A TRITICON KNOWLEDGEWORLD GUIDE · FREE

12

CDISC terms you need to understand

Clinical data standards arrive as a wall of acronyms, and most explanations of them are written for people who already know. Here are the twelve terms worth understanding before your next standardised study — and the traps hiding inside several of them.

MODULE A07 · INTRODUCTION TO CDISC

FIRST EDITION · V0.1

TriTiCon

THE KNOWLEDGE SHARING COMPANY

A SHORT FIELD GUIDE · VOCABULARY

The hard part of CDISC is rarely the standards. It is knowing which question each one answers.

CDISC is not one standard. It is a family of them, built by different working groups over nearly three decades, and each sits at a different point in the life of clinical data. Learning them as a list of acronyms is the slowest possible route, because the list tells you nothing about why any of them exists.

This guide takes twelve terms and places each one on the journey data actually takes: from a question on a form, through organisation and analysis, to a submission a regulator can read. Each entry is an honest working definition — what the term means, why it is there, and where it tends to trip teams up. Several of these have a trap inside them that catches experienced people, not just newcomers; those are flagged as we go.

Drawn from TriTiCon's course module A07 — Introduction to CDISC. A pointer at the end tells you where to find the full course.

THE TWELVE TERMS

1 CDISC
FOUNDATIONS

2 CDASH
COLLECT

3 SDTM
TABULATE

4 Domain
TABULATE

5 Core designation
TABULATE

6 USUBJID
TABULATE

7 Controlled Terminology
ACROSS THE CHAIN

8 Annotated CRF
COLLECT TO TABULATE

9 ADaM
ANALYSE

10 Traceability
ANALYSE

11 Define-XML
SUBMIT

12 Conformance
SUBMIT

THE TWELVE TERMS · FIELD GUIDE

FOUNDATIONS

CDISC

1

The organisation, not the standard.

The Clinical Data Interchange Standards Consortium is a global non-profit, founded in 1997, that builds and maintains clinical data standards. The word consortium is doing real work: the standards are developed collaboratively by sponsors, CROs, vendors, academics and regulators, not handed down by any one of them. CDISC does not sell the standards and does not police their use — it stewards them, maintaining and versioning them so they stay coherent. The standards themselves are free of charge, though most downloads require an account.

COLLECT

CDASH

2

How the question gets asked in the first place.

Clinical Data Acquisition Standards Harmonization governs data collection: which fields to put on a case report form, what to call them, how to phrase them. Its defining feature is that it was designed backwards from what comes next — collect in CDASH shape, and the move into tabulation is close to mechanical. Unlike the standards further down the chain, CDASH is not itself a submission requirement, which is why adoption varies: widely used, but with real differences in how strictly teams follow it.

TABULATE

SDTM

3

The structure a reviewer expects to see.

The Study Data Tabulation Model organises collected data into a standard shape for review. It does not change the data or analyse it; it puts it where a reviewer knows to look. This is the standard that makes submitted data comparable across studies, sponsors and decades, and it is the one with the least room for local interpretation. The common misreading is to treat SDTM as an analysis format. It is not — it is organised for review, not arranged for analysis, which is precisely why ADaM exists.

TABULATE

Domain

4

One table, one kind of information.

An SDTM domain is a table holding one type of data, identified by a two-letter code: DM for demographics, AE for adverse events, VS for vital signs, EX for exposure, and so on. Domains belong to classes — events, interventions, findings, special purpose — and the class determines the shape of the table. One record per subject in DM; one record per event in AE; one record per measurement per visit in VS. Recognising the class tells you what a domain's rows mean before you have read a single value.

TABULATE

Core designation

5

Required, Expected, Permissible — and not what they sound like.

Every SDTM variable carries a core designation, and this is the term most often misread. Required means the variable must be present and must never be null. Expected means it must be present in the dataset, but individual records may be null — a genuinely different thing. Permissible means include it if you collected it, and omit it entirely if you did not. The trap is reading these as mandatory, optional and forbidden. Omitting a Permissible variable is correct behaviour; populating an Expected variable with nulls is not the same as leaving it out.

TABULATE

USUBJID

6

The value that holds the whole submission together.

The unique subject identifier appears in every domain and identifies one participant unambiguously across all of them. It is what lets a reviewer take an adverse event and know which patient it belongs to, what treatment they received, and what their laboratory values were doing that week. It must be unique across the entire study, not merely within a site — which is why it is normally built by concatenating study, site and subject number rather than reusing the number the site writes on the form.

ACROSS THE CHAIN

Controlled Terminology

7

Fixed vocabulary, on a release cycle.

Where a field takes a value from a set, Controlled Terminology fixes exactly which values are allowed. It is maintained with the US National Cancer Institute and published free of charge. Each term has a submission value — the exact text that goes into the dataset — and a code identifying the concept. This is what stops one study recording 'mild' while another records 'Grade 1' for the same observation. Two things catch people: terminology is released on a regular cycle, so the version you work to is a deliberate decision rather than a default; and a codelist marked not extensible means exactly that — you may not add values to it.

COLLECT TO TABULATE

Annotated CRF

8

The bridge, and a deliverable in its own right.

The annotated case report form — the aCRF — is your blank collection form marked up to show where every field lands in SDTM, naming the target variable and the codelist that governs it. It answers the question a reviewer asks first: where did this value come from? Two things are commonly missed. It is built from the blank form, not a completed one, because it documents the design rather than a patient. And it is not internal documentation — it is part of what you submit.

ANALYSE

ADaM

9

Data arranged for the analysis you are actually going to run.

The Analysis Data Model takes SDTM and produces analysis-ready datasets, with derived variables, groupings and flags already built in, so a statistical programmer is not re-deriving the same logic in every output. ADSL, the subject-level dataset, is the backbone: one row per patient carrying treatment, demographics and population flags, which every other analysis dataset links back to. That link is how an adverse event record comes to know which treatment group it belongs to — the information is not in the SDTM event record at all.

ANALYSE

Traceability

10

The property that makes any of this worth doing.

Traceability means any value in a result can be followed back through the analysis dataset, to the tabulation dataset, to the field on the form where it was recorded. It is ADaM's defining principle and the reason derived variables must have documented rules rather than just plausible values. Treat it as a live test rather than a claim: take a number from a key output, and walk it back. If you cannot, a reviewer will not be able to either — and that is the point at which standardised data stops being trustworthy, however conformant it looks.

SUBMIT

Define-XML

11

The data about your data.

Define-XML is a standardised, machine-readable description of your datasets: which datasets exist, which variables they contain, where each value came from, and how anything derived was derived. It is built on ODM, CDISC's underlying exchange standard, and it travels with the submission as the data definition document. It has two faces, which is worth seeing once: opened in a browser with its stylesheet it is a readable set of tables, and opened in a text editor it is XML. A reviewer's guide normally travels alongside it, explaining in plain language what the metadata cannot.

SUBMIT

Conformance

12

Following the rules, and explaining where you do not.

Conformance checking verifies that datasets follow the rules of the standards. Findings come with a severity: errors must be resolved or formally explained before submission; warnings may be legitimate but need a considered answer; notices are informational. Pinnacle 21, now part of Certara, has long been the de-facto industry checker, and CDISC's open-source CORE engine is emerging alongside it. Two habits matter more than the tool. Run checks continuously, because an issue found during collection can be fixed at source while the same issue found near submission usually cannot. And remember a clean report is not the goal — trustworthy data is. Some findings should be explained rather than engineered away.

ONE MORE THING WORTH KNOWING

None of these standards is a fixed target. Versions change, terminology is re-released, and which version applies to your study depends on the agency, the submission type and when the study started. Pin that decision down at start-up and record it — it is far cheaper than discovering a mismatch during submission preparation.

IF THIS WAS USEFUL · THERE'S MORE

Twelve definitions in, one good question out: could you trace it back?

This guide is a teaser for module A07 of TriTiCon Knowledgeworld — a structured introduction to CDISC that maps the standards onto the journey clinical data actually takes.

The full course

Five chapters: the case for standards; the standards landscape and the value chain; data in, from CDASH to SDTM; data out, from ADaM to submission; and a recap with where CDISC is heading. With a complete worked example you can open and inspect.

Or a conversation

If you are standing up CDISC capability, preparing a first standardised submission, or trying to work out what you actually need — sponsor, CRO or vendor, it doesn't matter — we are happy to talk. Enablement is one of TriTiCon's core values; we'd rather help you do it well than do it for you.

TriTiCon — The Knowledge Sharing Company

www.TriTiCon.com · info@TriTiCon.com · +45 52 19 34 99

V0.1 · FIRST EDITION · FREE TO SHARE WITH ATTRIBUTION

Disclaimer: this guide is informal and must not be seen as a reference for authorities' interpretation of regulatory guidelines or laws. Descriptions reflect common practice and the views of the author; specific regulatory positions, standard versions and licence terms should always be confirmed with the relevant stewards and authorities for your trial. © TriTiCon ApS. All rights reserved.