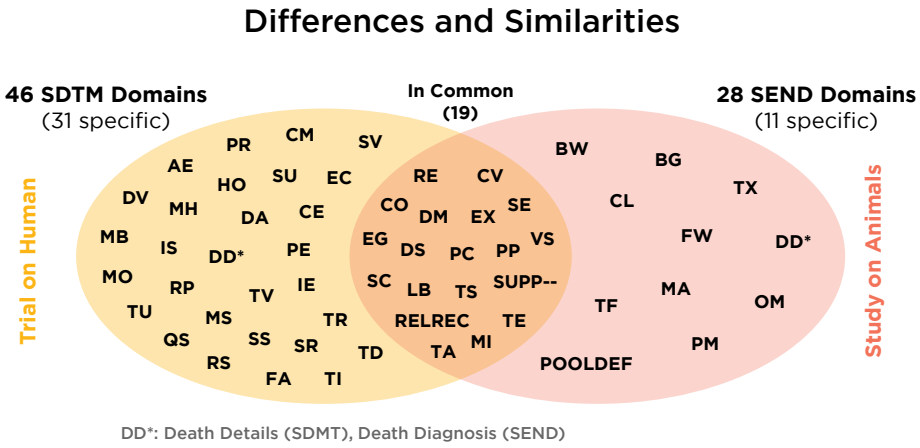
At Altasciences, we have observed an improvement in approval time for Phase I studies submitted to the FDA since the introduction of CDISC standards.

CLIENT CONSIDERATIONS FOR NONCLINICAL AND CLINICAL DATA STANDARDIZATION

Current SEND Model

After the successful implementation of SDTM for submissions to the FDA for clinical studies, a standardized electronic format was also requested for nonclinical studies. The nonclinical model is named Standard for Exchange of Nonclinical Data (SEND), and the goal is to standardize submissions from all parties. This standardization allows for the warehousing and comparison of drugs in similar families, ultimately allowing for better analysis of drug candidates and shortening regulatory decision time.

Figure 2. Venn Diagram depicting the similarities and differences between SDTM and SEND domains.



DD*: Death Details (SDMT), Death Diagnosis (SEND)

SEND 3.0

SEND version 3.0 was the first version to be accepted by the FDA for nonclinical submissions. Compared to the clinical datasets, SEND has a narrow scope of supported study types (eCTD Filing Study Type):

- Single-dose general toxicology (eCTD Section 4.2.3.1)
- Repeat-dose general toxicology (eCTD Section 4.2.3.2)
- Carcinogenicity studies (eCTD sections under Section 4.2.3.4)

The narrow scope of SEND allows for all nonclinical studies being submitted outside these three eCTD filings to be excluded from SEND format requirements.

The implementation date requirements for SEND 3.0 for FDA submissions can be found in the table below.

Table 1. SEND version 3.0 requirement dates for FDA

SUBMISSION TYPE	SUPPORT START	REQUIREMENT START	REQUIREMENT ENDS
NDA, ANDA, and certain BLA submissions	March 13, 2011	Studies which start after December 17, 2016	Studies which start after March 15, 2019
Commercial INDs and amendments		Studies which start after December 17, 2017	Studies which start after March 15, 2020

SEND 3.1

SEND version 3.1 was released by CDISC on June 27, 2016. It expands on the previous version in both scope and minor improvements, and supports the following study types (eCTD Filing Study Type):

- Single-dose general toxicology (eCTD Section 4.2.3.1)
- Repeat-dose general toxicology (eCTD Section 4.2.3.2)
- Carcinogenicity studies (eCTD sections under Section 4.2.3.4)
- Respiratory and cardiovascular safety pharmacology studies (eCTD Section 4.2.1.3)

Along with the addition of safety pharmacology studies to the scope, this version of SEND allows for the introduction of custom domains. This means that if a study contains information that has not been modeled in an established domain, a domain can be created or modified from a SDTM domain to fit the needs of the specific information.

The implementation date requirements for SEND 3.1 for FDA submissions can be found in the table below.

Table 2. SEND version 3.1 requirement dates for FDA

SUBMISSION TYPE	SUPPORT START	REQUIREMENT START	REQUIREMENT ENDS
NDA, ANDA, and certain BLA submissions	15 March 2019	• March 15, 2019 – NDA, ANDA, and certain BLA submissions • March 15, 2020 – IND submissions	Ongoing
Commercial INDs and amendments		Studies which start after December 17, 2017	

Clinical Considerations - Quick Reference

The clinical datasets are comprehensive and complex, and there are many considerations that can impact successful data capture. Our CDISC experts have compiled the following abbreviated guide. Complete information can be found on the CDISC website.

Protocol development

- Controlled terminology[1, 2]
  - Verify which protocol terminology version should be used

External data transfer planning

- Controlled terminology (CT)
- SDTM Implementation Guide (SDTMIG)
- Therapeutic Areas Guide

CRF building and annotation

- Controlled terminology
  - Verify which SDTM/CDASH terminology version should be used
  - Ensure that the terms can be mapped to the controlled terminology (CT)
  - Verify if you need to send a request to the CDISC organization to add new terms to the CT
- CDASH Implementation Guide (CDASHIG)[3]
  - Verify with regulatory agencies which CDASHIG version is supported for submission to the agency
  - Ensure the data can fit into existing domain; otherwise, plan for the creation of a custom domain
  - Ensure the variables are followed as per the established standards in CDASHIG
- Coding Dictionaries
  - Identify all the required coding dictionaries to be used (MedDRA[4], WHO Drug Dictionaries[5], etc.)
- SDTM Metadata Submission Guideline[6]
  - Create SDTM annotations on the CRF by following the standards established in SDTM Metadata Submission Guideline
- SDTM Implementation Guide (SDTMIG)[7]
  - Verify with regulatory agencies which SDTMIG version is supported for submission to the agency
  - Ensure the data can fit into existing domain; otherwise, plan for the creation of a custom domain
  - Variable naming as per SDTMIG
  - Ensure Required (Req) and expected (Exp) variables as defined by SDTMIG are collected
  - Ensure remaining variables are permissible (Perm) variable. If it is not Req, Exp, or Perm, then the variable should be mapped to to supplemental qualifier (SUPP--) domain.
  - Ensure each record includes identifier and timing variable as well as topic variable.
- Therapeutic Areas Guide[8]
  - For certain therapeutic areas, ensure the standards established in the associated therapeutic areas guide (TAUG) are followed when applicable

SDTM datasets development

- Controlled terminology (CT)
- SDTM Model
  - Verify with regulatory agencies which SDTM version is supported for submission to the agency
- SDTM Implementation Guide (SDTMIG)
- Therapeutic Areas Guide
- Validation with regulatory validator (Pinnacle 219, etc.)
  - Verify that the SDTM datasets are regulatory compliant
  - Check that there are no unexplained or unacceptable errors or warning in the validation report
- Coding dictionaries
- Dataset files
  - Trial domain datasets (TA, TE, TV, TS, etc.)
  - Special purpose domain datasets (CO, DM, SE, SV, and SM)
  - Finding domain datasets (EG, LB, PC, PP, VS, etc.)
  - Finding about events domain datasets (FA and SR)
  - Intervention domain datasets (CM, EX, SU, etc.)
  - Event domain datasets (AE, DS, DV, MH, etc.)
  - Relationship domain datasets (RELREC, SUPP--, etc.)

ADaM datasets development

- Controlled terminology (CT)
  - Verify which ADaM terminology version should be used
- Coding Dictionaries
- ADaM Model[10]
  - Verify with regulatory agencies which ADaM version is supported for submission to the agency
- ADaM Implementation Guide (ADaMIG)[11]
  - Verify with regulatory agencies which ADaMIG version is supported for submission to the agency
  - Variable naming as per ADaMIG
  - Timing variable conventions
  - Date and time imputation flag variables
  - Flag variable conventions
  - Record-level treatment and dose variables for basic data structure (BDS) datasets
- Validation with regulatory validator (Pinnacle 21, etc.)
  - Ensure the ADaM datasets are regulatory compliant
  - Ensure all unacceptable errors and warning are resolved or explained
- Datasets files
  - Subject level analysis datasets (ADSL)
  - BDS datasets (ADVS, ADAE, etc.)

Table, listings and figures (TLFs) creation

- Predecessors and traceability
  - Data sources for TLFs: Everything in the TLFs can be traced back to the SDTM/ADaM datasets

Submission deliverable

- SDTM Define
  - aCRF: Blank CRF with SDTM annotations
  - Clinical Study Data Reviewer's Guide (cSDRG)[12]
    - Additional information for regulatory reviewer
    - Data conformity
  - SDTM submission datasets (XPT format)
  - Define.XML
    - Metadata intended to describe the format and content of the study data (datasets, variables, value-level metadata, and codelist)
- ADaM Define
  - Analysis Data Reviewer's Guide (ADRG)[13]
    - Additional information for regulatory reviewer
    - Data conformity
    - Duplicates limited information found in other submission documentation (Protocol, SAP, CSR, define.xml)
  - ADaM submission datasets (XPT format)
  - ADaM, tables, and figures programs
  - Define.XML
    - Metadata intended to describe the format and content of the study data (datasets, variables, value-level metadata, and codelist)

1 https://www.cdisc.org/standards/terminology/controlled-terminology
2 https://datascience.cancer.gov/resources/cancer-vocabulary/cdisc-terminology
3 https://www.cdisc.org/standards/foundational/cdash
4 https://www.meddra.org/
5 https://who-umc.org/whodrug/whodrug-global/
6 https://www.cdisc.org/standards/foundational/sdtm
7 https://www.cdisc.org/standards/foundational/sdtmig
8 https://www.cdisc.org/standards/therapeutic-areas
9 https://www.pinnacle21.com/products/validation
10 https://www.cdisc.org/standards/foundational/adam
11 https://www.cdisc.org/standards/foundational/adam
12 https://advance.phuse.global/display/WEL/Clinical+Study+Data+Reviewer%27s+Guide+%28cSDRG%29+Package
13 https://advance.phuse.global/display/WEL/Analysis+Data+Reviewer%27s+Guide+%28ADRG%29+Package

CASE STUDY – REALIZED EFFICIENCY

Altasciences conducted a Phase I study in 36 participants to establish a clinical bridge via bioequivalence between the proposed drug product and the listed drugs to establish the scientific basis for reliance on the FDA's previous findings of safety and efficacy in support of the proposed 505(b)(2) New Drug Application (NDA). Data was collected, compiled, and analyzed according to CDISC standards.

The clinical conduct of the study was completed in 2020, with results submitted to the FDA in February 2021. The NDA was approved in December 2021.

Between 2009 and 2015 (prior to CDISC), a sample of 284 applications demonstrated mean approval time of 17 months.[3] Our recent experience, since the advent of CDISC standards, demonstrates approval time at approximately 12 months. Although the demonstrated efficiency is likely not 100% attributable to use of the CDISC standards, we are confident that the ease of processing and analysis contributed significantly.

Ability to Reuse Data

CDISC standards ensures that your data remains useable over time. For example, a sponsor that is building a 505(b)(2) application using both previous and current data will find the file simpler and easier to build if all the datasets are in the same structure, and can be compared and compiled across studies. Relatively simple fields such as gender or vital signs can be captured with different headings, depending on the design of the initial database. For example, gender fields may be designated as Male/Female, M/F, Man/Woman, Men/Women, M/W, etc. Not only can there be a difference in the label, but the way the data under that label is captured can also be different.

An example is presented below, comparing two studies, one of which has two sites. All three of the sites are conducting studies on the same compound, and capturing the same information (subject ID, sex, and date of informed consent).

CDISC

Study 001 Compound X Site 01– CRO 1

USUBJID	SEX	RFICDTC
001-01-001	F	2021-03-04T08:00
001-01-002	M	2021-03-05T09:00
001-01-003	F	2021-03-04T10:30

Study 002 Compound X Site 01– CRO 2

USUBJID	SEX	RFICDTC
002-01-1005	M	2021-03-06T10:00
002-01-1006	M	2021-03-07T08:00
002-01-2000	F	2021-03-04T12:00

Study 002 Compound X Site 02 – CRO 3

USUBJID	SEX	RFICDTC
002-02-1005	F	2021-03-04T09:00
002-02-1006	F	2021-03-07T10:00
002-02-2000	M	2021-03-04T12:00

Non-CDISC

Study 001 Compound X – CRO 1

SUBJECT NUMBER	GENDER	ICF DATE-TIME
001	F	20210304-0800
002	M	20210305-0900
003	F	20210304-1030

Study 002 Compound X Site 01– CRO 2

ID	SEX	DDMMYYY	HHMM
1005	M	06032021	1000
1006	M	07032021	0800
2000	F	04032021	1200

Study 002 Compound X Site 02– CRO 3

SUBJID	GEN_SUBJ	ICF_DATE_TIME
1005	0	202103040900
1006	0	202103071000
2000	1	202103041200

While all the data is present, aggregating or comparing will require additional programming to align the database structure and allow for analysis. We can clearly see that amalgamating the data in the non-CDISC format will require manipulation of both the header and data formats, to create a common structure.

Facilitating the reuse of older, non-CDISC data that may be relevant for a different indication, formulation, or to compare across molecules within a therapeutic area, is a definite advantage of having standard datasets available. As the industry transitions to the CDISC standards, legacy study data conversion to standardized study data is often necessary and requires expert support to align the legacy data with SDTM and ADaM standards.

At Altasciences, the Data Services team has vast expertise and experience with database programming, and can support sponsor needs in reformatting previous data, as well as ensuring that all current data is pristine.

CASE STUDIES – LEGACY DATA CONVERSION

For regulatory purposes: A sponsor had conducted Phase I studies at Altasciences prior to the implementation of CDISC standards. The sponsor requested the legacy data conversion to standardized study data format for older studies being included in the FDA submission, to facilitate and simplify the analysis of their submission package by the regulatory authorities. Altasciences was able to remap the data to CDISC standards so that the sponsor was able to provide all data to the regulatory authorities in a clear, consistent format, and help minimize requests for clarification or explanation.

For internal use: several sponsors have requested to remap the databases of multiple older studies performed at Altasciences to SDTM format for their internal use. Having the ability to combine older and newer data to analyze results over the longer term, and across multiple studies, is value-added during the lifecycle of a product, and when comparing data for multiple products over time. Altasciences has the skilled team in place to perform such remapping as requested for any of our clients.

FOCUS ON THE FUTURE

CDISC standards are designed to benefit artificial intelligence and machine learning algorithms, due to their well-defined standard datasets that follow a common structure, making the data easier to aggregate and analyze. The standards are publicly available for those who register for a free account. For a fee, detailed information is accessible via the CDISC API, so that technical staff and software developers can create new tools and technical integrations for benefit of the entire research community, now and into the future.

Currently, the FDA in the U.S. as well as the PMDA in Japan require that data be submitted in conformity with CDISC standards. In the EU, the EMA accepts CDISC-compliant data, not as a mandatory approach, and is actively looking towards developing data standardization policies.

On May 18, 2021, the EMA held a virtual stakeholder workshop regarding the launch of an initiative to develop a data standards strategy, with a view to addressing the HMA/EMA Big Data Task Force recommendation to engage in international initiatives related to data standardization. During this session, the president of CDISC made a detailed presentation.

The EMA plans to consult a wide range of stakeholders to understand their data standardization needs related to the submission, receipt, use, and re-use of scientific data at each stage of the medicinal product lifecycle.

ALTASCIENCES' CAPABILITIES

We ensure that our team is always up to date with the latest trends and guidelines in regard to CDISC standards. We utilize a suite of proprietary validation checks, as well as Pinnacle 21 Enterprise version technology, to load, review, and validate SDTM and ADaM data and Define.xml files. Our CDISC experts are available on a per-project or full-time equivalent (FTE) basis, according to your needs.

Our team has converted data from clinical trials, in diverse therapeutic areas, to CDISC-compliant SDTM and ADaM datasets, as well as designed CDASH-compliant forms and databases for streamlined data management and reporting.

To ensure your standards are successfully adopted in every phase of development, we provide:

- CDISC-compliant SDTM datasets for your legacy study data and the migration to submission-ready format of data from ongoing trials
- Clinical Data Acquisition Standards Harmonization (CDASH)-compliant CRFs for paper-based clinical trials and eCRFs for electronic data capture (EDC) clinical trials, in their proper design format
- Analysis Data Model (ADaM), CDISC-compliant datasets for statistical analysis
- Additional CDISC services
- Gap analysis of study documents and data
- CDASH annotated case report forms
- SDTM annotated case report forms
- Trial design compliant datasets
- Generation of Define.xml and Define.pdf files
- Comprehensive quality control measures
- Strategic consulting

REFERENCE

1 CDISC website, https://www.cdisc.org/ accessed April 28, 2022.

2 Importance of Data Standards in the European Health Data Space. Presentation by CDISC to the EMA. https://www.ema.europa.eu/documents/presentation/presentation-importance-data-standards-european-health-data-space-l-kustra-mano-dg-sante_en.pdf accessed April 28, 2022

3 Regulatory Affairs Professionals Society (RAPS). https://www.raps.org/regulatory-focus%E2%84%A2/news-articles/2017/3/505(b)(2)-approval-pathway-not-necessarily-shorter-approval-times#:~:text=The%20Tufts%20report%20offers%20similar,days%20for%20expedited%2Dreview%20drugs Accessed April 28, 2022.

ALTASCIENCES' RESOURCES

Webpages

E-bulletin, Navigating SEND, includes links to several related resources

Research Support Services, CDISC

Research Support Service, Biostatistics

Fact Sheet

SEND

ADDITIONAL INDUSTRY RESOURCES

FDA Study Data Standards Resources

ABOUT ALTASCIENCES

Altasciences is an integrated drug development solution company offering pharmaceutical and biotechnology companies a proven, flexible approach to preclinical and clinical pharmacology studies, including formulation, manufacturing, and analytical services. For over 25 years, Altasciences has been partnering with sponsors to help support educated, faster, and more complete early drug development decisions. Altasciences' integrated, full-service solutions include preclinical safety testing, clinical pharmacology and proof of concept, bioanalysis, program management, medical writing, biostatistics, clinical monitoring, and data management, all customizable to specific sponsor requirements. Altasciences helps sponsors get better drugs to the people who need them, faster.