

Do's and Don'ts of Define.xml

Best practices for creating define.xml that is clear and useful in a regulatory review

David Roulstone / Associate Director, Clinical Data Services
May 15, 2018

Presenter – David Roulstone



- › SME on FDA Standards Testing
- › Data Fitness Analyst on FDA CBER CASA-CV Project
- › 10+ years Industry Experience
- › Primarily focused on data conversions and data standards

Agenda

- › Introduction – users and content
- › Origins – Derived vs Assigned
- › Sponsor specific conventions
- › References to raw data
- › Incorrect information
- › Formatting issues
- › Scattered information
- › Codelist issues
- › Conclusions

“The data definition file describes the metadata of the submitted electronic datasets, and is considered arguably the most important part of the electronic dataset submission for regulatory review.”

FDA's Study Data Technical Conformance Guide, section 4.1.4.5

Data Definition Files for SDTM, SEND, and ADaM

<https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>

Users

Assume that the end consumers of your define.xml...

- › have never seen the study before
- › have not read the protocol in detail
- › have no information about the raw database
- › have little understanding of CDISC standards
- › are attempting to **gain familiarity** with the **study, data, and documents**, and are turning to the define.xml for explanations

Content

Your define.xml should...

- › explain the data
- › concisely define all the derivations used
- › not be too complicated
- › not require programming knowledge to read
- › not be overloaded with unnecessary information

Origins of Derived versus Assigned

- › Origins of derived require a derivation
- › Don't cut corners and list a derived variable as
Origin = Assigned, simply to avoid providing
the derivation
- › Quick define check – see if the Epoch
derivation is provided

Missing information

- › Sponsor and CDISC standards are sometimes confused
- › Fundamental CDISC standards end up being documented
- › Sponsor specific standards are neglected and not properly documented

Raw Data References

- › The FDA and PMDA do not see your raw data
- › Do NOT copy-paste ALL the info from your mapping spec into the define.xml
- › References to raw data should be updated to SDTM and/or ADaM language

Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier		text	200		Derived	SDTM.DM.STUDYID. Merge by Site Number and SUBJID.
DOMAIN	Domain Abbreviation		text	2	DOMAIN	Assigned	Set as AE
USUBJID	Unique Subject Identifier	1	text	200		Derived	SDTM.DM.STUDYID "-" DM.SITEID "-" DM.SUBJID
AESEQ	Sequence Number		integer	8		Derived	Sort by Primary Keys and assign sequence number in the ascending order. Restart numbering with each new subject.
AESPID	Sponsor-Defined Identifier		text	200		CRF Page	AE.RECORDPOSITION
AETERM	Reported Term for the Adverse Event	3	text	200		CRF Pages 140 148	1) AE.AETERM. 2) If YR.YRAEYN = Y then set YR.YRCOM and generate an alert message for this case.
AELLT	Lowest Level Term		text	200		Assigned	AE.AELLT
AELLTCD	Lowest Level Term Code		integer	8		Assigned	AE.AELLTCD
AEDECOD	Dictionary-Derived Term		text	200		Assigned	AE.AEDECOD
AEPTCD	Preferred Term Code		integer	8		Assigned	AE.AEPTCD

Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier		text	10		Protocol	
DOMAIN	Domain Abbreviation		text	2	DOMAIN	Assigned	
USUBJID	Unique Subject Identifier	1	text	19		Derived	DM.STUDYID "-" DM.SITEID "-" DM.SUBJID
AESEQ	Sequence Number		integer	8		Derived	Sort by Primary Keys and assign sequence number in the ascending order. Restart numbering with each new subject.
AESPID	Sponsor-Defined Identifier		text	12		eDT	Unique record identifier from the raw database. Provides traceability between raw, SDTM and ADaM data.
AETERM	Reported Term for the Adverse Event	3	text	21		CRF Pages 140 148	
AELLT	Lowest Level Term		text	12	Adverse Event Dictionary	Assigned	
AELLTCD	Lowest Level Term Code		integer	8	Adverse Event Dictionary	Assigned	
AEDECOD	Dictionary-Derived Term		text	12	Adverse Event Dictionary	Assigned	
AEPTCD	Preferred Term Code		integer	8	Adverse Event Dictionary	Assigned	

Incorrect Information

- › Protocol amendments often result in changes to the data collection
- › People focus on QC'ing data
- › The define.xml is often neglected

Formatting Issues

Define.xml does not support:

- › Bullet Points
- › Numbered Lists
- › New Line Characters / Carriage Returns

EXDOSE	Dose per Administration		float	5		Derived	<p>1. For EX.EXCAT='For records with STUDY DRUG ADMINISTRATION': a. EXTRT=ABC123 check the following: If SUPPEX.QVAL WHERE QNAM=DISPAMT is equal to 0 then EXDOSE=0. If SUPPEX.QVAL WHERE QNAM=DISPAMT is greater than 0 then check : When EX.EXTRT=ABC123 and EXROUTE=ORAL, EX.EXSPID= 1 then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*0.5, If EXSPID=2 then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*0.75. If EXSPID >2 and DM.ACTARM=contains 1 mg then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*1 If EXSPID >2 and DM.ACTARM=contains 1 mg then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*1 b. When EXTRT=Placebo and (SUPPEX.QVAL WHERE QNAM=DISPAMT) >0 then EXDOSE=0 2. EXCAT='ON SITE DRUG ADMINISTRATION' and EXSTDTC non missing a. if VISIT=BASELINE/V1M1 then EXDOSE=0.5 b.if ACTARM contains 1 mg otherwise check c. If ACTARM contains 0.75 mg then EXDOSE=0.75</p>
--------	-------------------------	--	-------	---	--	---------	---

EXDOSE	Dose per Administration		float	5		Derived	<p>Please see the complex algorithms document, linked below, to see the derivation with formatting applied. 1. For EX.EXCAT='For records with STUDY DRUG DOSING': a. EXTRT=ABC123 check the following: If SUPPEX.QVAL WHERE QNAM=DISPAMT is equal to 0 then EXDOSE=0. If SUPPEX.QVAL WHERE QNAM=DISPAMT is greater than 0 then check : When EX.EXTRT=ABC123 and EXROUTE=ORAL, EX.EXSPID= 1 then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*0.5, If EXSPID=2 then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*0.75. If EXSPID >2 and DM.ACTARM=contains 1.25 mg then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*1.25 If EXSPID >2 and DM.ACTARM=contains 1.25 mg then EXDOSE=(EXENDTC-EXSTDTC)*(SUPPEX.QVAL WHERE QNAM=DISPAMT)*1.25 b. When EXTRT=Placebo and (SUPPEX.QVAL WHERE QNAM=DISPAMT) >0 then EXDOSE=0 2. EXCAT='ON SITE DRUG DOSING' and EXSTDTC non missing a. if VISIT=BASELINE/V1M1 then EXDOSE=0.5 b.if ACTARM contains 1.25 mg otherwise check c. If ACTARM contains 0.75 mg then EXDOSE=0.75</p> <p>Complex Algorithms</p>
--------	-------------------------	--	-------	---	--	---------	---

Dataset	Variable	Derivation
EX	EXDOSE	<ol style="list-style-type: none"> If DM.ACTARM = "0.75 mg ABC123" and SUPPEX.DISPAMT > 0 then set EXDOSE as follows: <ol style="list-style-type: none"> If EXSTDY < 8 then EXDOSE = SUPPEX.DISPAMT * 0.5; If EXSTDY > 7 then EXDOSE = SUPPEX.DISPAMT * 0.75; Else If DM.ACTARM = "1.25 mg ABC123" and SUPPEX.DISPAMT > 0 then set EXDOSE as follows: <ol style="list-style-type: none"> If EXSTDY < 8 EXDOSE = SUPPEX.DISPAMT * 0.5; If EXSTDY > 7 EXDOSE = SUPPEX.DISPAMT * 1.25. For records where EXTRT = "Placebo": <ol style="list-style-type: none"> If SUPPEX.DISPAMT or SUPPEX.TAKEDOSE = "Yes" then set EXDOSE equal to "0"; Else set EXDOSE to Null. For records where EXTRT = "COMP": <ol style="list-style-type: none"> If SUPPEX.TAKEDOSE = "Yes" then set EXDOSE equal to "30"; Else set EXDOSE to Null. For records where EXCAT = "ON SITE DOSING": <ol style="list-style-type: none"> If EX.VISIT = "Month 1" and EX.EXSTDTC ^= null then set EXDOSE equal to "0.5". For subjects records where EX.VISIT ^= "Month 1" and EX.EXSTDTC ^= null and DM.ACTARM = "0.75 mg ABC123", set EXDOSE equal to "0.75". For subjects records where EX.VISIT ^= "Month 1" and EX.EXSTDTC ^= null and DM.ACTARM = "1.25 mg ABC123", set EXDOSE equal to "1.25". For subjects records where EX.VISIT ^= "Month 1" and EX.EXSTDTC ^= null and DM.ACTARM = "Comparator", set EXDOSE equal to "0". For records where EXTRT = "Placebo": <ol style="list-style-type: none"> If SUPPEX.DISPAMT > 0 or SUPPEX.TAKEDOSE = "Yes" then Set EXDOSE equal to "0"; Else set EXDOSE to Null. For records where EXTRT = "COMP": <ol style="list-style-type: none"> If SUPPEX.TAKEDOSE = "Yes" then Set EXDOSE equal to "30"; Else set EXDOSE to Null.

Scattered Information

- › Not all define.xml tools are created equal
- › Some systems can add to the confusion

Exposure Dataset (EX)						ex.xpt
Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
STUDYID	Study Identifier	text		ASSIGNED	Identifier	See Computational Method: DER229
DOMAIN	Domain Abbreviation	text	CODELISTC7	ASSIGNED	Identifier	See Computational Method: DER104
USUBJID	Unique Subject Identifier	text		DERIVED	Identifier	See Computational Method: DER271
EXSEQ	Sequence Number	integer		DERIVED	Identifier	See Computational Method: DER220
EXTRT	Name of Treatment	text		CRF Page 59	Topic	See Computational Method: DER301
EXDOSE	Dose	float		CRF Page 59	RecordQualifier	See Computational Method: DER302
EXDOSU	Dose Units	text	CODELISTC43	CRF Page 59	VariableQualifier	See Computational Method: DER109
EXDOSFRM	Dose Form	text	CODELISTC14	ASSIGNED	VariableQualifier	See Computational Method: DER107
EXDOSFRQ	Dosing Frequency per Interval	text	CODELISTC13	CRF	VariableQualifier	See Computational Method: DER108
EXROUTE	Route of Administration	text	CODELISTC37	CRF Page 59	VariableQualifier	See Computational Method: DER111
EXADJ	Reason for Dose Adjustment	text		CRF Page 59	RecordQualifier	See Computational Method: DER105
VISITNUM	Visit Number	float		DERIVED	Timing	See Computational Method: DER114
VISIT	Visit Name	text		CRF	Timing	See Computational Method: DER272

Computational Algorithms (DER104)	
Reference Name	Computation Method
DER104	Set as EX
Computational Algorithms (DER105)	
Reference Name	Computation Method
DER105	EX.EXADJ
Computational Algorithms (DER302)	
Reference Name	Computation Method
DER302	EX.EXDOSE. Amount of EXTRT when numeric.
Computational Algorithms (DER107)	
Reference Name	Computation Method
DER107	Set the value using protocol info if its not available directly from CRF.
Computational Algorithms (DER108)	
Reference Name	Computation Method
DER108	Set the value using protocol info if its not available directly from CRF.
Computational Algorithms (DER109)	
Reference Name	Computation Method
DER109	EX.EXDOSU.
Computational Algorithms (DER110)	
Reference Name	Computation Method
DER110	Set using SDTM.EX.EXSTDTC.
Computational Algorithms (DER111)	
Reference Name	Computation Method
DER111	EX.EXROUTE

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier		text	200			SDTM.DM.STUDYID. Merge by Site Number and SUBJID.
DOMAIN	Domain Abbreviation		text	2	DOMAIN		Set as EX
USUBJID	Unique Subject Identifier	1	text	200			SDTM.DM.STUDYID "-" DM.SITEID "-" DM.SUBJID
EXSEQ	Sequence Number		integer	8			Sort by Primary Keys and assign sequence number in the ascending order. Restart numbering with each new subject.
EXTRT	Name of Treatment	4	text	200		CRF Page 59	Set using EX.EXTRT
EXDOSE	Dose		float	8		CRF Page 59	EX.EXDOSE. Amount of EXTRT when numeric.
EXDOSU	Dose Units		text	200	UNIT	CRF Page 59	EX.EXDOSU.
EXDOSFRM	Dose Form		text	200	["SOLUTION" = "Solution"] < FRM >		Set the value using protocol info if its not available directly from CRF.
EXDOSFRQ	Dosing Frequency per Interval		text	200	["ONCE" = "Once"] < FREQ >	CRF Page	Set the value using protocol info if its not available directly from CRF.
EXROUTE	Route of Administration		text	200	["ORAL" = "Intraoral Route of Administration; PO"] < ROUTE >	CRF Page 59	EX.EXROUTE
EXADJ	Reason for Dose Adjustment		text	200		CRF Page 59	EX.EXADJ
VISITNUM	Visit Number	2	float	8			Set using SDTM.SV.VISITNUM. Numeric version of VISIT.
VISIT	Visit Name		text	200		CRF Page	Corresponding domain.VISIT

Codelist Issues

FDA has stated that a

“separate UNIT codelist **not** used for each variable” is a “common issue that **impacts review**”*

*DeYett Law, Crystal Allard, Mary Doi, Lilliam Rosario, Barbara Witczak, Jesse Anderson, Kathryn Matto, Austin Taylor, Jeno Pizarro, Margo Cohen, “Data Quality Findings from JumpStart” PhUSE CSS March 2017

http://www.phusewiki.org/docs/2017_CSS_US/PP29_Draft.pdf

Codelists

A codelist should show the
intended data collection
for that variable, or value
level metadata

Codelist Example

CRF collects 3 values for Race

DEMOGRAPHY		DM=Demographics														
Sex:	SEX	<input type="checkbox"/> Male <input type="checkbox"/> Female														
Date of birth:	BIRTHDTC	<table border="1"><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>DD</td><td>MM</td><td colspan="5">YYYY</td></tr></table>								DD	MM	YYYY				
DD	MM	YYYY														
Race:	RACE	<input type="checkbox"/> WHITE <input type="checkbox"/> BLACK OR AFRICAN AMERICAN <input type="checkbox"/> ASIAN														

2 Distinct Race Values in SDTM

subjid	sex	race
1001	M	BLACK OR AFRICAN AMERICAN
1002	F	BLACK OR AFRICAN AMERICAN
1004	M	BLACK OR AFRICAN AMERICAN
1005	M	WHITE
1006	M	WHITE
1007	M	BLACK OR AFRICAN AMERICAN
1008	M	WHITE
1009	F	WHITE
1010	M	WHITE
1011	M	WHITE
1003	M	WHITE
1012	M	WHITE
1013	M	WHITE
1014	M	WHITE
1015	M	WHITE
1016	M	WHITE

5 values in CDISC CT

Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value
	No	Race	RACE
C74457		Race	AMERICAN INDIAN OR ALASKA NATIVE
C74457		Race	ASIAN
C74457		Race	BLACK OR AFRICAN AMERICAN
C74457		Race	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER
C74457		Race	WHITE

Define.xml Race codelist holds all 3 CRF values

Race [CL.RACE, C74457]	
Permitted Value (Code)	
ASIAN	[C41260]
BLACK OR AFRICAN AMERICAN	[C16352]
WHITE	[C41261]

Do not...

- › Have one UNIT codelist for all unit variables/values across the data package
- › Some systems can add to the confusion
- › Overcrowd your define.xml with codelists meant for other data packages
- › Create codelists with all values from CDISC CT, when many values are irrelevant to your data package

EGORRESU	Original Units		text	200	["msec" = "Millisecond"] < UNIT >
----------	----------------	--	------	-----	--

EGSTRESU	Standard Units		text	200	["msec" = "Millisecond"] < UNIT >
----------	----------------	--	------	-----	--

EXDOSU	Dose Units		text	200	["mg" = "Milligram"] < UNIT >
--------	------------	--	------	-----	--

LBORRESU	Original Units	6	text	200	UNIT
----------	----------------	---	------	-----	----------------------

LBSTRESU	Standard Units		text	200	UNIT
----------	----------------	--	------	-----	----------------------

PCORRESU	Original Units		text	200	PKUNIT
----------	----------------	--	------	-----	------------------------

PCSTRESU	Standard Units	5	text	200	PKUNIT
----------	----------------	---	------	-----	------------------------

UNIT [CL.EG_UNIT, C71620]

Permitted Value (Code)	Display Value (Decode)
msec [C41140]	Millisecond

UNIT [CL.EXDOSU, C71620]

Permitted Value (Code)	Display Value (Decode)
mg [C28253]	Milligram

UNIT [CL.LBORRESU, C71620]

Permitted Value (Code)	Display Value (Decode)
% [C25613]	Percentage
10 ¹² /L [C67308]	1/pL; 10 ⁶ /mm ³ ; 10 ⁶ /uL; T/L;
10 ⁹ /L [C67255]	1/nL; 10 ³ /mm ³ ; 10 ³ /uL; 10 ³ /
fL [C64780]	Cubic Micrometer; Cubic Micron;
g/L [C42576]	Gram per Liter; Kilogram per Cub
/HPF [C96619]	Per High Powered Field
IU/L [C67376]	IE/L; IU/L; International Unit per
mg/dL [C67015]	Milligram per Deciliter; mg%
mL/min [C64777]	ML/Min
mmol/L [C64387]	Micromole per Milliliter; Millimole
pg [C64551]	Picogram
/uL [C67254]	/mm ³ ; Per Microliter
umol/L [C48508]	Micromole per Liter

UNIT [CL.LBSTRESU, C71620]

Permitted Value (Code)	Display Value (Decode)
% [C25613]	Percentage
10 ¹² /L [C67308]	1/pL; 10 ⁶ /mm ³ ; 10 ⁶ /uL; T/L;
10 ⁹ /L [C67255]	1/nL; 10 ³ /mm ³ ; 10 ³ /uL; 10 ³ /
fL [C64780]	Cubic Micrometer; Cubic Micron; f
g/L [C42576]	Gram per Liter; Kilogram per Cub
/HPF [C96619]	Per High Powered Field
IU/L [C67376]	IE/L; IU/L; International Unit per
mL/min [C64777]	ML/Min
mmol/L [C64387]	Micromole per Milliliter; Millimole
pg [C64551]	Picogram
/uL [C67254]	/mm ³ ; Per Microliter
umol/L [C48508]	Micromole per Liter

PKUNIT [CL.PCORRESU, C85494]

Permitted Value (Code)	Display Value (Decode)
% [C25613]	%
g [C48155]	g
mBq/mL [*]	mBq/mL
ng equivalents/g [*]	ng equivalents/g
pg eq/mL [*]	pg eq/mL
pg/mL [C67327]	pg/mL

* Extended Value

PKUNIT [CL.PCSTRESU, C85494]

Permitted Value (Code)	Display Value (Decode)
h*pg/mL [C85635]	h*pg/mL
h*pmol/L [C85612]	h*pmol/L
L [C48505]	L
mL [C28254]	
mL/min [C64777]	mL/min
ng [C48516]	
pg [C64551]	
pg/mL [C67327]	pg/mL
pmol/L [C67434]	pmol/L

PKUNIT [CL.PP_UNIT, C85494]

Permitted Value (Code)	Display Value (Decode)
1/h [*]	1/h
h [C25529]	h
h*pg/mL [C85635]	h*pg/mL
L [C48505]	L
mL/min [C64777]	mL/min
pg/mL [C67327]	pg/mL

Study: CDISC01

Import Export Compare

Properties Datasets Variables Value Level Codelists Codelist Items Methods

Search

Columns

Issues 329

Dataset	Variable	Label	Data Type	Origin	Pages
AE	STUDYID	Study Identifier	text	Protocol	
AE	DOMAIN	Domain Abbreviation	text	Assigned	
AE	USUBJID	Unique Subject Identifier	text	Derived	
AE	AESQ	Sequence Number	integer	Derived	
AE	AESPID	Sponsor-Defined Identifier	text	CRF	21
AE	AETERM	Reported Term for the Adverse Event	text	CRF	21
AE	AEMODIFY	Modified Reported Term	text	Assigned	
AE	AEDECOD	Dictionary-Derived Term	text	Assigned	
AE	AEBODSYS	Body System or Organ Class	text	Assigned	
AE	AESV	Severity/intensity	text	CRF	21
AE	AESER	Serious Event	text	CRF	21
AE	AEACN	Action Taken with Study Treatment	text	CRF	21
AE	AEREL	Causality	text	CRF	21
AE	AESTDTC	Start Date/Time of Adverse Event	date	CRF	21
AE	AEENDTC	End Date/Time of Adverse Event	date	CRF	21
AE	AESTDY	Study Day of Start of Adverse Event	integer	Derived	
AE	AEENDY	Study Day of End of Adverse Event	integer	Derived	
AE	AEENRF	End Relative to Reference Period	text	CRF	21
CM	STUDYID	Study Identifier	text	Protocol	
CM	DOMAIN	Domain Abbreviation	text	Assigned	
CM	USUBJID	Unique Subject Identifier	text	Derived	

Study Datasets
Excel Spec
Define.xml
Annotated CRF
CSV

Pinnacle 21 Define.xml Designer

Provide reviewers with a clear understanding of your data.
Pinnacle 21 Enterprise creates define.xml files quickly and easily,
while helping you follow all the do's and avoid all the don'ts.

Further reading

CDISC Define-XML standard

<https://www.cdisc.org/standards/data-exchange/define-xml>

PhUSE Define-XML v2.0 Completion Guidelines &
Style Sheet Recommendations

http://www.phusewiki.org/wiki/index.php?title=Define-XML_V2.0_Completion_Guidelines_%26_Style_Sheet_Recommendations

Common Define.xml File Issues Seen During FDA's
JumpStart Service - PhUSE CSS 2018

<http://www.phusewiki.org/docs/2018%20US%20CSS/posters/PP19.pdf>

Summary & Questions

- › The define.xml is for everyone
 - › Medical degrees are not required
 - › Programming experience is not required
 - › Keep it simple
 - › Ensure your derivations are correct and understandable


David Roulstone
droulstone@pinnacle21.net

Get updates about Pinnacle 21

 linkedin.com/company/pinnacle-21

 twitter.com/pinnacle_21

 facebook.com/pinnacle21

 pinnacle21.com/blog