I-Mind

Turning Your Data into Knowledge

Challenges with the interpretation of CDISC

-

Who can we trust?

Linda Palm Simonsson

linda.simonsson@i-mind.se

CD01 – #PhUSE14 - London 2014-10-13

Abstract

Many smaller companies have none or very little knowledge of biometrics and turns to one or several CROs to help them conduct their trials or a submission. A small company expects the CRO to be the expert and since many small companies lacks a biometric department they also lack internal resources, routines and expertise to do a qualified QC of the SDTM and the ADaM packages received from a CRO. Sometimes the difficulties for the sponsor begins already at understanding what's included in the proposal vs what's needed according to FDA or other agency. Other difficulties may be inconsistency between e.g. SDTM IG and FDAs requirements.

This presentation will address some errors, misunderstandings and difficulties seen in the interpretation of different documents and showing some inconsistencies between different sources.

Background

- Sponsors expects the CROs to be experts
  - and the CROs are selling themselves as experts
- Lack transparency in proposals of what's included vs needed
  - Does the proposal include what the Sponsor need according to agencies requests?
- Inconsistency between CDISC and FDAs recommendations
  - Makes it hard to know what to do or what to expect
- Programmers with none or little CDISC experience/training
  - Shouldn’t we expect at least one experienced person in each project?
- Should the Sponsor need to do a detailed QC of the deliveries?

I-Mind

Turning Your Data into Knowledge

Empty Datasets

Ambiguous documentation of how we should document empty dataset

Empty Datasets

All documentation seems to agree that empty datasets should not be submitted.

But how should we document that we didn't submit an expected dataset?

Empty Datasets

SDTM IG – Section 3.2

In the event that no records are present in a dataset, the empty dataset should not be submitted and **should not be described in the define.xml** document. The annotated CRF will show the data that would have been submitted had data been received; **it need not be re-annotated to indicate that no records exist.**

Empty Datasets

FDA - Supplemental Information for Planning a CDISC Formatted Submission

If no data is collected for a specific domain, **annotate on the CRF and the define.xml** but do NOT submit an empty dataset (Update 8/12/10)

Empty Datasets

FDA & PhUSE – Study Data Reviewer's Guide

Were any domains planned, but not submitted because no data were collected?

Yes

If yes, list domains not submitted:

IE – All subjects met inclusion/exclusion criteria.
SUPPDM – For all subjects, race was one of those pre-specified on the CRF. Specification for "other" race was not needed.

I-Mind

Turning Your Data into Knowledge

ARM and ARMCD

Ambiguous documentation

ARM and ARMCD

The variables ARM and ARMCD have a very central role in SDTM, and the SDTM IG has already a rule for how screening failures should be handled.

But in February 2014 FDA released a new draft guideline (which aren't for implementation – Yet (!)) that contradicts the SDTM IG. So please watch out for the final version before you start to make any changes

SDTM IG v3.2 (and previous versions)

Data for screen failure subjects, if submitted, should be included in the Demographics dataset, with ARMCD = "SCRNFAIL" and ARM = "Screen Failure".

STUDY DATA TECHNICAL CONFORMANCE GUIDE Technical Specifications Document (Draft feb 2014)

Screen failures, when provided, should be included as a record in DM with the ARM field left blank.

I-Mind

Turning Your Data into Knowledge

Trial Design

TS, TA, TE, TI, TV (and TD)

Trial Design

- What different vendors includes:
  - None
  - Not specified in proposal, but created all but the Trial Summary (TS)
  - Clearly states that TS is out of scope
  - All

Trial Design

FDA - Supplemental Information for Planning a CDISC Formatted Submission

Trial Design datasets provide a standard way to describe the planned conduct of a clinical trial. **At a minimum, the Trial Summary (TS) domain should be submitted whenever possible**.

Also OpenCDISC gives an ERROR if TS is not provided.

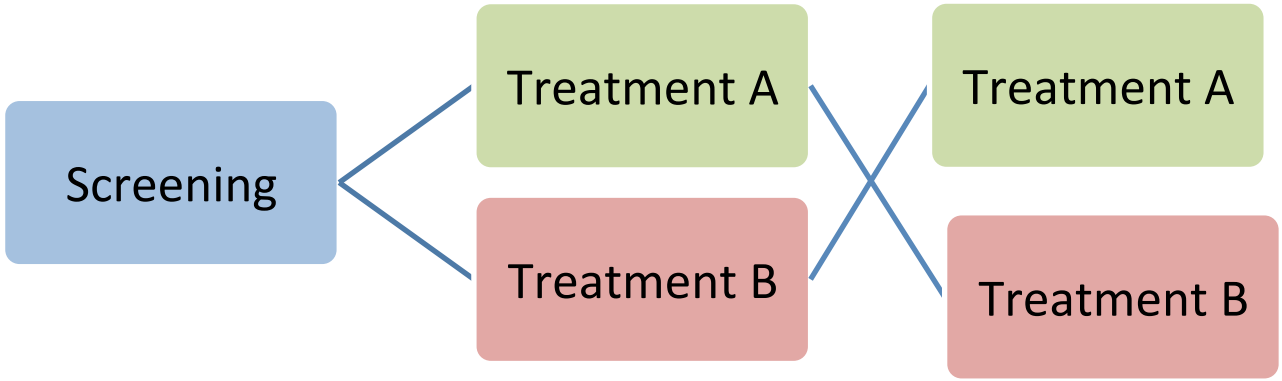
I-Mind

Turning Your Data into Knowledge

Subject Elements

SE - Subject Elements

A sponsor needed to convert a legacy study to SDTM that included two periods (cross-over)



The proposal stated

- SE is out of scope

SE - Subject Elements

FDA - Supplemental Information for Planning a CDISC Formatted Submission

Subject Elements (SE) and Subject Visits (SV) describe actual elements and visits for each subject. These datasets **should be submitted whenever possible**. This is **especially important when the trial includes multiple treatment periods**.

I-Mind

Turning Your Data into Knowledge

Creating a New Domain

Creating a New Domain

- #1 Created a custom made Finding domain for "Cardiac History"

Cardiac History

- High blood pressure? ☐ Yes ☐ No
- Chest pain on exertion? ☐ Yes ☐ No
- Previous heart attack? ☐ Yes ☐ No ____/_____/_______
  dd mmm yyyy
- Valvular heart disease? ☐ Yes ☐ No ____/_____/_______
  dd mmm yyyy
- Heart Surgery? ☐ Yes ☐ No ____/_____/_______
  dd mmm yyyy

I-Mind
Turning Your Data into Knowledge

Creating a New Domain

- #2 Another CRO followed IG 3.1.2 and had to create SDTM for an Oncology study.
  (At this time point IG 3.1.3 was already finalized)
  - They created a custom made domain starting with Y for the tumor result data

Creating a New Domain

SDTM IG – Section 2.6

Confirm that none of the existing published domains will fit the need. **A custom domain may only be created if the data are different in nature and do not fit into an existing published domain. Group and separate data within the domain using --CAT, --SCAT, --METHOD, --SPEC, --LOC, etc.** as appropriate.

Creating a New Domain

CDER Common Data Standards Issues Document

- Prior to creating a custom domain, **sponsors should confirm that the data do not fit an existing domain and also check the CDISC website for domains added after the most recent published implementation guide.**

Creating a New Domain

Two errors in #1:

1. Mistook “Cardiac History” for data that will fit into a Finding domain
2. Didn’t recognize it as Medical History of Special Interest

Should have been in Medical History (MH) with
MHCAT=“Cardiac History”,
while the ordinary Medical History could have had MHCAT=“Relevant Medical History”

Use variables like MHTERM, MHPRESP, MHOCCUR and MHSTDTC

Creating a New Domain

One error in #2:

1.They did not check the CDISC website for domains added after IG 3.1.2

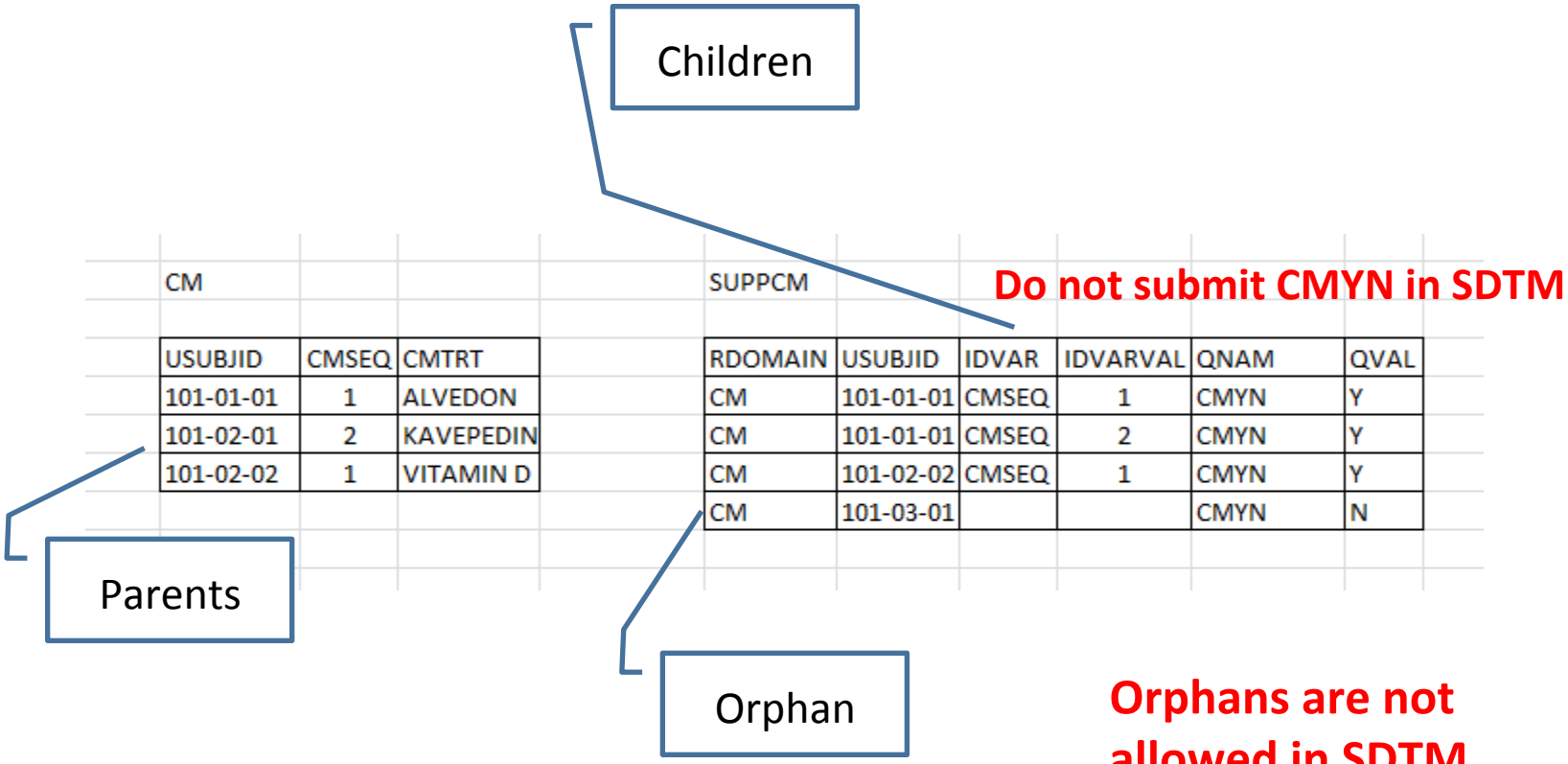
If a new domain has been defined in a later version (even if the later version is only in draft), you should use the later version instead of creating your own.

I-Mind

Turning Your Data into Knowledge

Orphans

Orphans



USUBJID	CMSEQ	CMTRT
101-01-01	1	ALVEDON
101-02-01	2	KAVEPEDIN
101-02-02	1	VITAMIN D

RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QVAL
CM	101-01-01	CMSEQ	1	CMYN	Y
CM	101-01-01	CMSEQ	2	CMYN	Y
CM	101-02-02	CMSEQ	1	CMYN	Y
CM	101-03-01			CMYN	N

Orphans

One mapping specification had the CDASH-variable “Any Concomitant Medications (Y/N)?” in SUPPCM

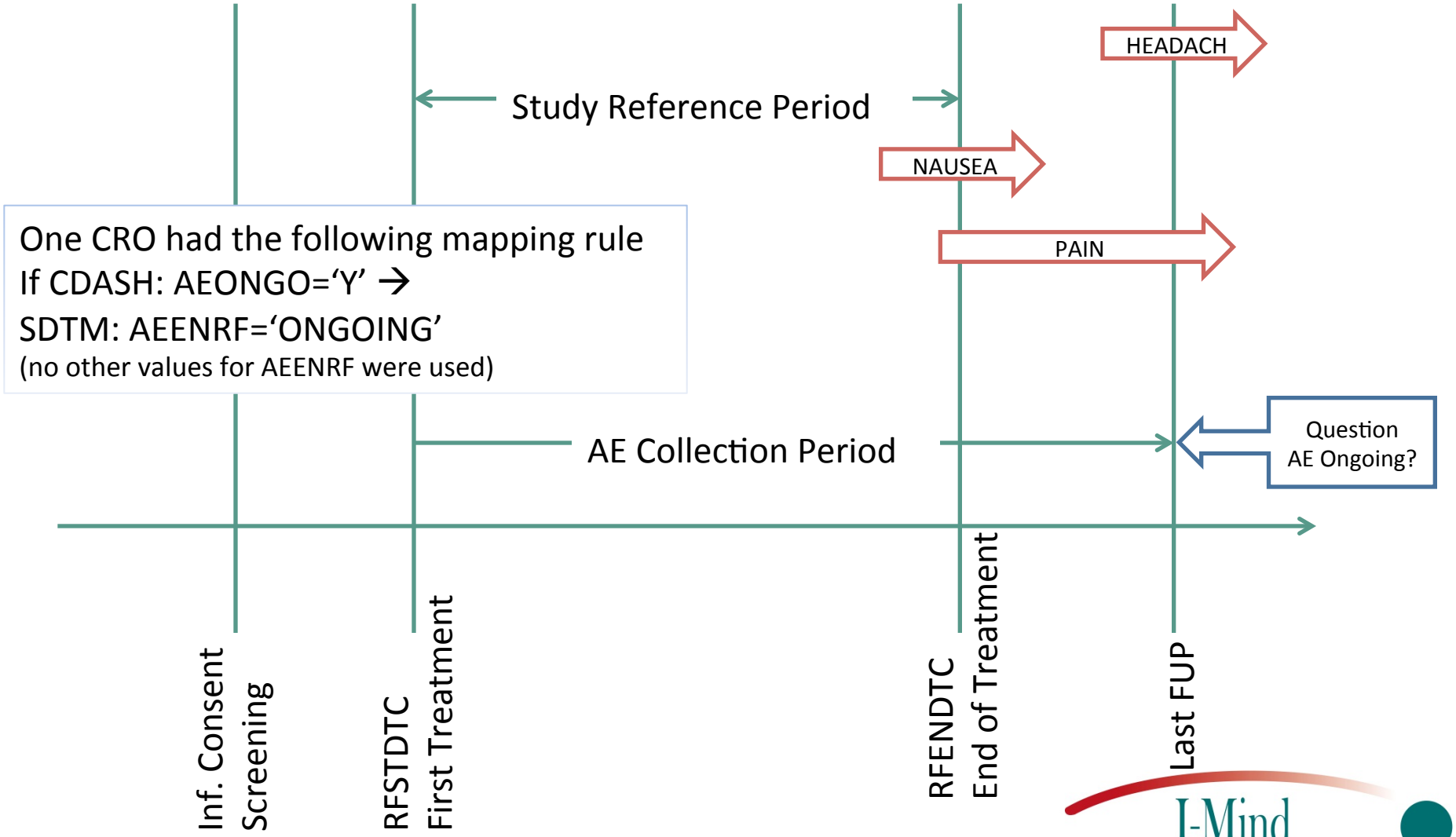
- → Every time they saved the answer N, they created an orphan
- → All entries in a SUPP-- must have linked data in the parent domain

I-Mind

Turning Your Data into Knowledge

Sponsor-Defined Reference Period

Sponsor-Defined Reference Period



I-Mind
Turning Your Data into Knowledge

Sponsor-Defined Reference Period

- AEENRF should NOT have been used.
- Better to use AEENRTPT and AEENTPT
  - AEENTPT='Last Follow-Up'
  - AEENRTPT=
    - 'BEFORE' (if AEONGO='N' and AENEDTC<Last Follow-Up)
    - 'COINCIDENT' (if AEONGO='N' and AENEDTC=Last Follow-Up)
    - 'ONGOING' (if AEONGO='Y')

Use of --PRESP without --OCCUR

Use of --PRESP without --OCCUR

Used pre-specified variable in Substance Use without specifying if it occurred or not in SU

USUBJID	SUTRT	SUPRESP
101-01	CIGARETTES	Y
101-01	PIPE	Y
101-02	CIGARETTES	Y
101-02	PIPE	Y
101-03	CIGARETTES	Y
101-03	PIPE	Y

SUPPSU.QNAM	SUPPSU.QVAL
USER	No
USER	Current User
USER	Current User
USER	No
USER	Ex-Smoker
USER	No

Use of --PRESP without --OCCUR

SDTM IG – Section 4.1.5.7

- –PRESP: should only be used when the topic variable values come from a pre-specified list.
- –OCCUR: may be omitted from the dataset if no topic-variable values were pre-specified.

Use of --PRESP without --OCCUR

SUPPSU is ok, but the variable(s) SUOCCUR, (SUENTPT and SUENRTPT) should have been added

```
if  USER in ('Current User', 'Ex-Smoker') then do;
    SUOCCUR='Y';
    SUENTPT='SCREENING';
    if USER='Current User' then SUENRTPT='ONGOING';
    else if USER='Ex-Smoker' then SUENRTPT='BEFORE';
end;
else if USER in ('No') then SUOCCUR='N';
```

USUBJID	SUTRT	SUPRESP
101-01	CIGARETTES	Y
101-01	PIPE	Y
101-02	CIGARETTES	Y
101-02	PIPE	Y
101-03	CIGARETTES	Y
101-03	PIPE	Y

SUPPSU.QNAM	SUPPSU.QVAL
USER	No
USER	Current User
USER	Current User
USER	No
USER	Ex-Smoker
USER	No

SUOCCUR	SUENTPT	SUENRTPT
N		
Y	SCREENING	ONGOING
Y	SCREENING	ONGOING
N		
Y	SCREENING	BEFORE
N		

I-Mind

Turning Your Data into Knowledge

Derive or Impute?

Derive or Impute?

- One CRO left AGE, --STRESC and –STRESN blank or copied e.g. VSORRES into VSSTRESC without converting from e.g. F to C for TEMP, with the motivation that all derivations/imputations should be left for ADaM

Derive or Impute?

- SDTM IG
  - Data stored in SDTM datasets include both raw (*as originally collected*) *and derived values* (e.g., *converted into standard units*, or computed on the basis of multiple values, such as an average).
  - When --ORRES is populated, --STRESC must also be populated, regardless of whether the data values are character or numeric.
  - AGE (Expected): Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC.
  - VSSTRESC (Expected): Contains the result value for all findings, copied or derived from VSORRES *in a standard format or standard units*.

Derive or Impute?

- CDER Common Data Standards Issues Document
  - *For a given test, all values of --STRESU should be the same*. In some cases --TESTCD may not be sufficient to uniquely identify a test.
  - *SDTM should not include any imputed data*. If there is a need for data imputation, this should occur in an analysis dataset, and the relevant supporting documentation to explain the imputation methods must be provided.

Derive or Impute?

- Derive
  - You derive the temperature from F to C like C=(F-32)*5/9
- Impute
  - When we only have partial dates we may need to impute the date with rules described in the SAP

I-Mind

Turning Your Data into Knowledge

Standard Questionnaires

Standard Questionnaires

Another common error is to handle questionnaires like this:

QSTEST	QSORRES	QSSTRESC	QSSTRESN
Global Improvement	No change	No change	4
Global Improvement	Much Improved	Much Improved	2
Global Improvement	Minimally Improved	Minimally Improved	3

When the correct way is like this:

QSTEST	QSORRES	QSSTRESC	QSSTRESN
Global Improvement	No change	4	4
Global Improvement	Much Improved	2	2
Global Improvement	Minimally Improved	3	3

--STRESN should only be populated if --STRESC contains a number (and only a number)

Standard Questionnaires

SDTM IG - Section 4.1.5.1.1

- In some cases where the code values in the codelist are statistically meaningful standardized values or scores, which are defined by sponsors or by valid methodologies such as SF36 questionnaires, the --ORRES variables will contain the decoded format, whereas, the --STRESC variables as well as the --STRESN variables will contain the standardized values or scores.

SDTM -> ADaM traceability

SDTM -> ADaM traceability

One ADaM mapping specification created ADSL.RACE like this:

```
If DM.RACE='OTHER' then
ADSL.RACE=propcase(strip(DM.RACE)||': '||SUPPDM.QVAL )
    (when SUPPDM.QNAM='RACEOTH')

Else ADSL.RACE=propcase(DM.RACE);
```

ADaM IG v1.0:

Any ADaM variable whose name is the same as an SDTM variable must be a copy of the SDTM variable, and its label, meaning, and values must not be modified.

- "same name, same meaning, same values."

The race of the subject is a required variable in ADSL. If the variable is not a copy of DM.RACE, then an additional differently named variable must be added.

I-Mind

Turning Your Data into Knowledge

Who can we trust?

Who can we trust?

- ? CDISC (SDTM IG etc)
- ? FDA and other agencies
- ? PhUSE
- ? OpenCDISC
- ? CROs
- ? Me
- ? You

What should be expected of a CRO?

- all required and highly recommended items should be included in the proposal
  - If not, clearly stated that it is expected in e.g. a submission
- OpenCDISC or equivalent must be used for validation before every delivery (both draft and final)

Conclusion

- Don’t be afraid to ask your vendor to give you a more detailed proposal of what’s included and if they have omitted anything that is required or recommended by e.g. FDA.
- Every company (sponsor and vendor) needs at least one in-house or contracted CDISC specialist
- It would be good if we could have a common place where we linked all information together (PhUSE WIKI??)

References

1. CDISC: SDTM IG 3.1.2, 3.1.3 and 3.2
2. CDISC: SDTM Terminology 2013-12-20
3. CDISC: ADaM IG 2.1
4. FDA: Study Data Specifications (SDS) v2.0
5. FDA: CDER Common Data Standards Issues Document v1.1
6. FDA: Supplemental Information for Planning a CDISC Formatted Submission
7. FDA: Study Data Technical Conformance Guide (Feb 2014 – Only for review)