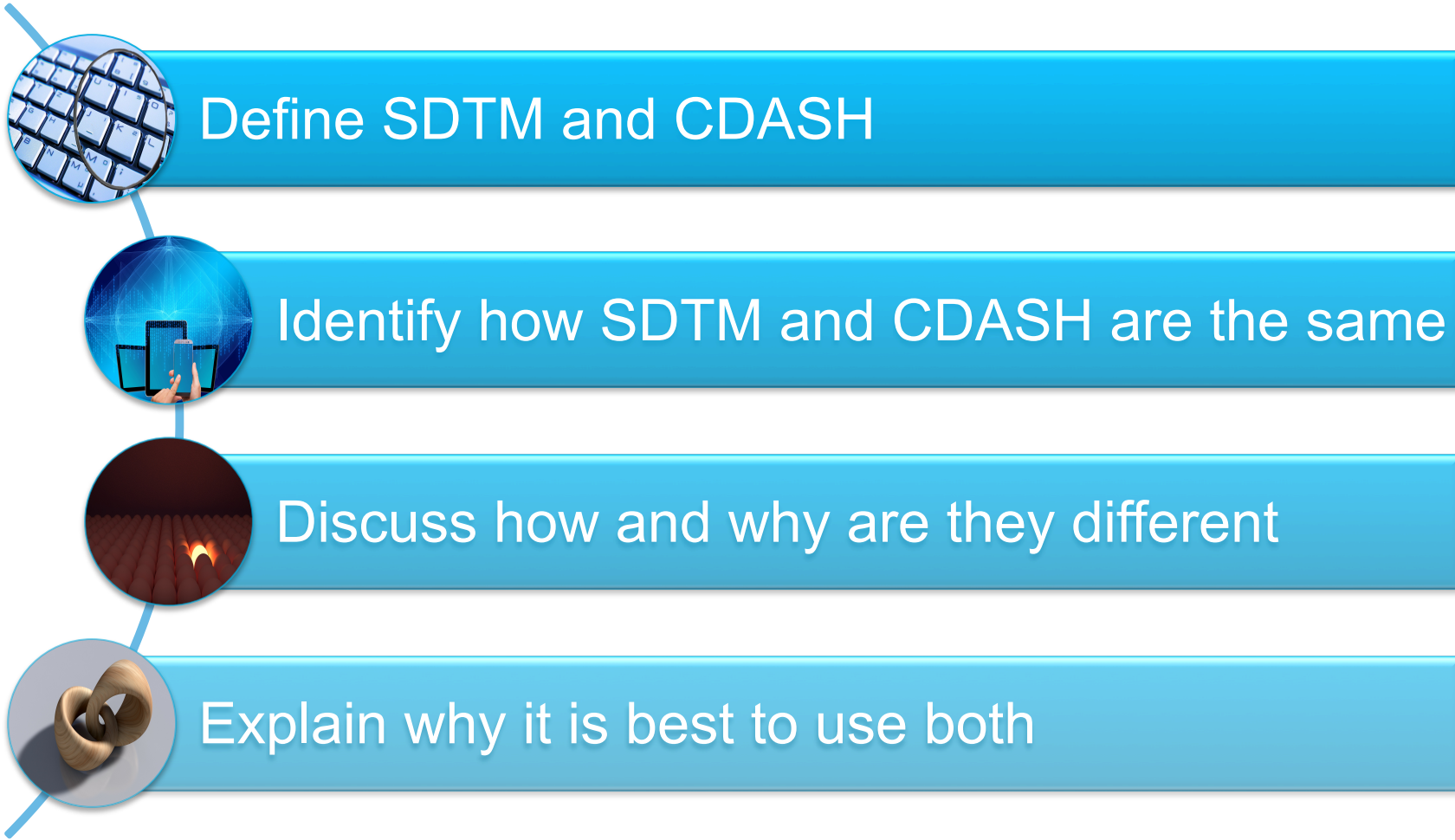


CDASH and SDTM: Why do We Need Both Standards?

Shannon Labout, Interim Chief Standards Officer, CDISC
Kit Howard, Director Education, CDISC

Strength through Collaboration

Learning Outcomes



Section 1: Define SDTM and CDASH

SDTM/IG

Defines predictable representation of collected data

- Logically groups data into topics
- Naming conventions, data types, controlled terminology
- All clean collected data plus some derived

Supports

- Data aggregation for submissions, data warehouses, registries
- Regulatory review, IRB reporting, safety surveillance
- Hypothesis generation

CDASH/IG

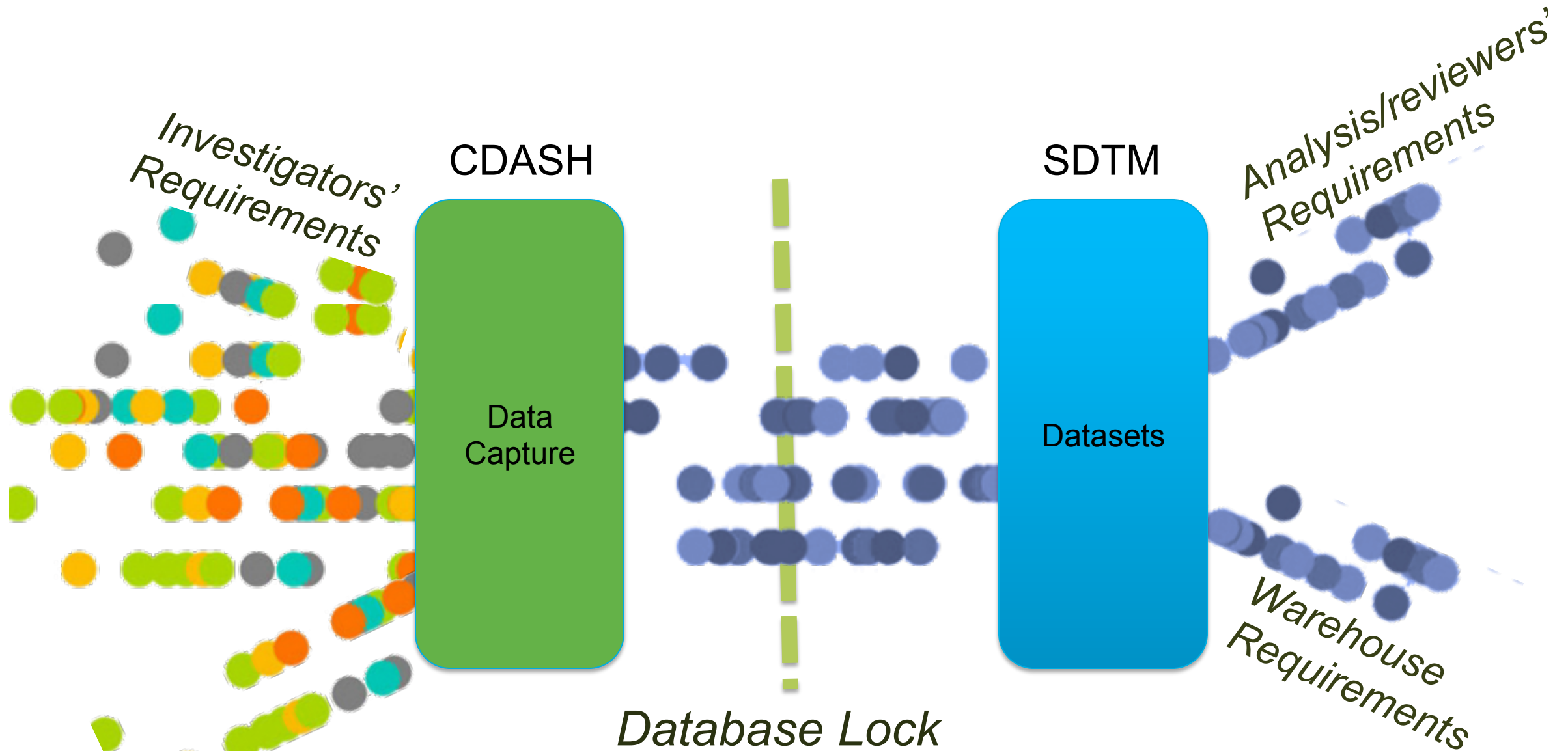
Data Collection & Management

- Consistent CRFs – questions, answers, assumptions
- Ease of use for investigative sites
- Data quality rules that enable SDTM's quality assumptions

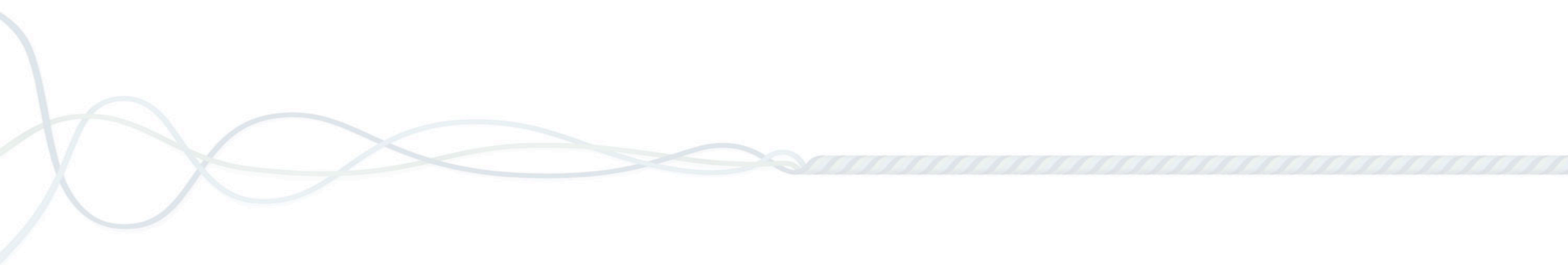
Supports

- Beginning with the end in mind
- Smooth flow of data from collection to submission
- Data integrity and traceability back to the source

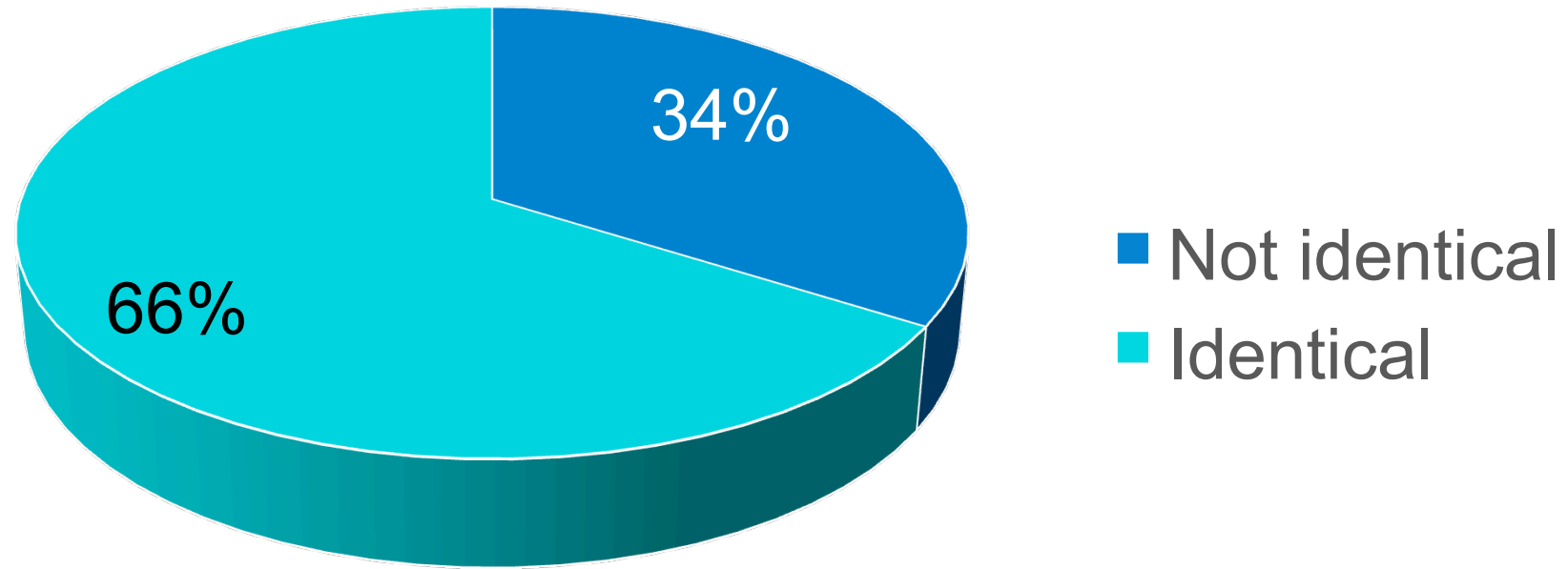
Data Flow



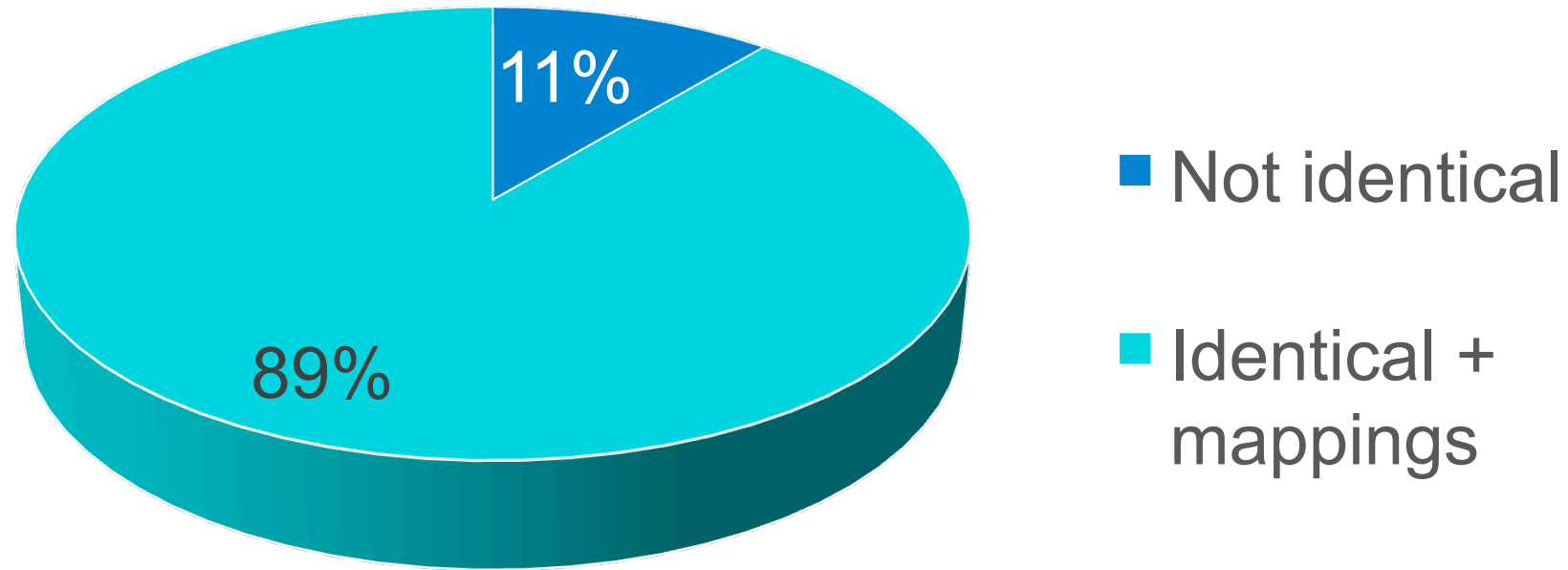
Section 2: Similarities



Variable Overlap: Exact Match



Variable Overlap: Exact Match + Standard Mappings



More Similarities



Same Controlled Terminology

Same data assumptions

- Only include IE criteria not met
- DA is for tracking controlled materials, not for treatment compliance

Same organization of data into domains

Section 3: Differences

Philosophy drives design

Subject ID: 1 | 0 | 2 | 3 SUBJID AE

Were Any Adverse Events Experienced?

☒ No AEYN Not Submitted

☐ Yes - (please complete the form)

Adverse Event - (please use precise medical terminology)

AETERM

Start Date and Time

Day Month Year

AESTDAT
AESTDTC

Missing Data

SDTM

Show me the data, not
lack of data

CDASH

Absence of evidence is
not evidence of
absence: must check
that missing data is
missing

Rationale

SDTM assumes that if there is no record then nothing happened. This only works if it was checked in data capture, which requires a question and record (e.g., Were there any AEs?)

SDTM

USUBJID	AETERM	AESTDTC	AEENDTC
2006-34-1022	HEADACHE	2006-10-14	2006-10-20
2006-34-1022	RASH	2006-10-14	2006-10-21
2006-34-1024	TIRED	2006-09-24	2006-09
2006-34-1024	HEADACHE	2006-09-24	2006-09-24

Human vs Machine Readable Data

SDTM

USUBJID	HOTERM	HOSTDTC	HODUR
ABC1201	Hospitalization	2011-08-06	P6H

CDASH

SUBJID	HOTERM	HOSTDAT	HOCDUR	HOCDURU
1201	Hospitalization	06-AUG-2011	6	HOURS

HOSPITALIZATION

HOTERM

HO

Was the subject hospitalized during this study?

☐ No

HOYN

Not Submitted

☒ Yes If "Yes," please complete the following:

Admission Date:

06

AUG

2011

Day

Month

Year

HOSTDAT

HOSTDTC

Duration of Hospitalization:

HOCDUR

6

HOCDURU

Number

☒ Hours

☐ Days

HODUR

SDTM

Machine-readable:
Dates/Times: ISO
8601, 1 variable,
YYYY-MM-
DDThh:mm:ss
Duration: P1M3D

CDASH

Human-readable:
Dates/Times: 2 or
more variables, DD-
MM-YYYY,
HH:MM:SS
Duration: 1 month,
3 days

Rationale

SDTM machine-readable formats for variables such as dates are good for data reusability but are not user-friendly for data capture. There is more chance for error when people record data in unfamiliar formats

Data Organization

SDTM

CDASH

Data must be organized into datasets by domain

Data on each CRF is driven by what is captured together, not necessarily by domain

Rationale

Each CRF should have the data that makes sense to collect together. Per CDASH, this can mix domains if standard variable names are used, and in the end the data appear in the right domains.

In SDTM, data **MUST** appear only in the correct domain.

Demographics

Birth Date

1	3	D	E	C	1	9	4	8
Day		Month			Year			

Domain = DM

BRTHDTC

BRTHDAT

Sex

☐ Male

☐ Female

SEX

Race

☐ White

☐ Black

☐ Other, (please specify) _____

RACE

RACE1

RACE2

RACEOTH

Reproductive Status (if female)

☒ Premenarchal

☐ Postmenarchal

☐ Postmenopausal

Domain = RP

RPORRES when RPTSTCD = MENOSTAT or (CDASH) MENOSTAT

DM CDASH

RP CDASH

DM SDTM

USUBJID	SUBJID	BRTHDTC	SEX	RACE
ABC1201	01	1948-12-13	M	WHITE
ABC1202	02	1955-03-22	M	BLACK

RP SDTM

USUBJID	RPTSTCD	RPORRES	RPSTRESC
ABC1201	MENOSTAT	Premenarchal	Premenarchal
ABC1201	NUMLIV	2	2

EDC CRF

Hamilton Depression Scale 21 Items (HAM-D21)

1. DEPRESSED MOOD (<i>Sadness, hopeless, help-</i>) ☐ Absent ☐ These feeling states indicated only on questioning ☐ These feeling states spontaneously reported verbally	4. INSOMNIA EARLY ☐ No difficulty falling asleep ☐ Complains of occasionally difficulty falling asleep ☐ Complains of nightly difficulty falling asleep
5. INSOMNIA MIDDLE ☐ No difficulty ☐ Complains of being restless and disturbed ☐ Waking during night - any getting out of bed rates 2	

RSORRES when RSTESTCD = HAMD101 or (CDASH) HAMD101

CDASH

SUBJID	HAMD101	HAMD104	HAMD105
1	Absent	No difficulty falling asleep	No difficulty

SDTM

USUBJ ID	RSTESTCD	RSTEST	RSORRES	RSSTRESC
2324-P0001	HAMD101	HAMD1-Depressed Mood	Absent	0
2324-P0001	HAMD104	HAMD1-Insomnia Early - Early Night	No difficulty falling asleep	0

Horizontal vs Vertical Data

SDTM

Findings data must be in a normalized or vertical structure; answers are already known

CDASH

Findings data may have to be horizontal for each test to have its own code list; SDTM CT is used for variable names & CRF prompts

Rationale

In normalized data, each test is on a different record and may need different controlled terms, called “value-level metadata,” which cannot be modeled in most data capture systems; some EDC systems can’t handle normalized data at all

Metadata Content

CDASH Metadata

Data Collection Scenario	CDASHIG Variable	Question Text	Prompt	Data Type	CRF Completion Instructions
--------------------------	------------------	---------------	--------	-----------	-----------------------------

SDTM Metadata

Variable Name	Variable Label	Type	Controlled Terms Codelist or Format	Role	CDISC Notes
---------------	----------------	------	-------------------------------------	------	-------------

SDTM

Metadata includes variable labels and roles

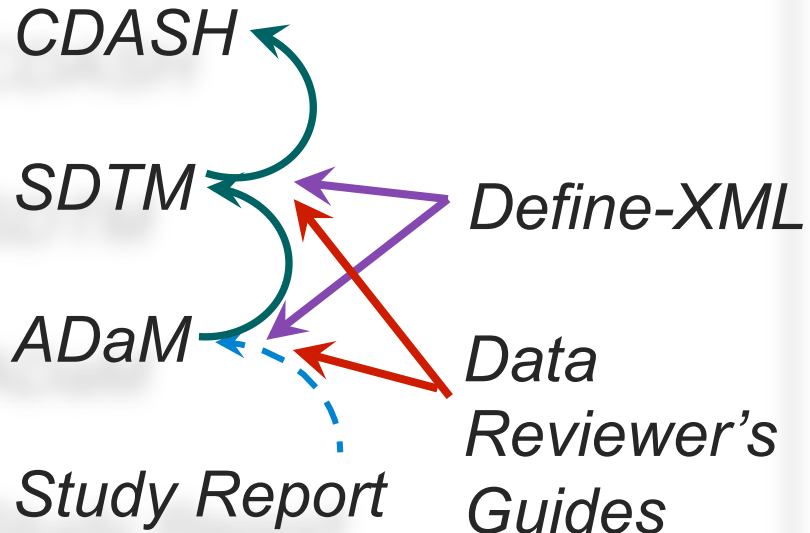
CDASH

Metadata includes question text/prompt, CRF completion instructions

Rationale

SDTM focuses on tabulating data. CDASH addresses data capture needs, designed to ensure clear questions that produce consistent answers. Because they have different purposes, some of their metadata differs

Traceability



SDTM

Provides traceability
back to an SDTM-
variable annotated
CRF

CDASH

Structure can provide
link to source data at
the site

Rationale

CDASH uses the same variable names and assumptions
as SDTM except where data capture needs require a
different approach. CDASH also provides mapping
algorithms to SDTM

Unavailable Variables

SDTM

CDASH

Exposure as Collected EC CRF

Exposure EC			
Line #	Dose	Dose Frequency	Dosing Time Period
1	EC DSTXT Capsules EC DOSU	☐ BID ☐ TID EC DOSFRQ	Start Date EC STDAT Day Month Year End Date EC ENDAT Day Month Year

Some variables can only be derived after studies are complete, e.g., USUBJID, STRESC/STRESN, EXDOSE

Some variables have meaning only within the context of the study, e.g., SUBJID, ECDOSE

Rationale

USUBJID is unique across a submission;
STRESC & STRESN standardize findings results across the submission;
In doubleblind studies, treatment dose is unknown until the blind is broken;
SDTM primary focus is pooling data for submission;
CDASH primary focus is individual studies

USUBJID	EXTRT	EXDOSE	EXDOSU	EXDOSFRQ	EXSTDTC	EXENDTC
ABC1201	BestDrug	350	mg	BID	2011-08-06	2011-09-01

Some Benefits of Using Both

Ensure asking same question using the same answer lists as analyzing

Optimize site data requirements *and* structure for transmission & analysis

Supports traceability back through data collection

Minimizes programming and validation resources and increases quality when transferring data from capture to tabulation

Can address the points in all the previous slides

Helps to “future-proof” the data for warehousing



Unlocking Cures

