CDISC Italian User Network 2020

Milan, Italy | 7 October 2020

Re-mastering the define-xml and its “brother” the Reviewer Guide

Presented by Angelo Tinazzi
Senior Director, Standards, Systems and CDISC Consulting - Statistical Programming

Cytel Inc. - **angelo.tinazzi@cytel.com**

7 October 2020

>20 Years of Experience

Passion for Standards

Member of CDISC-EU Committee

CDISC Authorized Instructor for ADaM

Agenda

1. Introduction
   - Quality in Data Submission
   - Intro to the Reviewer Guide and the define-xml
   - Initial (Final) Recommendations
   - Official and Unofficial References
2. (Unveiled?) Tips
3. Conclusions

Introduction

Quality in Data Submission

Intro to the Reviewer Guide and the define-xml

Initial (Final) Recommendations

Official and Unofficial References

Introduction

Quality in Data Submission

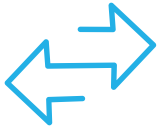
What do we mean by "Quality"

- Any "piece" submitted to HA should be of **Good Quality**
- Quality in the **Data**
- Quality in the **Results**
- (But also) Quality in the **Documentation**
  - **define-xml**
  - **Reviewer Guide**
  - aCRF
  - Any other "attached" submitted document, including scripts e.g. SAS code

*"The efficacy and safety of your drug are of course what matter, but **lack of traceability, poor or insufficient documentation might trigger questions and concerns from the reviewer.** You may think these are minor issues because they do not ultimately impact any results. However, **you are risking your credibility with the FDA reviewer, who may conclude that your package is not of good quality**"* A. Tinazzi "The CDISC Stupidario (the CDISC Nonsense)", CDISC-EU 2019

Introduction

Intro to the Reviewer Guide and the define-xml



Regulatory Interaction
They facilitate the communication with the reviewer

Reviewer Guide
Single point of orientation, nsdrg (SEND) csdrg (SDTM) adrg (ADaM)

define-xml
Set of machine readable Metadata, required by FDA and PMDA for SEND, SDTM and ADaM, Platform Independent and Vendor Neutral

Introduction

Initial (Final) Recommendations

Start with the « end » in Mind	• Do not wait the end to generate define-xml • The reviewer guide is a working document, it could be your programming « notebook »
define-xml can drive Automation	• Make sure your metadata are accurate e.g. reference the correct Ig/CT versions
Clarity of Explanations	• Reduce the risk to have questions back from the reviewer because something is unclear
Establish Conventions	• Naming conventions for code-list, derivations and comments • Standard wording for methods and comments
Review Process	• Make sure there is an internal review process, automatic and manual • Be pragmatic and Use Common Sense • Educate, there is "more" than the CDISC standards when it's time to submit to HA

Introduction

Intro to the Reviewer Guide and the define-xml

Official Standards

- define-xml 2.0 + Analysis Results Metadata (ARM) Specification 1.0
- define-xml 2.1 which includes ARM

Agency Recommended Templates

- PhUSE WG, Clinical Study Data Reviewer Guide Template and Guidance
- PhUSE WG, Analysis Data Reviewer Guide Template and Guidance

Good Recommendations

- PhUSE WG, "Best Practices for Documenting Dataset Metadata: Define-XML vs Reviewer's Guide"
- PhUSE WG, "Define-xml Version 2.0 Completion Guidelines"

(Unveiled?) Tips

1. When do I need to create ValueLevel Metadata
2. When do I need to assign / specify a codelist?
3. Use of subset codelist
4. When origin=CRF and you have to link to Multiple Pages
5. Origin=Assigned vs Origin=Derived
6. Reviewer Guide vs define-xml
7. SDTM Mapping Specifications are not needed in define-xml
8. Good vs Bad Computational Algorithms
9. Consistency between SDTM and ADaM define-xml

(Unveiled?) Tips

1. When do I need to create ValueLevel Metadata?

- The SDTM /ADAM models are highly normalized data structures

  e.g. Findings for SDTM or BDS for ADaM.

- As a result there are some cases where the **content of a column or variable cannot be unambiguously defined through Variables Metadata**

Value Level Metadata - VS [VSORRES]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
VSORRES	VSTESTCD EQ DIABP (Diastolic Blood Pressure)	integer	2		CRF Page 11	
VSORRES	VSTESTCD EQ SYSBP (Systolic Blood Pressure)	integer	3		CRF Page 11	
VSORRES	VSTESTCD EQ PULSE (Pulse Rate)	integer	2		CRF Page 11	
VSORRES	VSTESTCD EQ WEIGHT (Weight)	float	5.1		CRF Page 11	
VSORRES	VSTESTCD EQ HEIGHT (Height)	float	5.1		CRF Page 11	
VSORRES	VSTESTCD EQ FRMSIZE (Body Frame Size)	text	6	["LARGE", "MEDIUM", "SMALL"] <Size>	CRF Page 11	

(Unveiled?) Tips

1. When do I need to create ValueLevel Metadata?

- All SDTM Supplemental Qualifiers √

Value Level Metadata - SUPPCM [QVAL]

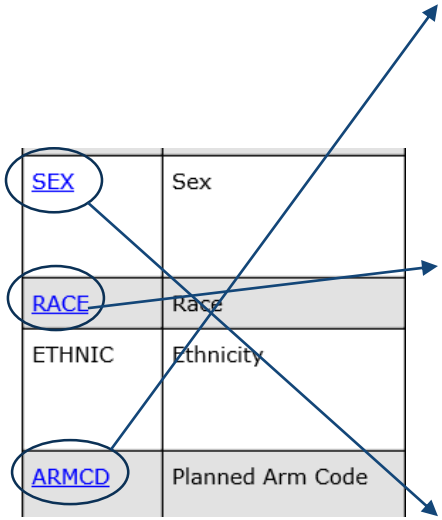
Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin
QVAL	QNAM = "APPETHIS" (History of appetite suppressant use)	text	18	History of Appetite suppresant use	CRF Page 13
QVAL	QNAM = "CATHLTYP" (Catheter Lumen Type)	text	12	["Double lumen" = "Double lumen", "Single lumen" = "Single lumen", "U" = "Unknown"] <Catheter Lumen Type>	CRF Page 65
QVAL	QNAM = "CATHTYPE" (Catheter Type)	text	8	Catheter Type	CRF Page 65
QVAL	QNAM = "CLOSEHUB" (Patient Using Closed Hub System?)	text	1	["N" = "No", "U" = "Unknown", "Y" = "Yes"] <No Yes Response>	CRF Page 65
QVAL	QNAM = "CMACN" (Medication Action)	text	11	["Change", "Discontinue", "Start"] <PAH Medication Action>	CRF Page 63

- Trial Summary (TS) √
- All findings? √ √ Inclusion / Exclusion Criteria (IE)? √
- All ADaM BDS? √ √

(Unveiled?) Tips

1. When do I need to create ValueLevel Metadata (cont)

- Is it really needed here?



Value Level Metadata - DM [ARMCD]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
ARMCD	ARMCD = "ENROLLED" (Enrolled)	text	200		Assigned	Set to ENROLLED when patient is enrolled into the study (no screen failure, eligible for the study)
ARMCD	ARMCD = "SCRNFAIL" (Screen Failure)	text	200		Derived	Set to SCRNFAIL when reason for discontinuation is "Protocol Deviation: Ineligible Patient" or "Duplicate data entry"

Value Level Metadata - DM [RACE]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
RACE	ETHNIC = "HISPANIC OR LATINO" (Hispanic or Latino)	text	200		Derived	RACE=OTHER if ETHNIC = HISPANIC OR LATINO
RACE	ETHNIC = "NULL"	text	200		CRF Page 4	

Value Level Metadata - DM [SEX]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
SEX	SEX = "F" (Female)	text	200		CRF Page 4	
SEX	SEX = "M" (Male)	text	200		CRF Page 4	
SEX	SEX = "U" (Unknown)	text	200		Derived	Set to "U" when SEX is not reported in the CRF

(Unveiled?) Tips

2. When do I need to assign / specify a codelist?

- Whenever in the Ig a variable has a **CDISC-CT associated**
- SDTM variables with **pre-printed code-list** in the CRF
- In general variables or VLMs with a **« finite » set of values** e.g. it is not applicable to free-text
- **ADaM variables copied from SDTM** when the SDTM variables have a codelist defined (traceability)

Make use of subset-codelist

e.g. see example later for CM and LB Unit

→ **See tip nr. 3**

(Unveiled?) Tips

2. When do I need to assign / specify a codelist? (cont)

Numeric Variables with a decode or variables **containing abbreviated text**:

- VISITNUM with decode from VISIT
- QNAM with decode from QLABEL

VISITNUM Mapping

Permitted Value (Code)	Display Value (Decode)
10	Enrollment
120	1 Year Interval

Qualifier Variable Name for Supplemental Qualifiers for MH

Permitted Value (Code)	Display Value (Decode)
APAHDIA1	Associated PAH Diagnosis 1
APAHDIAG	Associated PAH Diagnosis
APAHTYPE	Associated PAH Type

- --TESTCD with decode from --TEST
- PARAM PARAMN with decode from PARAM
- PARAMCD with decode from PARAM

(Unveiled?) Tips

2. When do I need to assign / specify a codelist? (cont)

The following do **NOT need a codelist** to be defined in define-xml

- MedDRA, WHO-DD, etc. → External Dictionaries
- ISO 8601 (date/time/duration) → External Standard format handled by the stylesheet
- ISO 3166 (country) → Yes recommended to create a codelist with “applicable” countries

(Unveiled?) Tips

3. Use of subset codelist

- CDISC-CT can be a **“superset” of terms used across different variables**, datasets e.g. UNIT
- Not all terms are applicable to all variables where the same CDISC-CT is used e.g. CMDOSU and LBORRESU

Unit [CL.UNIT, *C71620*]

Permitted Value (Code)	Display Value (Decode)
% [*C25613*]	Percentage
10^12/L [*C67308*]	Million per Microliter
10^9/L [*C67255*]	Billion per Liter
AMPULE [*C48473*]	Ampule Dosing Unit
APPLICATION [*C25397*]	Application Unit
CAPFUL [*C102405*]	Capful Dosing Unit
CAPLET [*C64696*]	Caplet Dosing Unit
CAPSULE [*C48480*]	Capsule Dosing Unit
cm [*C49668*]	Centimeter
EU [*C96599*]	Ehrlich Unit
fL [*C64780*]	Femtoliter

CMDOSU	Dose Units		text	~~40~~	Unit

LBORRESU	Original Units		text	~~200~~	Unit

This unit very likely does not apply to Laboratory Tests

(Unveiled?) Tips

3. Use of subset codelist (cont)

- Do not create one UNIT CT for CMDOSU and LBORRESU
- Create **two separate unit codelists as a "subset" of the CDISC-CT UNIT**

Units for Laboratory Results [CL.LBRESU, *C71620*]

Permitted Value (Code)	Display Value (Decode)
g/L [*C42576*]	Kilogram per Cubic Meter
mg/dL [*C67015*]	Milligram per Deciliter
mol/L [*C48555*]	Mole per Liter
pg/mL [*]	Picogram per Milliliter
umol/L [*C48508*]	Micromole per Liter
ng/L [*C67327*]	Nanogram per Liter
pmol/L [*C67434*]	Picomole per Liter

Units for Concomitant Medications [CL.CMDOSU, *C71620*]

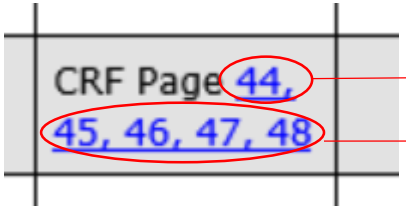
Permitted Value (Code)	Display Value (Decode)
L/min [*C67388*]	L/min
OTHER [*]	Other
mL/min [*C64777*]	Milliliter per Minute
mL/min/m2 [*]	Milliliter per Minute per Square Meter
mg [*C28253*]	Milligram
mg/L [*C64572*]	mg/L
mg/h [*C66969*]	mg/h

(Unveiled?) Tips

4. When origin=CRF and you have to link to Multiple Pages

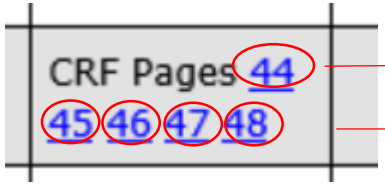
- Make use of space and not comma

```
<def:PDFPageRef Type="NamedDestination" PageRefs="44, 45, 46, 47, 48"/>
```



These links to individual acrf pages will not work once the define-xml is rendered by the stylesheet

```
<def:PDFPageRef Type="PhysicalRef" PageRefs="44 45 46 47 48"/>
```



These links to individual acrf pages will point to individual acrf pages once the define-xml is rendered by the stylesheet

(Unveiled?) Tips

5. Origin=Assigned vs Origin=Derived

Derived

- *"If the variable or set of parameter analysis values is calculated, then origin type is Derived"*. **Variables derived in the eDC are not considered derived**

Assigned: From CDISC *"Data that is determined by **individual judgment** (by an evaluator other than the subject or investigator)... This may include third party attributions by an adjudicator"* or *"Values that are set independently of any subject-related data values in order to complete SDTM fields such as DOMAIN and --TESTCD are considered to have an origin type of 'Assigned"*. Other examples

- **Secondary Variables** in ADaM e.g. SEXN (secondary of character SEX)
- **Logically Synonimous** e.g. PARAM/PARAMCD/PARAMN

(Unveiled?) Tips

6. Reviewer Guide vs define-xml - When it's time to find alternatives to define-xml

TRTA	Actual Treatment	text	14	TRT	Derived: Set to "SCREEN FAILURE" for all screen failures (check if ADSL.ACTARM = "SCREEN FAILURE") and set to "NOT TREATED" for all patients not treated (check if ADSL.ACTARM ="NOT TREATED"). For all other patients the following rules apply: If study is not in ("[redacted]","[redacted]-E") then set to ADTREAT.TRTA. Else If study ="[redacted]" then do: If SR.EPOCH in("SCREENING","BLINDED TREATMENT") set to ADTREAT.TRTA where ADTREAT.APHASE="BLINDED TREATMENT", else if SR.EPOCH in("FOLLOW-UP","OPEN LABEL TREATMENT","OPEN LABEL TREATMENT EXT") set to ADTREAT.TRTA where ADTREAT.APHASE="OPEN LABEL TREATMENT", else if SR.EPOCH is missing and SR.SRDTC is not missing then do: [if SR.SRDTC <= ADTREAT.AENDT where ADTREAT.APHASE="BLINDED TREATMENT" then set to ADTREAT.TRTA where ADTREAT.APHASE="BLINDED TREATMENT", else set to ADTREAT.TRTA where ADTREAT.APHASE="OPEN LABEL TREATMENT"]. Else if study = "[redacted]" then do: If SR.EPOCH in("SCREENING","OPEN LABEL FIRST TREATMENT") set to ADTREAT.TRTA where ADTREAT.APHASE="OPEN LABEL FIRST TREATMENT", else if SR.EPOCH in("FOLLOW-UP","OPEN LABEL SECOND TREATMENT") set to ADTREAT.TRTA where ADTREAT.APHASE="OPEN LABEL SECOND TREATMENT", else if SR.EPOCH is missing and SR.SRDTC is not missing then do: [if SR.SRDTC < ADTREAT.TRTSDT where ADTREAT.APHASE="OPEN LABEL SECOND TREATMENT" then set to ADTREAT.TRTA where ADTREAT.APHASE="OPEN LABEL FIRST TREATMENT" else set to ADTREAT.TRTA where ADTREAT.APHASE="OPEN LABEL SECOND TREATMENT"]

(Unveiled?) Tips

6. Reviewer Guide vs define-xml - When it's time to find alternatives (cont)

- define.xml has some **visual limitations**
- **Long text** might be **not always readable**
- If you see **your text could be not read**, than it's time to **find an alternative to define.xml**

AVISIT	Analysis Visit	text	25	[redacted]_AVISIT	Derived: See SAP 4.8.5 for the time frames Calculate the time frames (AVISIT) checking whether the related SR.SRDTC is in the frame of the period starting at Date of first exposure to treatment (ADSL.TRTSDT). [redacted]AVISIT Derivation Algorithm

(Unveiled?) Tips

7. SDTM Mapping Specifications are not needed in define-xml

ARM	Description of Planned Arm		text	14	["Enrolled", "Screen Failure"] <Arm>	Assigned	Taken from IVRS dataset RANDOM.TRT
ACTARMCD	Actual Arm Code		text	8	["ENROLLED" = "Enrolled", "SCRNFAIL" = "Screen Failure"] <Arm (Code)>	Derived	Same as ARMCD
ACTARM	Description of Actual Arm		text	14	["Enrolled", "Screen Failure"] <Arm>	Assigned	Assigned from TA.ARM based on ACTARMCD.
COUNTRY	Country		text	3	ISO 3166	Assigned	Derived from SITEINFO.CTRY

- RANDOM.TRT what?
- SITEINFO.CTRY what?

These are mapping specifications and they should be not included in the define.xml!!!

(Unveiled?) Tips

8. Good vs Bad Computational Algorithms (Methods)

- Avoid use of programming code or “only” programming code e.g. SAS

PCHG	Percent Change from Baseline	float	Analysis Parameter	12	Derived ((AVAL-BASE)*100)/BASE

This is acceptable and straightforward to understand

(Unveiled?) Tips

8. Good vs Bad Computational Algorithms (Methods) (cont)

- Avoid use of programming code or “only” programming code e.g. SAS

Not Acceptable. “I miss the rationale!!!!”

```
If PAANLFL = 'Y‘ then do;

If SCRNOCFL ne 'Y' then APHASE = 'From Day 1 to Week 12‘;

Else if SCRNOCFL ne 'Y' and PAANLFL ne 'Y' and FUPOCFL ne 'Y' then
APHASE = 'From Week 12 to Week 24' ;

End;

Else if EXTFL = "Y" then do;

If ASTDT > W24VISDT and FUPOCFL ne 'Y' then APHASE = "From
Week 24 to Week 52“;

Else if EXTFL ne 'Y' and FUPOCFL = 'Y' then APHASE = "From Week 24
to Week 48“;

End;
```

- What’s the difference?
- Which subset of subjects is selected here?
- …..

(Unveiled?) Tips

8. Good vs Bad Computational Algorithms (Methods) (cont)

- Avoid “concise” description e.g. repeating what is already stated in the SAP

“Daily average rescue medication consumption as per information collected in the concomitant medications page.”

(Unveiled?) Tips

8. Good vs Bad Computational Algorithms (Methods) (cont)

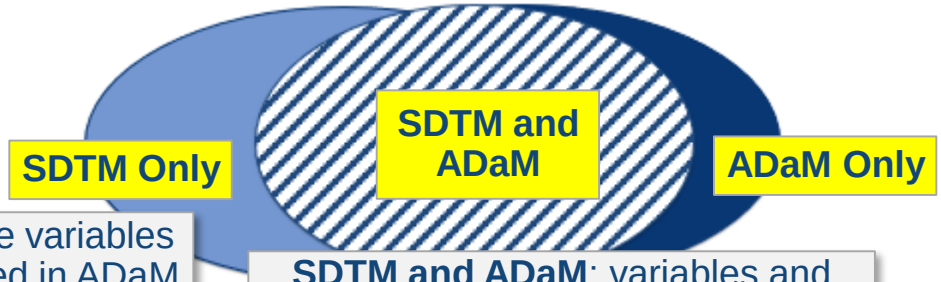
- Avoid “concise” description e.g. repeating what is already stated in the SAP

“Daily average rescue medication consumption as per information collected in the concomitant medications page.
It is derived from CM.CMDOSTXT where CM.CMDECOD='Paracetamol'.
More details can be found in the Rescue Medications Consumption Derivation document”

(Unveiled?) Tips

9. Consistency between SDTM and ADaM define-xml

(Traceability in SDTM/ADaM Metadata)



SDTM Only: source variables and records not used in ADaM i.e. Screen Failures, IE domain, Suppl. Lab. Data, etc.

SDTM and ADaM: variables and records copied from SDTM to ADaM

Traceability Issue 1

ADaM Only: derived or assigned variables / new records in ADaM for analysis purpose

Traceability Issue 2

TRACEABILITY ISSUE 1

Check for ADaM Origin=Predecessor

- Keep all variables attributes from SDTM
- Bring codelist defined in SDTM
- Content must be not changed

The same applies when Predecessor is another ADaM

Easy to check programmatically

TRACEABILITY ISSUE 2

Check for Clear Pattern of Derivations

Creation of Records if needed

e.g. impute missing observations

Clear description of derivations

It requires more Independent and Manual Review

(Unveiled?) Tips

9. Consistency between SDTM and ADaM define-xml (Traceability in SDTM/ADaM Metadata)

- Dataset Metadata e.g. Study Title
- Origin=Predecessor in ADaM
  e.g. SUPPAE.QVAL where QNAM=AETRTEM for ADAE.AETRTEM

Issues	Dataset	Variabl	Label	Codelis	Origin	Predecess	
[ERROR: Variable does not exist in predecessor dataset]	ADAE	AEHLTGCD	Group Term Code	MEDDRANUM	Predecessor	AE.AEHLTGCD	Di C
[ERROR: Variable does not exist in predecessor dataset]							
[ERROR: A codelist is referenced but the codelist	ADAE	AECAT	Category	AECAT	Predecessor	AE.AECAT	D

The following two variables do not exist in the SDTM while the ADAM specifications are indicating Origin=Predecessor

Conclusions

Conclusions

Do not cut corners! ***Try to imagine that you are the "recipient" of such package*** *and check, for example, if the* ***explanation of a derivation in the define.xml is clear enough****.*

Don't get bored, ***be patient, love and cure the standards!*** *Have a* ***passion for details*** *as they might matter when you submit your data to an agency.* ***That's*** *what I try to pass on to my colleagues almost every day, the passion for the data especially when they are organized in a standard way.*

References and Suggested Readings

- PhUSE WG, **Clinical Study Data Reviewer Guide and Analysis Data Reviewer Guide Template and Guidance**
- PhUSE WG, "**Best Practices for Documenting Dataset Metadata: Define-XML vs Reviewer's Guide**"
- PhUSE WG, "**Define-xml Version 2.0 Completion Guidelines**"
- PhUSE WG, "**Metadata Definitions**"

- D. Roulstone, "**Do's and Don'ts of Define.xml**", PhUSE-EU 2018, SA04
- S. Sirichenko, "**Diagnostic of technical errors in define-xml file**", PhUSE-US 2018, SI07
- J. Schoeman, N. Perry, "**Define'ing the Future**", PhUSE EU 2018, SI05
- S. Griffiths, "**ADaM Reviewer's Guide – Interpretation and Implementation**", PhUSE 2015, CD13
- M. Haloui and EL. Asam, "**High Quality Study Data Standards for Submission**", PhUSE-US 2019, SA03
- V. Debbeti, "**How to Prepare High-quality Metadata for Submission**", PhUSE-US 2018, SI12
- A. Tinazzi, "**The « CDISC Stupidario » (the CDISC Nonsense)**", PhUSE-EU 2018, PP26
- A. Tinazzi, "**How to Ensure Quality in Data Submission**", PharmaSUG-China 2019, DS-068

Thank You!

Backup slides

(Unveiled?) Tips

xx. Variable Metadata vs ValueLevel Metadata Consistency

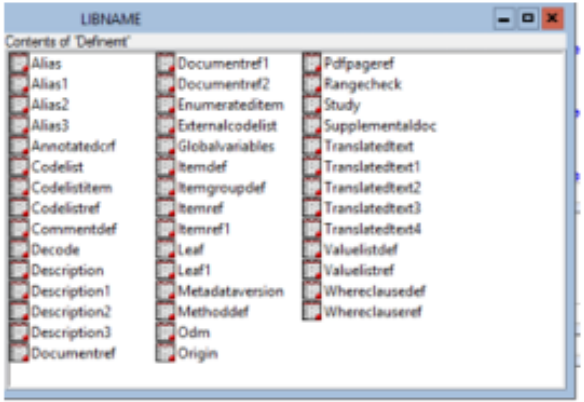
- When a variable has only one Origin Type for all its values and it has VLM, three options:
- Provide type only at variables metadata only, if all he same
- Provide type only at VLM metadata only, if diff
- Both but then the above hierarchy should be respected

Capture define-xml metadata in SAS

CAPTURE DEFINE-XML METADATA IN SAS

```
filename define          "…\define.xml" ;
filename map             "…\Work\defineXMLautomap.map";
libname define    xmlv2 automap=replace xmlmap=MAP;
libname defineMt      "…\Work";
proc copy inlib=define outlib=defineMt;
run;
libname define clear;
```

Accessing the Metadata from Define XML ", Lex Jansen, PharmaSUG 2018



(Unveiled?) Tips

xx. Clarity on Origin – define.xml v2.1

Type	Definition	Source (*)			
		Subject	Investigator	Vendor	Sponsor
Collected	Data that were actually observed or recorded by a person or received from an instrument.	ePro	CRF	Lab data, ECG	X
Derived	Data that is not directly collected, but is calculated by an algorithm or reproducible rule, which is dependent upon other data values.	X	X	Lab data, ECG	SDTM ADaM
Assigned	Data that is determined by individual judgment as provided by an evaluator other than the subject or investigator.	X	X	X	SDTM ADaM
Protocol	Data that is defined as part of the trial design preparation.	X	X	X	SDTM
Predecessor	Data that is copied from a variable in another dataset.	X	X	X	SDTM ADaM

```
<def:Origin Type="Collected"
            Source="Investigator">
```

```
<def:Origin Type="Predecessor"
            Source="Sponsor">
```

From “The Present and Future of Define-XML”, Lex Jansen, PhilaSUG 2018