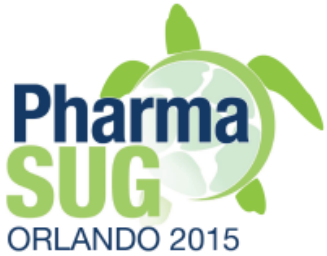


Creating Define.xml with OpenCDISC Community 2.0

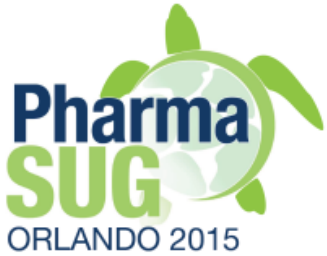
Sergiy Sirichenko, Pinnacle 21

PharmaSUG 2015
Paper HT04-Appendix

Abbreviations

- OC – OpenCDISC
- OCC – OpenCDISC Community
- OCE – OpenCDISC Enterprise
- CT – CDISC Control Terminology
- VL – Value Level

Topics

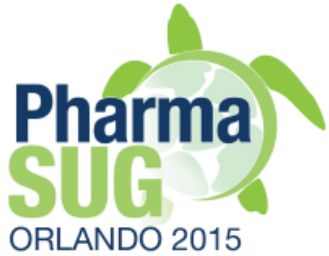


- Define.xml standard overview
- OCC vs. OCE
- Process flow
- Data Elements and Attributes
- MS Excel as a data entry tool
- Exercise and Q&A

PharmaSUG ORLANDO 2015

Define.xml standard overview

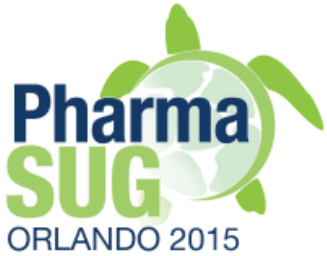
Define.xml v1.0.0



- Outdated
- 2005
- Has many limitations
- Cannot fully and correctly describe data

- OC does not support creation of v1.0
- OC provides migration of v1.0 to v2.0

Define.xml v2.0.0

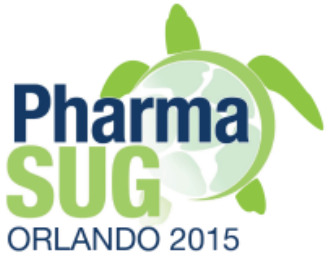


- 2013
- Robust Value Level
  - Reference to Variable
  - Multiple WhereClause conditions
- External documents
- Comments to everything
- More structural and logical standard
- Can handle ADaM Data

Standard content

- Define.xml specifications
  - Elements and attributes
  - Control terms
  - Business rules
  - Examples
- Data package implementation examples
  - SDTM
  - ADaM
- http://www.cdisc.org/define-xml

XML requirements



- Case-sensitive
  - “No” is not the same as “NO” or “no”
- Special characters
  - “>” should be replaced by “>”
  - “, ‘, <, >, &, …
  - See xml documentation for details
- Space character is still a real character!

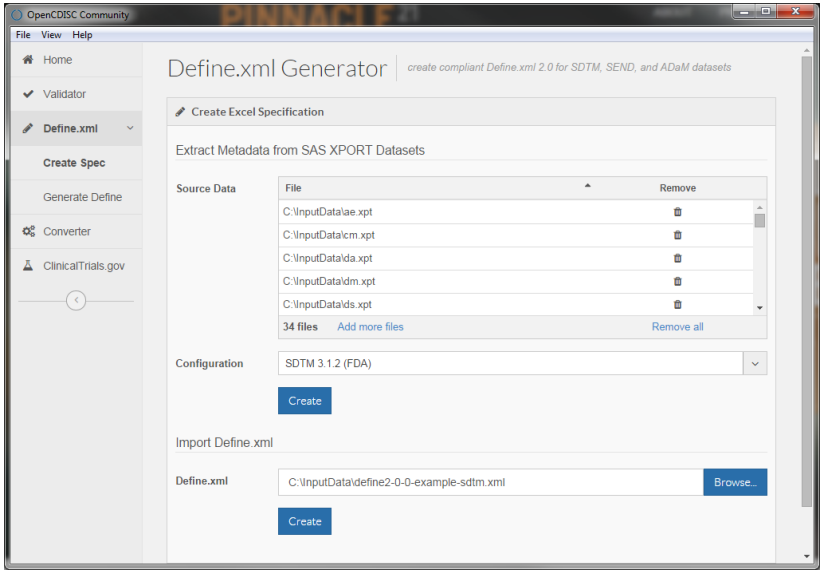
- Ensure consistency of values across IDs including character case and space characters

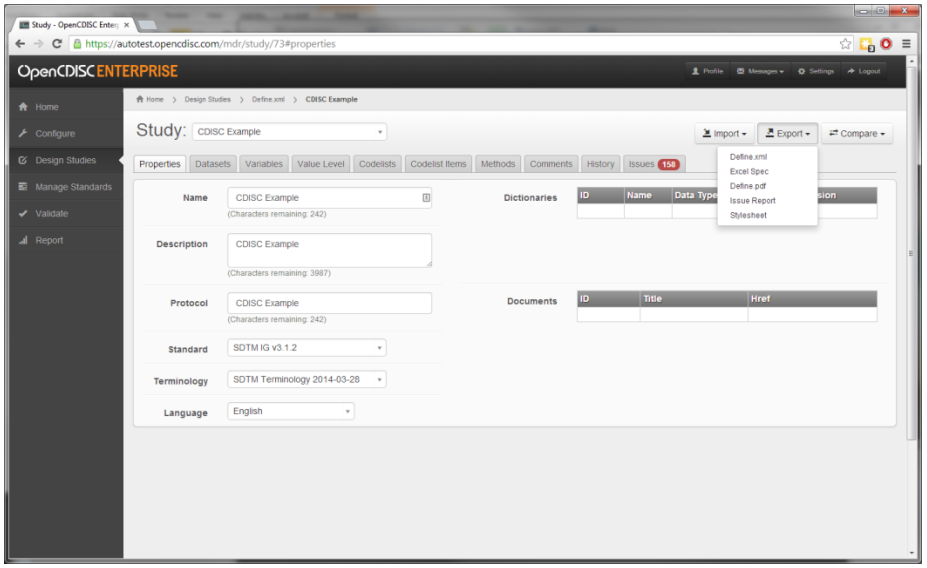
Define.xml is metadata

- It does not store any text format
  - Bold, Italic, Fonts Size or Color
- Special formatting characters will be removed
  - <New Line>, paragraph, bullets
- Metadata will be displayed as a continues plain text

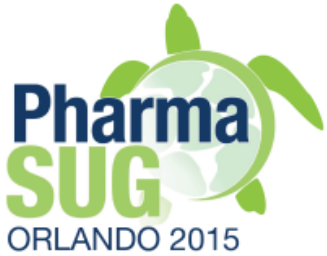
- Ensure to adjust existing external specifications to be “define.xml friendly”

OpenCDISC Community vs. Enterprise


OpenCDISC Community
File View Help
Home
Validator
Define.xml
Create Spec
Generate Define
Converter
ClinicalTrials.gov
Define.xml Generator
create compliant Define.xml 2.0 for SDTM, SEND, and ADaM datasets
Create Excel Specification
Extract Metadata from SAS XPORT Datasets
Source Data
File
Remove
C:\InputData\ae.xpt
C:\InputData\cm.xpt
C:\InputData\da.xpt
C:\InputData\dm.xpt
C:\InputData\ds.xpt
34 files
Add more files
Remove all
Configuration
SDTM 3.1.2 (FDA)
Create
Import Define.xml
Define.xml
C:\InputData\define2-0-0-example-sdtm.xml
Browse...
Create


https://autotest.opencdisc.com/mdr/study/73#properties
OpenCDISC ENTERPRISE
Home
Configure
Design Studies
Manage Standards
Validate
Report
Study: CDISC Example
Import
Export
Compare
Properties
Datasets
Variables
Value Level
Codelists
Codelist Items
Methods
Comments
History
Issues 158
Define.xml
Excel Spec
Define.pdf
Issue Report
Stylesheet
Name
CDISC Example
Description
CDISC Example
Protocol
CDISC Example
Standard
SDTM IG v3.1.2
Terminology
SDTM Terminology 2014-03-28
Language
English
Dictionaries
ID
Name
Data Type
Documents
ID
Title
Href

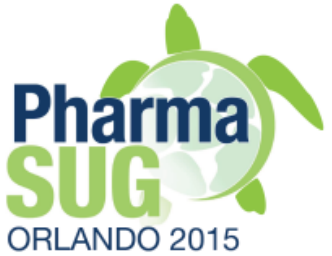
Define.xml tool OCC vs.OCE



- Designed for business users to generate content and not worry about XML syntax

- Common features:
  - Create define.xml for SDTM, SEND, ADaM
  - Convert define.xml v1.0 into define.xml v2.0
  - Extract metadata from SAS datasets
  - Export/Import specs in MS Excel format

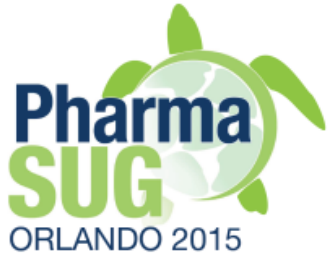
OCE additional features



- Utilize internal standards
- Scan Value Level metadata
- Create Variable and Value Level Codelists
- Extract Origin Page from aCRFs
- Merge metadata from external specs
- Generate define.PDF
- Validate define.xml content in real-time
- Compare content with standards and other studies
- Track changes between versions

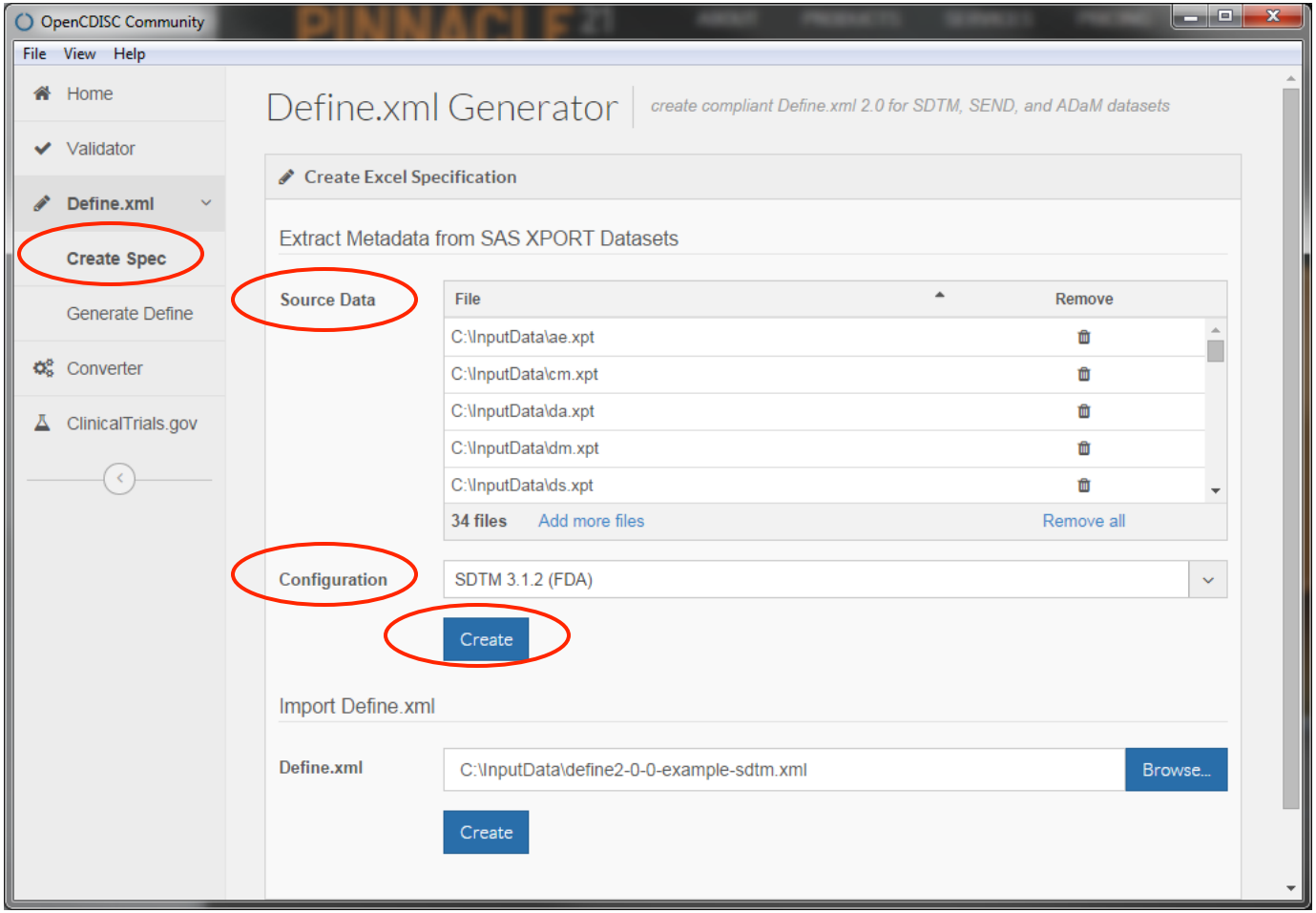
Process Flow

Descriptive

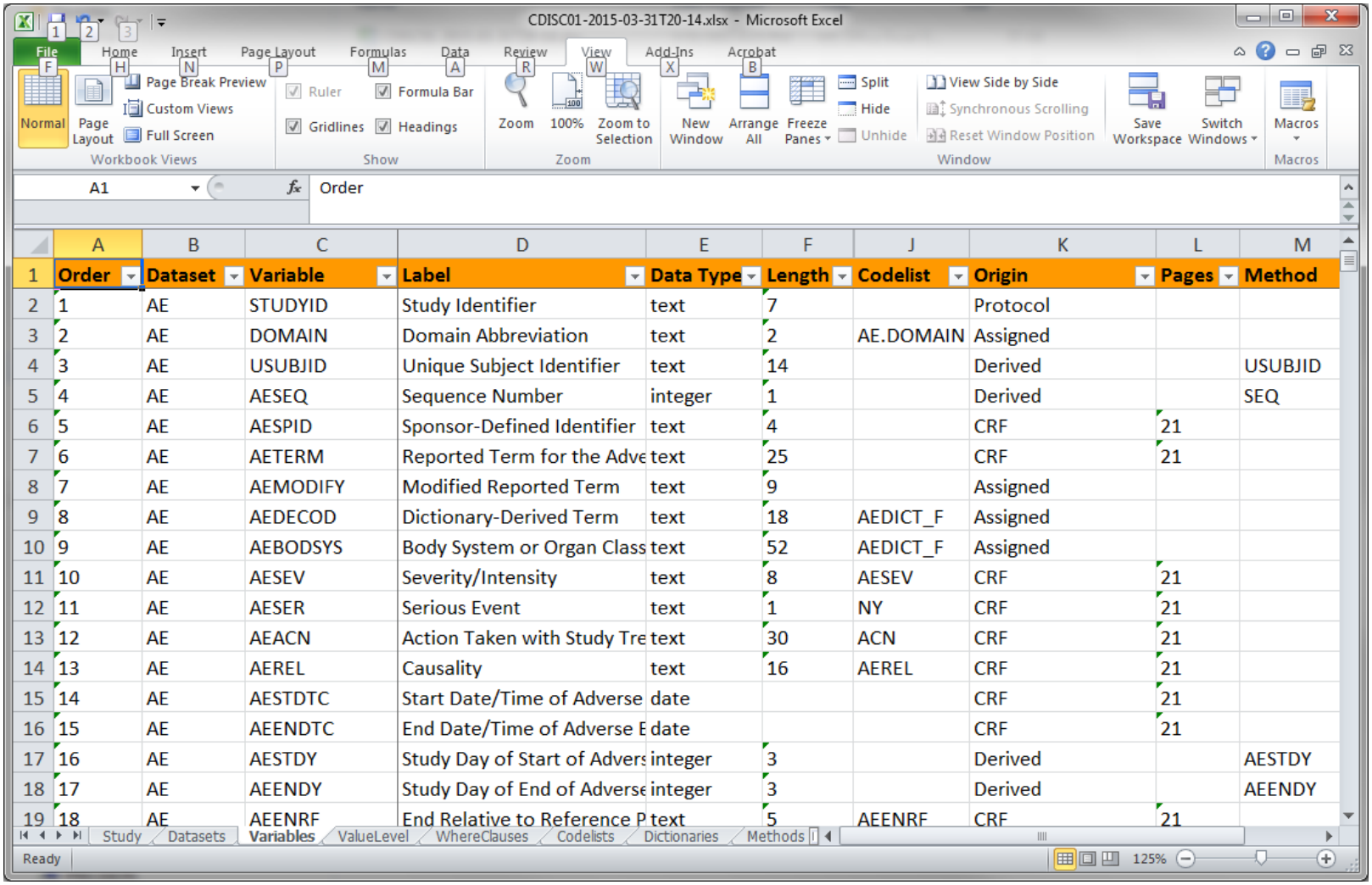


- Most common approach
- Finalize data
- Validate data and fix if needed
- Specify standard version, scan data, export to excel specs
- Populate missing metadata
- Generate define.xml from excel specs
- Validate define.xml and data vs. define.xml
- Correct errors

Scan

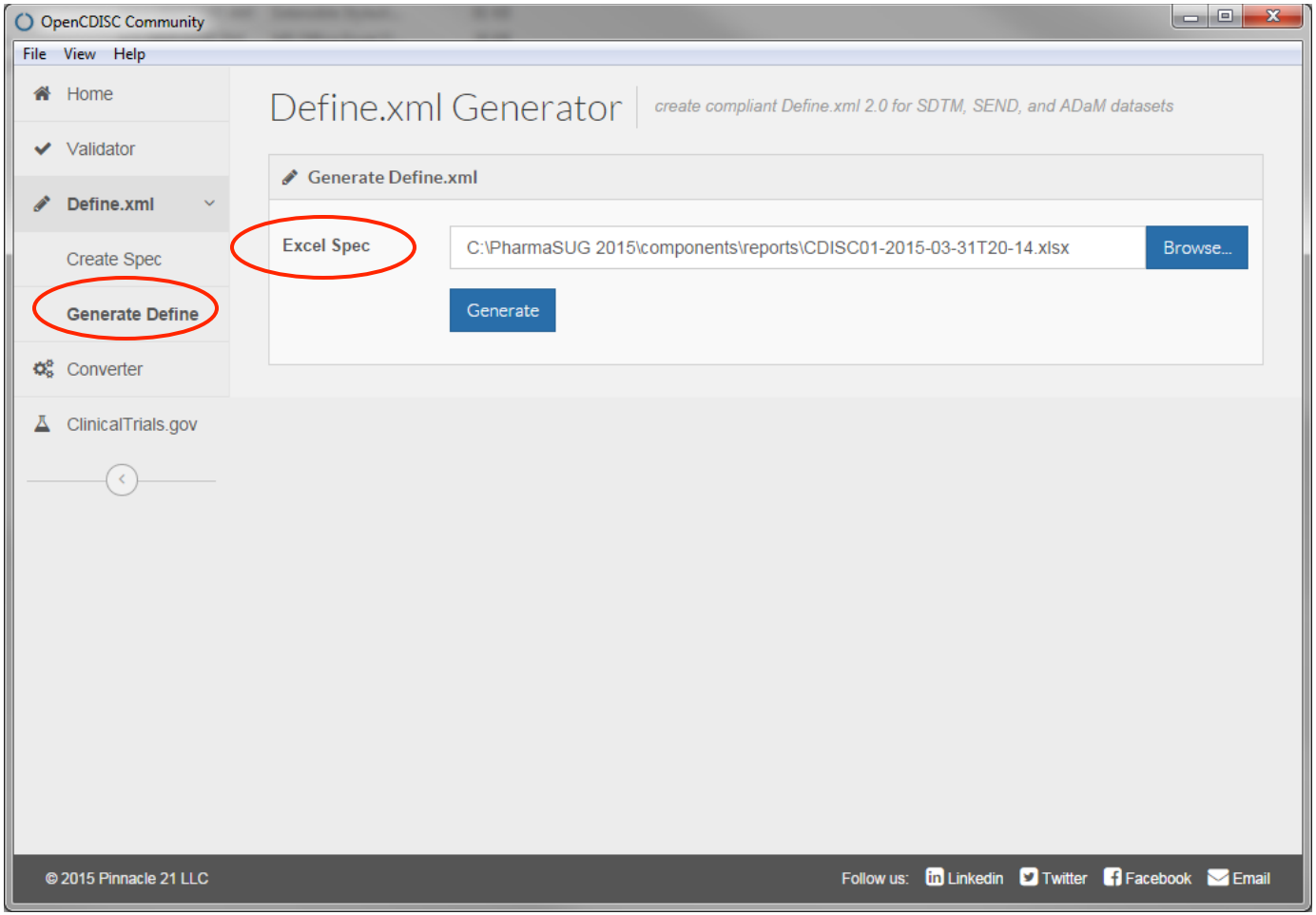

OpenCDISC Community
File View Help
Home
Validator
Define.xml
Create Spec
Generate Define
Converter
ClinicalTrials.gov
Define.xml Generator
create compliant Define.xml 2.0 for SDTM, SEND, and ADaM datasets
Create Excel Specification
Extract Metadata from SAS XPORT Datasets
Source Data
File
Remove
C:\InputData\ae.xpt
C:\InputData\cm.xpt
C:\InputData\da.xpt
C:\InputData\dm.xpt
C:\InputData\ds.xpt
34 files
Add more files
Remove all
Configuration
SDTM 3.1.2 (FDA)
Create
Import Define.xml
Define.xml
C:\InputData\define2-0-0-example-sdtm.xml
Browse...
Create

Populate

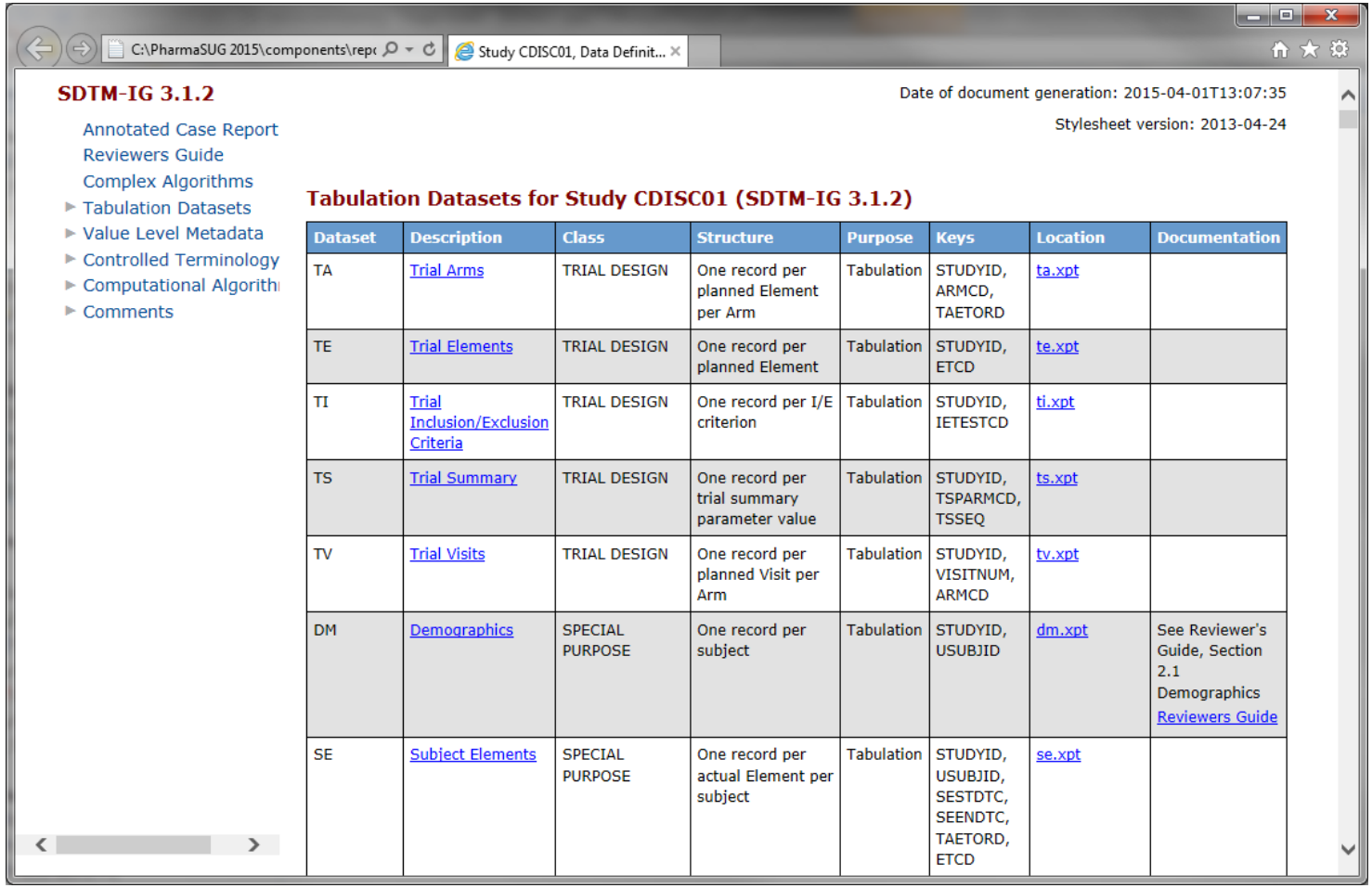


	A	B	C	D	E	F	J	K	L	M
1	Order	Dataset	Variable	Label	Data Type	Length	Codelist	Origin	Pages	Method
2	1	AE	STUDYID	Study Identifier	text	7		Protocol		
3	2	AE	DOMAIN	Domain Abbreviation	text	2	AE.DOMAIN	Assigned		
4	3	AE	USUBJID	Unique Subject Identifier	text	14		Derived		USUBJID
5	4	AE	AESEQ	Sequence Number	integer	1		Derived		SEQ
6	5	AE	AESPID	Sponsor-Defined Identifier	text	4		CRF	21	
7	6	AE	AETERM	Reported Term for the Adve	text	25		CRF	21	
8	7	AE	AEMODIFY	Modified Reported Term	text	9		Assigned		
9	8	AE	AEDECOD	Dictionary-Derived Term	text	18	AEDICT_F	Assigned		
10	9	AE	AEBODSYS	Body System or Organ Class	text	52	AEDICT_F	Assigned		
11	10	AE	AESEV	Severity/Intensity	text	8	AESEV	CRF	21	
12	11	AE	AESER	Serious Event	text	1	NY	CRF	21	
13	12	AE	AEACN	Action Taken with Study Tre	text	30	ACN	CRF	21	
14	13	AE	AEREL	Causality	text	16	AEREL	CRF	21	
15	14	AE	AESTDTC	Start Date/Time of Adverse	date			CRF	21	
16	15	AE	AEENDTC	End Date/Time of Adverse E	date			CRF	21	
17	16	AE	AESTDY	Study Day of Start of Advers	integer	3		Derived		AESTDY
18	17	AE	AEENDY	Study Day of End of Adverse	integer	3		Derived		AEENDY
19	18	AE	AEENRF	End Relative to Reference P	text	5	AEENRF	CRF	21	

Generate

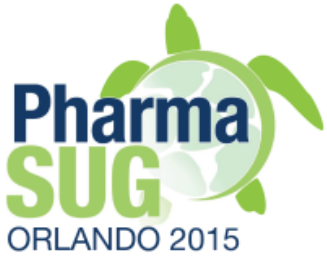

OpenCDISC Community
File View Help
Home
Validator
Define.xml
Create Spec
Generate Define
Converter
ClinicalTrials.gov
Define.xml Generator
create compliant Define.xml 2.0 for SDTM, SEND, and ADaM datasets
Generate Define.xml
Excel Spec
C:\PharmaSUG 2015\components\reports\CDISC01-2015-03-31T20-14.xlsx
Browse...
Generate
© 2015 Pinnacle 21 LLC
Follow us: Linkedin Twitter Facebook Email

Define.xml



Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt	
TE	Trial Elements	TRIAL DESIGN	One record per planned Element	Tabulation	STUDYID, ETCD	te.xpt	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt	
TS	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt	
TV	Trial Visits	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xpt	
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide
SE	Subject Elements	SPECIAL PURPOSE	One record per actual Element per subject	Tabulation	STUDYID, USUBJID, SESTDTC, SEENDTC, TAETORD, ETCD	se.xpt	

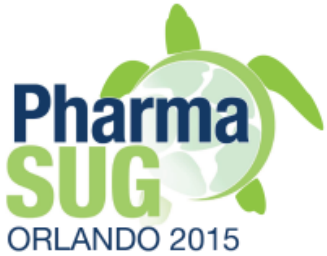
Prescriptive



- Use Standard or previous study as template
- Populate study specific metadata
- Use as enforcement for study data

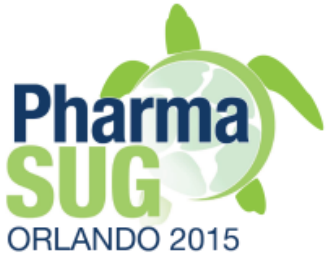
Data Elements and Attributes in OC specs

Datasets



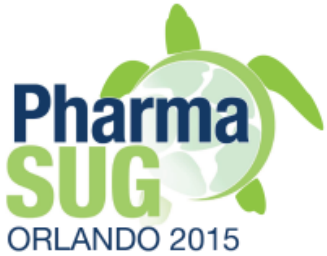
- Description is populated from XPT dataset labels
- Fix Labels in SAS datasets first, then re-scan data
- Structure and Key Variables are populated from "standard metadata"
  - Replace them with actual study specific info

Variables



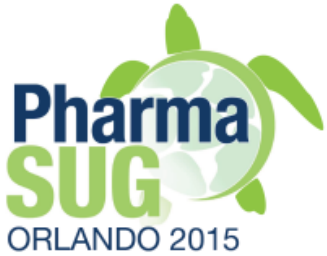
- Variables, Labels, Order, Length were taken from data
  - Change data first if corrections are needed, then re-scan data.
  - Length has missing value for "datetime" Datatype

Variables



- Datatype is generated based on data scanning
  - SAS Num -> float, integer
  - SAS Char -> text, datetime
- Review if corrections are needed
  - E.g., "2000-01" -> "text" instead of "datetime"
- Control terms in OCC:
  - text, datetime, float, integer
- Define.xml has additional Type terms:
  - E.g., time, partialDate, durationDatetime, etc.
- Populate actual Datatype for ValueLevel (consider ValueLevel as a new variable)

Variables



- SignificantDigits
  - Required for “float” Type
  - Is not used anywhere
- Format
  - Used in ADaM, not in SDTM
  - SAS format
    - E.g., Date9., YYMMDD10.
  - Limited to SAS native formats only, no custom formats are allowed in submission data

Variables

- Mandatory
  - For “Required” variables or ValueLevel – “Yes”
  - Otherwise – “No”
  - Attribute value is required (must be populated)

Variables

- Codelist
  - Expected for
    - SDTM variables with assigned standard Terminology. E.g., CDISC CT, MedDRA (OC specs use the same Codelist column for external dictionaries)
    - Variables which are "subject for study-specific CT"
      - Variables which collected using limited choices. E.g., Category, Subcategory
      - "Coded" variables. E.g., ARMCD, LBTESTCD, TRT01AN
  - Codelist must be populated in Codelists tab
    - Variables.Codelist = Codelists.ID

Variables

- Origin
  - One value per variable, use Value Level otherwise
  - Must be populated with control terms
  - Protocol
    - Examples: STUDYID, ELEMENT, EPOCH
  - CRF
    - Page value is required
    - Space character as a separator for pages
  - Derived
    - Method is required
    - Examples: --DY, --BLFL, RFSTDTC, EPOCH, --STRESC, --STRESN, --NRIND

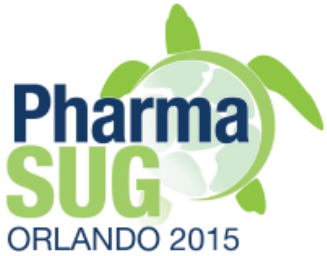
Variables

- Assigned
  - Coded variables: ARMCD, MedDRA, VISITNUM
  - --TEST, --PARM if there are no annotations
- eDT ("external data transfer")
  - Lab, IVRS
- Predecessor
  - Used in Analysis data
  - Exact copy of value in reference variable
  - Two major types of source:
    - SDTM
    - "Core" variables in ADSL
  - Examples:
    - DM.ARMCD, ADSL.SAFFL, ADSL.WEIGHT

Variables

- Pages
  - Required when Origin=CRF
  - Use reference to pdf document page (“Physical Reference”), not “CRF book (printed)” page
  - Space character as separator. E.g., “3 5 11”
  - OCE can scan and extract annotations from CRFs
- Method
  - Required if Origin=Derived
  - Should be defined in Methods tab
  - Variables.Method=Methods.ID

Variables



- Role
  - Not used anywhere
  - Populated from Standard metadata
- Comments
  - Optional
  - Should be defined in Comments tab
  - Variables.Comment=Comments.ID

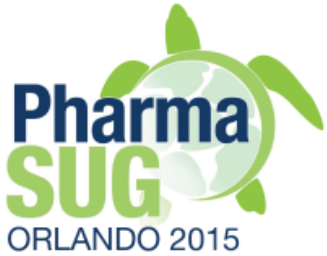
Value Level

- Many properties are similar to Variables
- Expected to be populated for
  - SUPPQUAL.QVAL (based on QNAM)
  - AVAL (in ADaM, based on PARAMCD, …)
  - Other common usage examples are
    - TSVAL (TSPARMCD)
    - --ORRESU, --STRESU, --ORRES, etc.
  - Can be used to specify different Origins for variable

Value Level

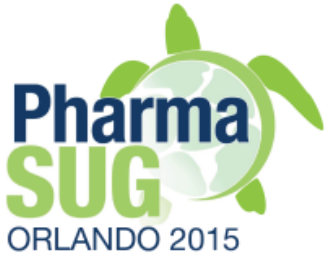
- Order is used for display sorting only
- Where Clause
  - Required
  - Defined in WhereClause tab
    - ValueLevel.Where Clause=WHereClauses.ID
    - It may be several records in WhereClause tab for a "multiple" condition
- Data Type
  - Data Type of Value Level, not "hosted" Variable
    - VSORRES is text, but VSORRES.DIABP is integer
    - Describe Value Level as "a new variable"

Value Level



- All attributes related to Value Level, not Variable
  - E.g., SUPPAE.QVAL has 200 char Length. However SUPPAE.QVAL.AETRTEM may have only 1 char Length (only “Y” and “N” values are used)

WhereClauses



- Multiple conditions are linked by the same ID
  - VSTEST=HEIGHT and DM.COUNTRY in (CAN, MEX)
- Dataset
  - Current dataset or DM (Demographics)
- Variable
  - Within specified dataset
- Comparator
  - EQ, NE, IN, NOTIN, LT, LE, GT, GE

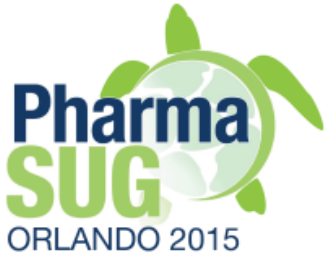
Codelists

- Populate Codelist for each Variable or Value Level which was collected or derived based on pre-specified terms
- Include only terms and all terms which were used as options during data collection process
  - Do not include full CT codelist. E.g., for EXDOSU
  - Remember to include terms from CRF, which are not present in actual collected data (limitation of data scanning process)

Codelists

- It's better to not have codelist for Variable/ CRF Field which was collected as free text
- Ensure to populate study specific codelits!
  - --CAT, --SCAT
  - ARM, EPOCH, ELEMENT
  - --GRID, --SPID if applicable
- Name of codelist
  - Should be unique and data specific
    - Recommend to create a separate codelist rather than generic. E.g., "Units for Exposure Dose"

Codelists



- Data Type should match Variable or Value Level
- Order must be unique or missing within codelist
- NCI Codelist Code, NCI Term Code
  - As specified by CDISC CT
  - A flag that particular term is based on CDISC CT
- Require exact match to standard term
- Decoded Value is optional, but expected for "coded" variables. E.g.,
  - ARMCD, --TESTCD, --PARMCD, QNAM (use QLABEL as Decoded Value), SEX, ETCD, coded Vars in ADaM

Dictionaries

- MedDRA, WHODrug, ISO3166 (Country Code)

Methods

- Used for Derived Variables and Value Level
- Type
  - Computation
  - Imputation (only in ADaM)
- Description should be done in terms of available data, no reference to external sources like EDC data
- Reference to additional external Documents are useful in case of complex algorithms
  - Methods.Document=Documents.ID
- Page is physical page of pdf document

Documents

- Href is a link to document
  - Relative links can be utilized
  - E.g., "../../analysis/adam/complex_algorithms.pdf"

- Annotated CRFs
  - A special case (Element) of Documents
  - Href values should be either "blankcrf.pdf" or "acrf.pdf" as the only two options specified by FDA. These two values are hardcoded in OC to create aCRF Element

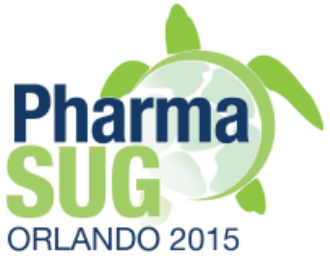
MS Excel as data entry tool

VLOOKUP example


```
IF(
ISNA(
VLOOKUP(CONCATENATE(B2, ".", C2),
Methods!$A$1:$A$428, 1, FALSE) ),
"",
VLOOKUP(CONCATENATE(B2, ".", C2),
Methods!$A$1:$A$428, 1, FALSE))
```

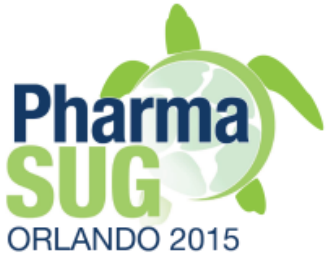
- IF and ISNA functions are optional to “clean” results

VLOOKUP function



- Reference in Table Array must be a value, not formula
  - Use formula, then copy results as values
- Usage examples:
  - Add/merge external specs
  - Populate NCI codelist code
- Additional columns and tabs in Excel specs are ignored by OC Define.xml tool
  - Use additional columns as reference specs
  - Add new tabs for look-up info

Exercise and Q&A



Name: Sergiy Sirichenko
Organization: Pinnacle 21
Address: 531 Plymouth Road, Suite #508
City, State ZIP: Plymouth Meeting, PA 19462
Work Phone: 908.781.2342
E-mail: ssirichenko@pinnacle21.net
Web: pinnacle21.net