

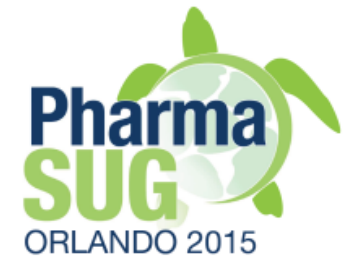
Usage of OpenCDISC Community Toolset 2.x for Clinical Programmers

Sergiy Sirichenko, Pinnacle 21

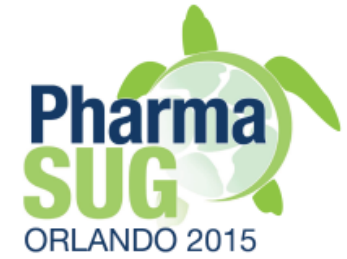
Michael DiGiantomasso, Pinnacle21

Travis Collopy, Pinnacle 21

PharmaSUG 2015
Paper #HT04



What is OpenCDISC?



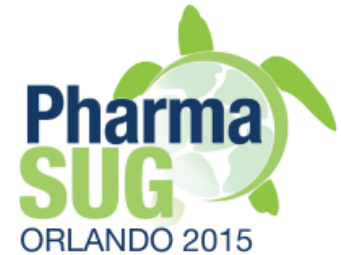
What is OpenCDISC?

Toolset for pharma



- ▶ Free, open source
- ▶ Widely used across the industry including FDA
- ▶ Easy to install and use
- ▶ Java based, metadata driven
- ▶ Validation, data conversion, creation of define.xml
- ▶ Available commercial version and support

History

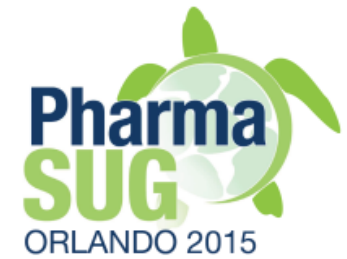


- ▶ 2008 – Validator alpha
- ▶ 2010 – FDA started using for validation of submission data
- ▶ 2011 – Pinnacle 21 was founded to provide commercial support
- ▶ 2011 – FDA DataFit (OpenCDISC Enterprise)
- ▶ 2014 – OpenCDISC Community v2.0.0
- ▶ 2015 – Deploying at PMDA

OCC vs. OCE

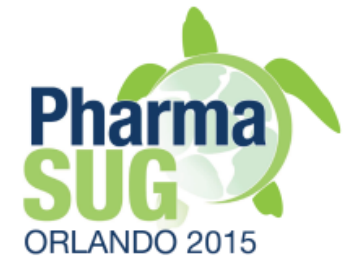


- ▶ Share the same engine
- ▶ Designed for different use cases
- ▶ OCC – personal desktop application
 - QC of your own work
 - Small organization and teams. E.g., pre-clinical
- ▶ OCE – web-based collaboration environment
 - Manage data quality across teams
 - Additional functionality
 - “Database on data issues”



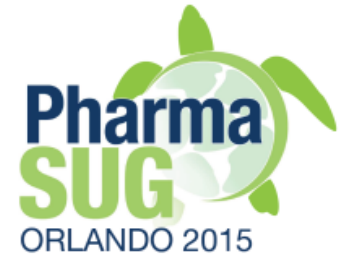
Installation and Tuning

Documentation



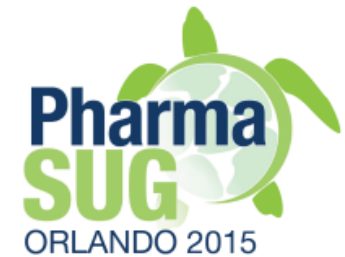
The screenshot shows a web browser window displaying the OpenCDISC Documentation page. The browser's address bar shows "www.opencdisc.org/documentation". The page features the OpenCDISC logo, navigation tabs (HOME, ABOUT, TOOLS, RULES, DOWNLOAD, DOCUMENTATION, CONTRIBUTE, FORUM), and a search bar. On the left, there are two user-related boxes: "SERGIY" with a star icon and links for "My account", "Newsletter Subscription", "Create content", and "Log out"; and "SUBSCRIBE" with icons for "Email", "Twitter", and "RSS Feed". The main content area has "VIEW" and "EDIT" links above the "Documentation" heading. Below this, a paragraph describes the collection of documents. The page lists "OpenCDISC Community Documentation" with sub-items: Installation, GUI Usage, CLI Usage, MedDRA and SNOMED Setup, CDISC CT Setup, Performance and Scalability, and Application Framework. It also lists "Publications and Articles" with links to "Dec 2010 Validator Review by Meta-Xceed, Inc.", "PharmaSUG 2012 - SDTM, ADaM and define.xml with OpenCDISC", "Controlling OpenCDISC Validator using R", and "FDA Datafit - Data Standards in Clinical Trials, A Regulatory Perspective". Finally, it lists "Regulatory Documents" with links to "CDER Study Data Standards", "CBER Study Data Standards", "CDER Common Data Standards Issues Document v1.1", "FDA Study Data Specifications v2.0", and "FDA Study Data Standards Resources".

Downloading and Installation



- ▶ Windows (both 32- and 64-bit versions)
- ▶ OS X, UNIX/Linux
- ▶ Installation package includes Java package
- ▶ “Download, unzip and run” installation
- ▶ Can be copied to and run from USB drive
- ▶ No administrative IT permissions are needed
- ▶ opencdisc.org/download
- ▶ `opencdisc-community.exe`

Download page



The screenshot shows a web browser window with the URL www.opencdisc.org/download. The page features the OpenCDISC logo and a navigation menu with links for HOME, ABOUT, TOOLS, RULES, DOWNLOAD (highlighted), DOCUMENTATION, CONTRIBUTE, and FORUM. A search bar is located in the top right corner.

OpenCDISC ENTERPRISE

Used by FDA to review submission data, Enterprise is the next evolution of CDISC compliance software. This comprehensive solution does everything — from validating data to creating Define.xml — to support your end-to-end lifecycle.

[LEARN MORE](#)


SUBSCRIBE


- Email
- Twitter
- RSS Feed

VIEW EDIT

DOWNLOADS

The following is latest release of OpenCDISC Community, which includes Validator, Define.xml Generator, Data Converter, and ClinicalTrials.gov Miner. The download also contains the latest set of standard validation configurations.

 **OpenCDISC Community 2.0.1**
[Download for Windows](#)

 **OpenCDISC Community 2.0.1**
[Download for Mac OS X](#)

Release Notes

INSTALLATION AND USAGE

The following are basic installation and usage instructions. For additional information, please refer to documentation.

- Installing OpenCDISC Community
- Using OpenCDISC Community
- Configuring OpenCDISC Validator for MedDRA

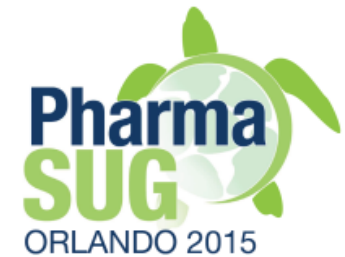
LICENSE

OpenCDISC Community is free software licensed under the OpenCDISC Open Source Software License. To learn more about our license check out the [FAQ](#).

SOURCE CODE REPOSITORY

OpenCDISC uses the [Bitbucket](#) source code control system for all of its source code. Although only active contributors get write access to the repository, anyone can browse our repository anonymously at <https://bitbucket.org/opencdisc/>

Application



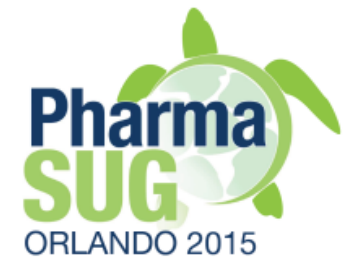
The screenshot shows the OpenCDISC Community application window. The title bar reads "OpenCDISC Community". The menu bar includes "File", "View", and "Help". A left sidebar contains navigation options: "Home", "Validator", "Define.xml", "Converter", and "ClinicalTrials.gov". The main content area features the "OpenCDISC COMMUNITY" logo and the tagline "an open source toolkit for the CDISC professional". Below this, a heading asks "Where would you like to begin?" and presents four options, each with an icon and a brief description: 1. Validator (blue checkmark icon): "Check compliance with SDTM, SEND, ADaM, and Define.xml". 2. Define.xml Generator (orange pencil icon): "Create compliant Define.xml 2.0 for SDTM, SEND, and ADaM datasets". 3. Data Converter (teal gears icon): "Convert data between SAS XPORT, Excel, CSV, and Dataset-XML". 4. ClinicalTrials.gov Miner (red flask icon): "Aggregate clinical outcomes and site info across trials and therapeutic areas". A "Recent Updates" section follows, listing five items with dates: "OpenCDISC Community 2.0.1 is now available. Download now" (19 days ago), "Everyone knows OpenCDISC. (Or so they think.) Read the true story." (a month ago), "FDA Final Guidance on Study Data Standards webinar - recap and presentation slides are now available" (2 months ago), "FDA Validation Rules webinar - recap and presentation slides are now available." (3 months ago), and "We have entered a new era: FDA finalized guidance for submission of standardized study data. Learn more" (3 months ago). The footer contains "© 2015 Pinnacle 21 LLC" and social media links for LinkedIn, Twitter, Facebook, and Email.

Auto-Update functionality



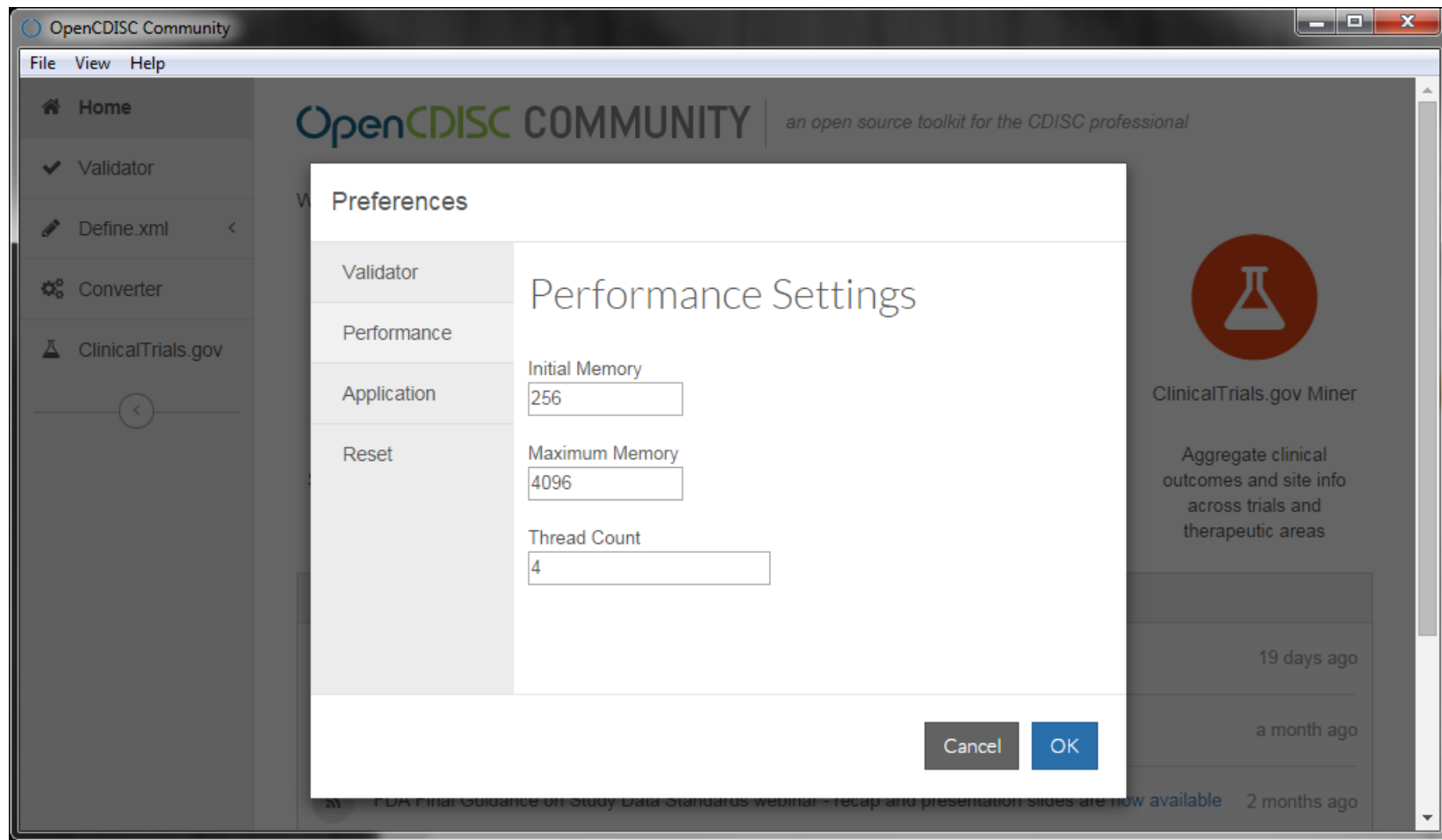
- ▶ Automatically checks for updates and install them
 - New versions of Control Terminology
 - New and updated validation configurations
 - New tools
- ▶ Alternative – check website and install new versions manually
 - Only most recent versions are available on website
 - Outdated versions should not be used and are not supported

Tuning

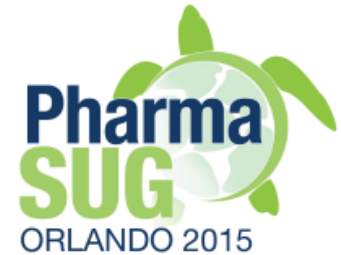


- ▶ Help → Preferences → Performance
- ▶ Maximum Memory
 - Recommended \geq 4GB
 - Depends on validated data
- ▶ Thread Count
 - Parallel processing of datasets
 - Recommended \geq 2
- ▶ Use “Reset to Default Settings” if needed

Tuning

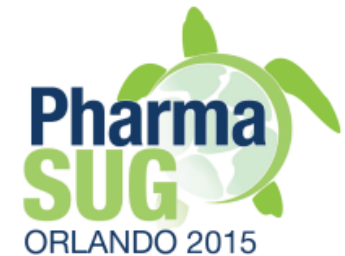


Dictionaries installation



- ▶ Supplied free dictionaries
 - CDISC Control Terminology
 - UNII, NDF-RT
- ▶ Pro-proprietary dictionaries
 - Manual installation using company's files
 - SNOMED
 - May be supplied by request for US companies (license limitation)
 - MedDRA
 - Create a folder with version number and copy MedDRA "ascii" files

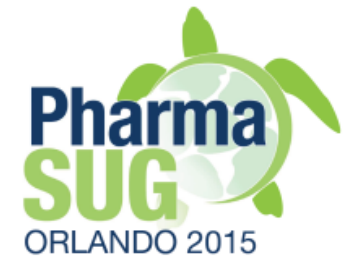
Exercise # 1



- ▶ Install MedDRA 7.1
 - Instructions on <http://www.opencdisc.org/projects/validator/configuring-opencdisc-validator-external-dictionaries>
 - Create a folder “7.1” in ... \components\config\data \MedDRA
 - Unzip MedDRA 7.1 files
 - Optional: keep only
 - pt.asc, llt.asc, hlt.asc, hlgt.asc, soc.asc
 - Start OpenCDISC Validator, check if version 7.1 is available in MedDRA drop-down box

Validator GUI

Validator GUI



OpenCDISC Community

File View Help

Validator

check compliance with SDTM, SEND, ADaM, and Define.xml

✓ Validate Data

Standard: SDTM Configuration: SDTM 3.1.2 (FDA)

Source Format: SAS® Transport (XPORT) Report Format: Excel

Source Data

File	Remove
C:\InputData\ae.xpt	Remove
C:\InputData\cm.xpt	Remove
C:\InputData\da.xpt	Remove
C:\InputData\dm.xpt	Remove
C:\InputData\ds.xpt	Remove

34 files [Add more files](#) [Remove all](#)

Define.xml: C:\InputData\define2-0-0-example-sdtm.xml [Browse...](#)

CDISC CT: 2014-03-28 MedDRA: 8.0 [More dictionaries](#)

[Validate](#)

Validation configurations



▶ SDTM

- SDTM 3.1.1 (FDA)
 - Executable version of FDA business rules for SDTM data. See [1]
- SDTM 3.1.2 (FDA)
- SDTM 3.1.3 (FDA)
- SDTM 3.2
 - SDTM version is not officially supported by FDA

▶ SEND

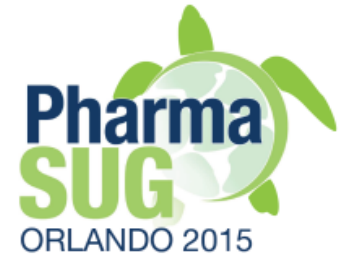
- SEND 3.0 (FDA)

Validation configurations



- ▶ **ADaM**
 - V1.0
 - Based on CDISC ADaM riles v1.3
 - ADSL, BDS
 - New: ADAE, TTE
- ▶ **Define.xml**
 - V1.0
 - V2.0
 - New more robust validation is introduced in v2.1.0

Source format



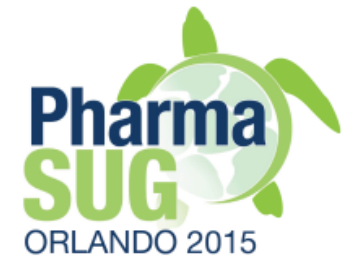
- ▶ Source format
 - SAS Transport v5 (XPORT)
 - Delimited (text)
 - CDISC Dataset–XML

Report format



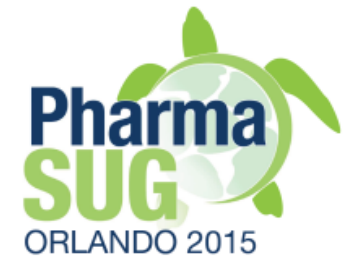
- ▶ MS Excel
 - Multi-tab document
 - Dataset Summary
 - Issue Summary
 - Details
 - Rules
- ▶ CSV
 - Details tab only
 - Preferable choice to read by SAS

Dataset Summary



	A	B	C	D	E	F	G	H
1	OpenCDISC Validator Report							
2								
3	Configuration: c:\PharmaSUG 2015\components\config\SDTM 3.1.2 (FDA).xml							
4	Define.xml: C:\InputData\define2-0-0-example-sdtm.xml							
5	Generated: 2015-03-31T19:11:18							
6	CDISC CT Version: 2014-03-28							
7	MedDRA Version: 8.0							
8	UNII Version: 2015-02-18							
9	NDF-RT Version: 2015-03-02							
10	Engine Version: 2.0.1							
11								
12	Processed Sources							
13	Domain	Label	Class	Source	Records	Errors	Warnings	Notices
14	GLOBAL	Global Metadata	--	--	--	0	0	0
15	AE	Adverse Events	EVENTS	ae.xpt	16	4	2	0
16	CM	Concomitant Medications	INTERVENTIONS	cm.xpt	36	26	3	0
17	DA	Drug Accountability	FINDINGS	da.xpt	16	3	1	0
18	DM	Demographics	SPECIAL PURPOSE	dm.xpt	5	2	2	3
19	DS	Disposition	EVENTS	ds.xpt	14	4	1	0
20	EG	ECG Test Results	FINDINGS	eg.xpt	56	4	1	0
21	EX	Exposure	INTERVENTIONS	ex.xpt	17	2	1	0
22	IE	Inclusion/Exclusion Criteria Not Met	FINDINGS	ie.xpt	1	0	0	0
23	LB	Laboratory Tests Results	FINDINGS	lb.xpt	83	5	36	0
24	MH	Medical History	EVENTS	mh.xpt	18	13	2	0
25	PE	Physical Examination	FINDINGS	pe.xpt	65	1	1	0
26	QS	Questionnaires	FINDINGS	qscg.xpt, qscs.xpt, qsmm.xpt	402	15	3	0
27	RELREC	Related Records	RELATIONSHIP	relrec.xpt	2	3	0	0

Issue Summary



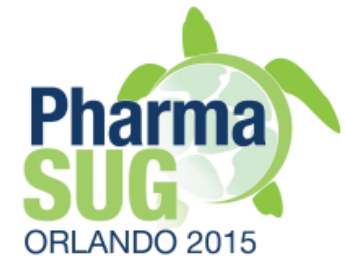
11						
12	Issue Summary					
13	Source	OpenCDISC ID	Publisher ID	Message	Severity	Found
14	AE					
15	SD0009	FDAC206		No qualifiers set to 'Y', when AE is Serious	Error	1
16	SD1082	FDAC036		Variable length is too long for actual data	Error	2
17	SD1089	FDAC130		AESTDY variable value is imputed	Error	1
18	SD1077	FDAC021		FDA Expected variable not found	Warning	1
19	SD1097	FDAC022		No Treatment Emergent info for Adverse Event	Warning	1
20	CM					
21	SD1082	FDAC036		Variable length is too long for actual data	Error	3
22	SD1089	FDAC130		CMSTDY variable value is imputed	Error	22
23	SD1093	FDAC135		CMENDY variable value is imputed	Error	1
24	SD1031	FDAC138		Value for CMENRF is populated, when RFENDTC is NULL	Warning	2
25	SD1077	FDAC021		FDA Expected variable not found	Warning	1
26	DA					
27	SD1082	FDAC036		Variable length is too long for actual data	Error	3
28	SD1077	FDAC021		FDA Expected variable not found	Warning	1
29	DM					
30	SD1082	FDAC036		Variable length is too long for actual data	Error	2
31	CT2002	FDAC341		RACE value not found in 'Race' extensible codelist	Warning	2
32	SKIP_SD0006			MB is missing or lacks necessary variables and cannot be used for this cross-dataset validation	Notice	1
33	SKIP_SD0006			MS is missing or lacks necessary variables and cannot be used for this cross-dataset validation	Notice	1
34	SKIP_SD0006			PC is missing or lacks necessary variables and cannot be used for this cross-dataset validation	Notice	1
35	DS					
36	SD1082	FDAC036		Variable length is too long for actual data	Error	1
37	SD1090	FDAC131		Incorrect value for DSSTDY variable	Error	3
38	CT2002	FDAC341		EPOCH value not found in 'Epoch' extensible codelist	Warning	1
39	EG					

Details



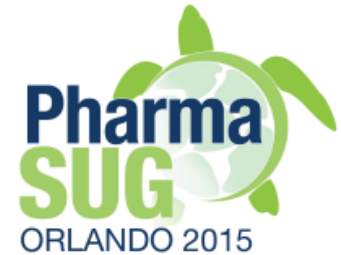
	A	B	C	D	E	F	G	H
1	Domain	Record	Count	Variables	Values	OpenCDISC ID	Publisher ID	Message
2	AE			VARIABLE, DATASET	EPOCH, AE	SD1077	FDAC021	FDA Expected variable not found
3	AE	8		AESER	Y	SD0009	FDAC206	No qualifiers set to 'Y', when AE is Serious
4	AE			Variable, Excess	AEACN, 14	SD1082	FDAC036	Variable length is too long for actual data
5	AE			Variable, Excess	AESPID, 3	SD1082	FDAC036	Variable length is too long for actual data
6	AE	1		SUB:RFSTDTC, AESTDTC, AESTDY	2003-04-29, 2003-05, 3	SD1089	FDAC130	AESTDY variable value is imputed
7	AE	8		AESEQ, USUBJID	5, CDISC01.100014	SD1097	FDAC022	No Treatment Emergent info for Adverse Event
8	CM			VARIABLE, DATASET	EPOCH, CM	SD1077	FDAC021	FDA Expected variable not found
9	CM	35		CMENRF, USUBJID	AFTER, CDISC01.200005	SD1031	FDAC138	Value for CMENRF is populated, when RFENDTC is NULL
10	CM	36		CMENRF, USUBJID	AFTER, CDISC01.200005	SD1031	FDAC138	Value for CMENRF is populated, when RFENDTC is NULL
11	CM			Variable, Excess	CMENRF, 1	SD1082	FDAC036	Variable length is too long for actual data
12	CM			Variable, Excess	CMDOSFRQ, 1	SD1082	FDAC036	Variable length is too long for actual data
13	CM			Variable, Excess	CMDECOD, 4	SD1082	FDAC036	Variable length is too long for actual data
14	CM	1		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1986, -5963	SD1089	FDAC130	CMSTDY variable value is imputed
15	CM	2		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1987, -5598	SD1089	FDAC130	CMSTDY variable value is imputed
16	CM	3		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1995, -2676	SD1089	FDAC130	CMSTDY variable value is imputed

Report Settings



- ▶ File Name Format
- ▶ Excel Message Limit
 - Default value is 1000
 - Used to programming, rather than data entry issues
 - Remember about MS Excel limitation on number of rows
 - Consider to use CSV format and SAS if all messages are needed

Dictionaries and data



- ▶ MedDRA is available only after set-up
- ▶ Use a link to additional dictionaries

- ▶ Selection of Source data
 - Select, Add more files, Remove, Remove all
 - Drug-and-drop
- ▶ Define.xml file for data validation
 - Optional for XPT
 - Required for Dataset-XML
 - To ensure consistency with metadata

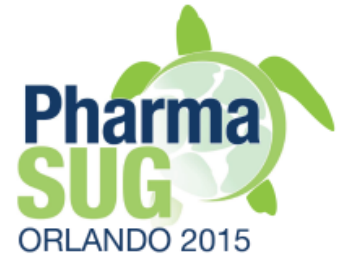
Exercise # 2



- ▶ Run validation of sample data from CDISC define.xml v2.0 package
 - What is a standard and version used?
 - Check MedDRA version
 - Issues interpretation Q&A

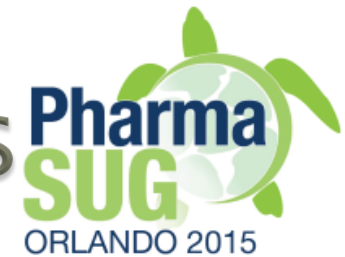
Overview of validation specifications

Metadata driven validation



- ▶ Validation specifications are extension of define.xml standard
 - Structure
 - domains, variables
 - Definition of rules
 - OpenCDISC specific syntax
 - Rule assignment to domain
 - ... \components\config\SDTM 3.1.2(FDA).xml

Structure and assignments

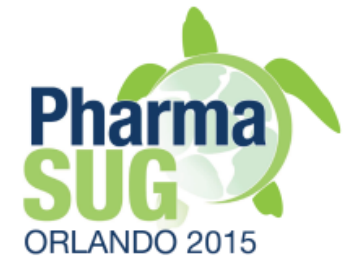


```
</ItemGroupDef>
<ItemGroupDef OID="IG.DM" Name="DM" Repeating="No" IsReferenceData="No" Purpose="Tabulation" def:Structure="One record per subject" def:C
<Description>
  <TranslatedText xml:lang="en">Demographics</TranslatedText>
</Description>
<ItemRef ItemOID="IT.DM.STUDYID" OrderNumber="1" Mandatory="Yes" KeySequence="1" Role="Identifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.DOMAIN" OrderNumber="2" Mandatory="Yes" Role="Identifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.USUBJID" OrderNumber="3" Mandatory="Yes" KeySequence="2" Role="Identifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.SUBJID" OrderNumber="4" Mandatory="Yes" Role="Topic" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.RFSTDTC" OrderNumber="5" Mandatory="No" Role="Record Qualifier" val:Core="Expected"/>
<ItemRef ItemOID="IT.DM.RFENDTC" OrderNumber="6" Mandatory="No" Role="Record Qualifier" val:Core="Expected"/>
<ItemRef ItemOID="IT.DM.SITEID" OrderNumber="7" Mandatory="Yes" Role="Record Qualifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.INVID" OrderNumber="8" Mandatory="No" Role="Record Qualifier" val:Core="Permissible"/>
<ItemRef ItemOID="IT.DM.INVNAM" OrderNumber="9" Mandatory="No" Role="Synonym Qualifier" val:Core="Permissible"/>
<ItemRef ItemOID="IT.DM.BRTHDTC" OrderNumber="10" Mandatory="No" Role="Record Qualifier" val:Core="Permissible"/>
<ItemRef ItemOID="IT.DM.AGE" OrderNumber="11" Mandatory="No" Role="Record Qualifier" val:Core="Expected"/>
<ItemRef ItemOID="IT.DM.AGEU" OrderNumber="12" Mandatory="No" Role="Variable Qualifier" val:Core="Expected"/>
<ItemRef ItemOID="IT.DM.SEX" OrderNumber="13" Mandatory="Yes" Role="Record Qualifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.RACE" OrderNumber="14" Mandatory="No" Role="Record Qualifier" val:Core="Expected"/>
<ItemRef ItemOID="IT.DM.ETHNIC" OrderNumber="15" Mandatory="No" Role="Record Qualifier" val:Core="Permissible"/>
<ItemRef ItemOID="IT.DM.ARMCD" OrderNumber="16" Mandatory="Yes" Role="Record Qualifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.ARM" OrderNumber="17" Mandatory="Yes" Role="Synonym Qualifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.COUNTRY" OrderNumber="18" Mandatory="Yes" Role="Record Qualifier" val:Core="Required"/>
<ItemRef ItemOID="IT.DM.DMXFN" OrderNumber="19" Mandatory="No" Role="Record Qualifier" val:Core="Model Permissible"/>
<ItemRef ItemOID="IT.DM.VISITNUM" OrderNumber="20" Mandatory="No" Role="Timing" val:Core="Model Permissible"/>
<ItemRef ItemOID="IT.DM.VISIT" OrderNumber="21" Mandatory="No" Role="Timing" val:Core="Model Permissible"/>
<ItemRef ItemOID="IT.DM.VISITDY" OrderNumber="22" Mandatory="No" Role="Timing" val:Core="Model Permissible"/>
<ItemRef ItemOID="IT.DM.DMDTC" OrderNumber="23" Mandatory="No" Role="Timing" val:Core="Permissible"/>
<ItemRef ItemOID="IT.DM.DMDY" OrderNumber="24" Mandatory="No" Role="Timing" val:Core="Permissible"/>
<ItemRef ItemOID="IT.DM.RFSTDY" OrderNumber="25" Mandatory="No" Role="Timing" val:Core="Model Permissible"/>
<ItemRef ItemOID="IT.DM.RFENDY" OrderNumber="26" Mandatory="No" Role="Timing" val:Core="Model Permissible"/>
<def:leaf ID="Location.DM" xlink:href="dm.xpt">
  <def:title>dm.xpt</def:title>
</def:leaf>
<val:ValidationRuleRef RuleID="CT2001" Active="Yes"/>
<val:ValidationRuleRef RuleID="CT2002" Active="Yes"/>
<val:ValidationRuleRef RuleID="CT2003" Active="Yes"/>
<val:ValidationRuleRef RuleID="CT2004" Active="Yes"/>
<val:ValidationRuleRef RuleID="CT2005" Active="Yes"/>
<val:ValidationRuleRef RuleID="CT2006" Active="Yes"/>
<val:ValidationRuleRef RuleID="SD0001" Active="Yes"/>
```

Structure

Check assignments

Rules definition



▼ val:Property ID=SD0001 PublisherID=FDAC014 Message=No records in data source Description=Domain table should have at least...	
▲ val:Required	
= ID	SD0002
= PublisherID	FDAC018
= Message	NULL value in %Variable% variable marked as Required
= Description	Required variables (where Core attribute is 'Req') cannot be NULL for any records
= Category	Presence
= Type	Error
= Variable	%Variables.Core:Required%
▲ val:Regex	
= ID	SD0003
= PublisherID	FDAC038
= Message	Invalid ISO 8601 value for %Variable% variable
= Description	Value of Dates/Time variables (*DTC) must conform to the ISO 8601 international standard
= Category	Format
= Type	Error
= Variable	%Variables[*DTC]%
= Test	%Variable.Type.Regex%
▼ val:Match ID=SD0004 PublisherID=FDAC056 Message=Inconsistent value for DOMAIN Description=Domain Abbreviation (DOMAIN)...	
▼ val:Unique ID=SD0005 PublisherID=FDAC044 Message=Duplicate value for %Domain%SEQ variable Description=The value of Seq...	
▼ val:Lookup ID=SD0006 PublisherID=FDAC113 Message=No baseline result in %P...	

“Grid view” in XML Spy

Config view in IE



OpenCDISC VALIDATOR

TABLE OF CONTENTS

- Global Rules Section (GLOBAL)
- Demographics (DM)
- Comments (CO)
- Subject Elements (SE)
- Subject Visits (SV)
- Interventions (INTERVENTIONS)
- Concomitant Medications (CM)
- Exposure (EX)
- Substance Use (SU)
- Events (EVENTS)
- Adverse Events (AE)
- Disposition (DS)
- Medical History (MH)
- Protocol Deviations (DV)
- Clinical Events (CE)
- Findings (FINDINGS)
- ECG Test Results (EG)
- Inclusion/Exclusion Criteria Not Met (IE)
- Laboratory Test Results (LB)
- Physical Examination (PE)
- Questionnaires (QS)
- Subject Characteristics (SC)
- Vital Signs (VS)
- Drug Accountability (DA)
- Microbiology Specimen (MB)
- Microbiology Susceptibility (MS)
- Pharmacokinetic Concentrations (PC)
- Pharmacokinetic Parameters (PP)
- Findings About Events or Interventions (FA)

Validator Configuration

Global Rules Section

Dataset Structure

Variable	Description	Required	Data Type	Length
STUDYID	Study Identifier	✓	text	200

Validation Rules

ID	Description	Validator	Message	Severity	Active
SD0061	Domains referenced in data definition document (define.xml) should be included in the submission	Find	Domain referenced in define.xml but dataset is missing	Warning	✓
SD0095	Supplemental Qualifiers special purpose dataset (SUPP--) can only be used to capture non-standard variables and their association to parent records in general-observation-class datasets (Events, Findings, Interventions) and Demographics	Condition	Invalid usage of SUPPQUAL for non-general-observation-class Domain	Error	✓
SD1020	Demographics (DM) dataset must be included in every submission	Find	Missing DM dataset	Error	✓
SD1061	Microbiology Specimen (MB) dataset should be included, when a Microbiology Susceptibility Test (MS) dataset is present	Find	Missing MB dataset, when MS dataset is present	Warning	✓
SD1063	Datasets included in study data must be described in the data definition document (define.xml)	Match	Dataset is not present in define.xml	Error	✓
SD1106	Adverse Events (AE) dataset should be included in every submission	Find	Missing AE dataset	Warning	✓

Exercise # 3



- ▶ Create validation specifications for MedDRA terms only
 - Make a copy of existing config file
 - Open by “plain” text editor. E.g., WordPad
 - Replace all *Active=“Yes”* to *Active=“No”*
 - Identify MedDRA checks
 - Set *Active=“Yes”* to MedDRA checks assignments only

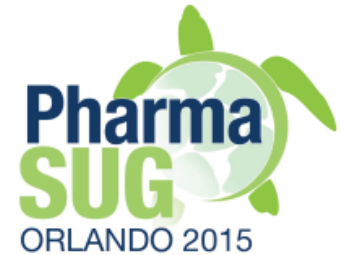
Validator CLI and SAS

Validator CLI



- ▶ Examples of use cases
 - Automation
 - Run Validator at the end of your SAS program as QC step
- ▶ Syntax documentation
 - <http://www.opencdisc.org/using-opencdisc-validator-cli>
 - Ensure to periodically check the page for updates
- ▶ SAS x command

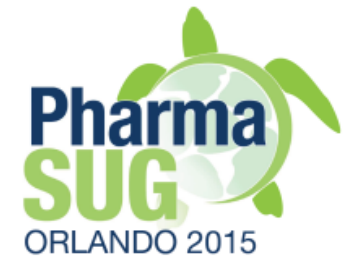
Example



x

```
java -jar
"C:\OpenCDISC\components\lib\validator-cli-2.0.1.jar"
-task=validate
-type=sdtm
-source:type=sas
-source="C:\InputData\*.xpt"
-config="C:\OpenCDISC\components\config\SDTM 3.1.2 (FDA).xml"
-config:cdisc=2014-03-28
-config:meddra=8.0
-report="C:\OpenCDISC\components\reports
\ValidationReport.xls"
-report:type=excel
-report:overwrite=yes
;
```

Exercises # 4, 5

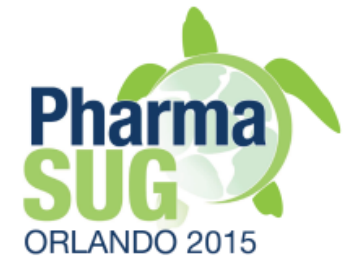


- ▶ Run OpenCDISC Validator from SAS, automate OC CLI command by adding dates into report, use SAS macro variables to specify input/output location and other parameters

- ▶ Create a SAS program to identify unknown MedDRA version used in study data
 - Create validation config for MedDRA terms only
 - Run validations with different versions of MedDRA
 - Find a validation report with no Errors
 - Optional: consider to use reports in SCV format, upload them into SAS to summarize validation findings

Converter

Data Converter



OpenCDISC Community

File View Help

Data Converter

convert data between SAS XPORT, Excel, CSV, and Dataset-XML

Convert Data

Source Format: SAS® Transport (XPORT) ▼

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](#)

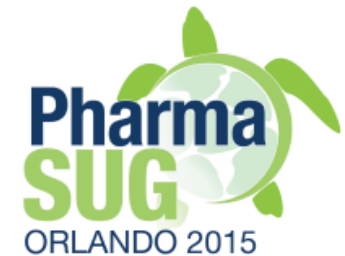
Output Format: Dataset XML ▼

Output Path: C:\OutputData [Browse...](#)

Define.xml: C:\InputData\define2-0-0-example-sdtm.xml [Browse...](#)

[Convert](#)

Input/Output



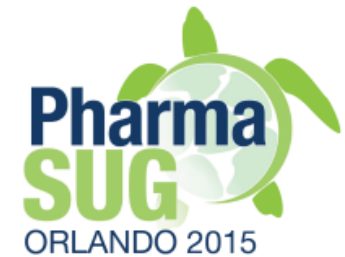
▶ Input

- SAS XPT

▶ Output

- MS Excel
 - 3 tabs: Dataset Metadata, Variables Metadata, Data records
 - Use cases
 - Clinical people without SAS skills
 - OS X users
- CSV
- Dataset–XML
 - Define.xml is required
 - Use existing one or create a basic define.xml with OpenCDISC Define.xml tool

Converter CLI example

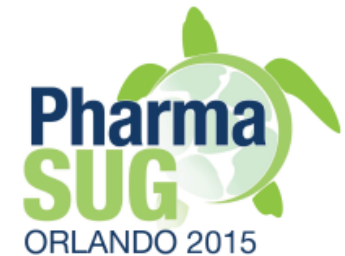


x

```
java
-jar C:\OpenCDISC\components\lib\data-converter-1.0.1.jar
-s=C:\InputData\*.xpt
-i=xpt
-o=C:\OutputData
-e=xml
-d=C:\InputData\define.xml
;
```

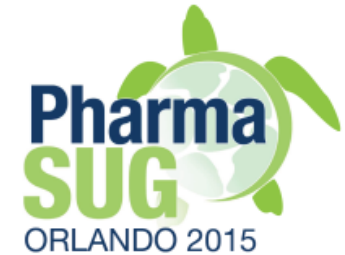
- It converts all SAS XPT files in C:\InputData folder into Dataset-XML format and put new files into C:\OutputData folder. Existing define.xml file is utilized and should be included into Dataset-XML data package.

Converter CLI parameters

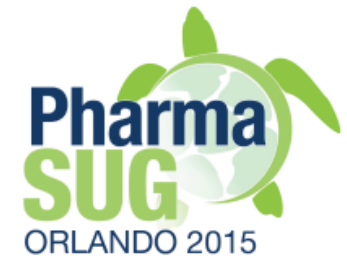


Parameter	Meaning	Description
-s	Source	Full path to source data files.
-i	Source Type	Data type of source files. Currently supported file types are [xpt] (default: xpt)
-o	Output	Full path to place output data files
-e	OutputType	Desired data type of output files. Currently supported file types are [xlsx, csv, xml] (default: xlsx)
-d	Define	For use in converting to Dataset-XML. Path to the define.xml for your datasets. If you do not have a define.xml you must specify the path to a configuration file so one can be generated.
-c	Config	For use in converting to Dataset-XML. Path to the configuration xml file for generating a define.xml. An incomplete define.xml will be generated based off of this configuration file. The configuration files are packaged with OpenCDISC Community.

Exercise # 6

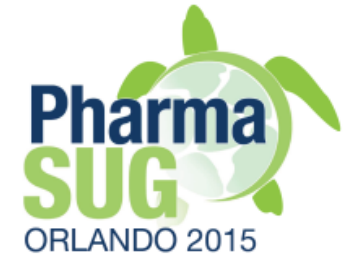


- ▶ Convert XPT data into Dataset–XML format and validate them

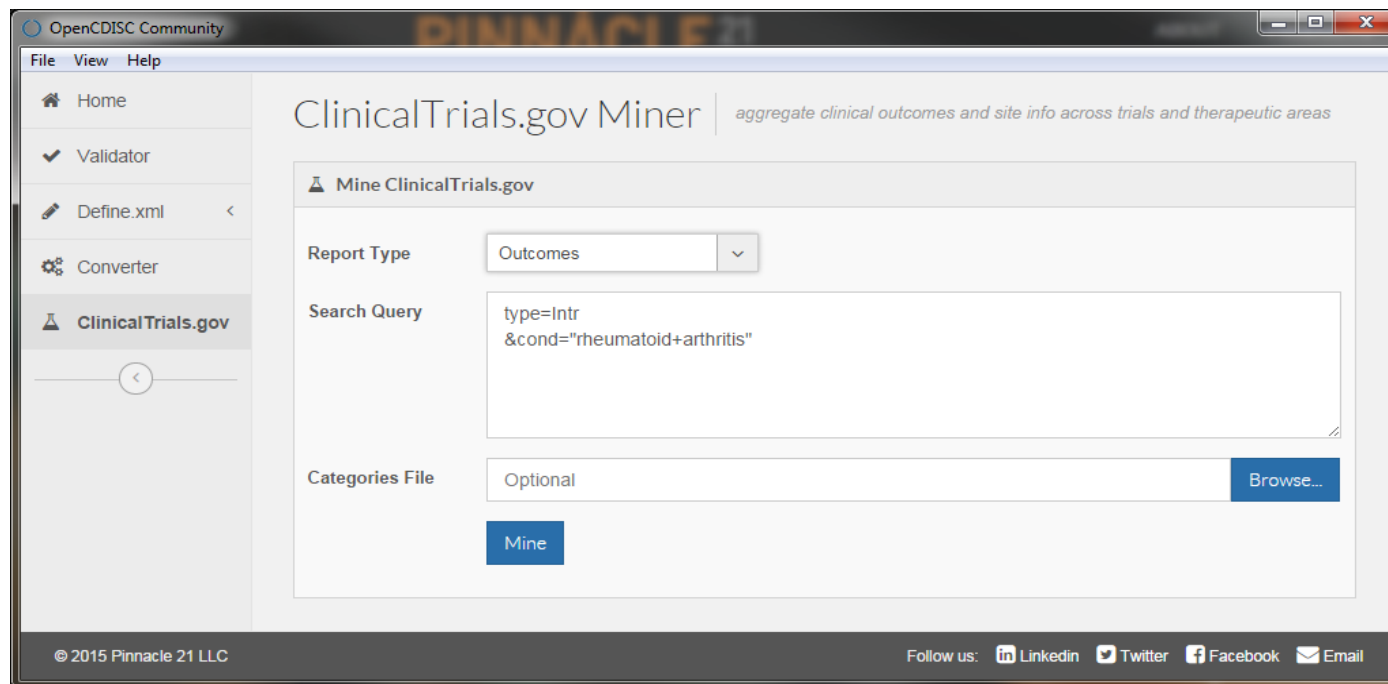


ClinicalTrials.gov tool

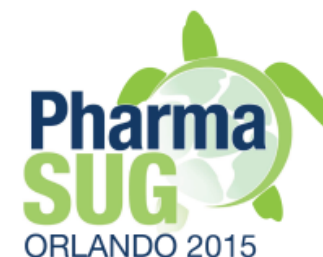
ClinicalTrials.gov tool



- ▶ Designed for clinical team
 - Return results of queries as summary table in MS Excel format, rather than separate XML files for Site or Trial



Outcome report

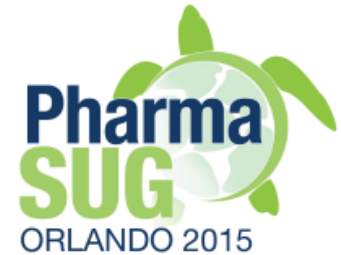


- ▶ One record per Study outcome with categories if applicable

Study_ID	Categorization	End_Point_Type	End_Point_Title	End_Point_Details	Conditions	Intervention
NCT00000395		Primary	Determine the effect of Foli		Rheumatoid Arthritis Adj	Methotrexate Folinic acid Folic acid
NCT00000395		Secondary	Determine the effect of folic		Rheumatoid Arthritis Adj	Methotrexate Folinic acid Folic acid
NCT00000395	DAS(28)	Secondary	Correlate disease activity wi		Rheumatoid Arthritis Adj	Methotrexate Folinic acid Folic acid
NCT00000401		Primary	Repeated measures analysis		Rheumatoid Arthritis	Oral bovine type II collagen
NCT00000401		Secondary	A Pearson correlation coeffic		Rheumatoid Arthritis	Oral bovine type II collagen
NCT00000416		Primary	job losses	Periods of work cessation	Rheumatoid Arthritis Syst	Rehabilitation counseling
NCT00000416		Secondary	limitation in ability to work	Extent of on the job limita	Rheumatoid Arthritis Syst	Rehabilitation counseling
NCT00000435	CLINICAL RESPONSE (ACR/EULAR)	Primary	Area under the curve or 'AUC		Rheumatoid Arthritis	dnaj peptide None-placebo
NCT00000435	CLINICAL RESPONSE (ACR/EULAR)	Secondary	Day 112 ACR 20 score		Rheumatoid Arthritis	dnaj peptide None-placebo
NCT00010335	MORTALITY/DEATH	Primary	Mortality		Systemic Sclerosis System	Stem Cell Transplantation CD34 select
NCT00010335		Secondary	Immune reconstitution, engi		Systemic Sclerosis System	Stem Cell Transplantation CD34 select
NCT00023205		Primary	Adherence to treatment	self report of medication	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Primary	Self efficacy	self report on self efficacy	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Primary	Satisfaction with medical car	self report of satisfaction	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Secondary	Heath Status	Health Assessment questi	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Secondary	Mental Health	SF-36 (5 items)	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Secondary	Understanding of medication	open ended questionnaire	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Secondary	Perceived usefulness of mat	Interview	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater
NCT00023205		Secondary	Appointment keeping	Clinical record	Rheumatoid Arthritis Pso	11th grade reading level arthritis mater

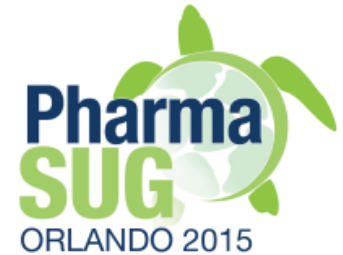
Define.xml tool

Overview



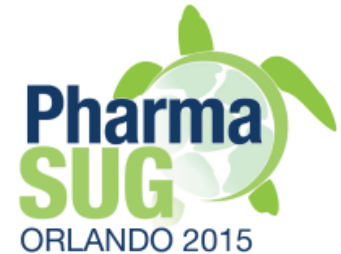
- ▶ Additional training materials
 - Training slides as reference
 - Videos on Youtube, Pinnacle 21 blog
- ▶ Solid knowledge if define.xml standard is expected

Define.xml versions



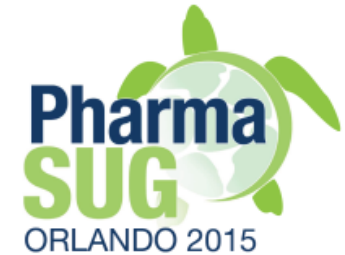
- ▶ OpenCDISC supports creation of define.xml v2.0 only
 - Define.xml v1.0 is outdated (>10y) standard, cannot correctly handle study metadata. E.g., Value Level
 - OpenCDISC provides migration of v1.0 to v2.0

XML requirements



- ▶ OpenCDISC provides interface to populate metadata content and do not worry about XML syntax
- ▶ However be aware of XML specific features
 - Case sensitive
 - E.g., “NO” is not the same as “No”
 - Special characters should be handled appropriately
 - “>” should be replaced by “>”
 - See XML documentation for details
 - Space character is still a real character
- ▶ Ensure consistency in ID values across Elements

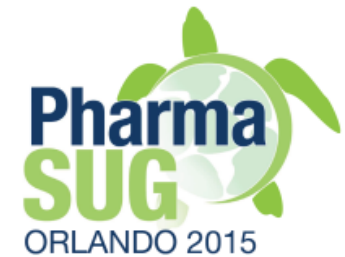
Descriptive vs. Prescriptive



- ▶ **Two basic approaches for creating define.xml**
 - Descriptive – after data is finalized
 - Most common one
 - Prescriptive – use define.xml for data collection, mapping

- ▶ **Process flow in Community version**
 - Scan data and create template in Excel format
 - Alternative – use existing define.xml v1.0
 - Populate missing metadata
 - Generate define.xml from completed specs

Generating specs



The screenshot shows the "Define.xml Generator" web application. The interface includes a sidebar with navigation options: Home, Validator, Define.xml (selected), Create Spec, Generate Define, Converter, and ClinicalTrials.gov. The main content area is titled "Define.xml Generator" and includes the subtitle "create compliant Define.xml 2.0 for SDTM, SEND, and ADaM datasets".

The "Create Excel Specification" section is active, showing "Extract Metadata from SAS XPORT Datasets". It features a table of source data files:

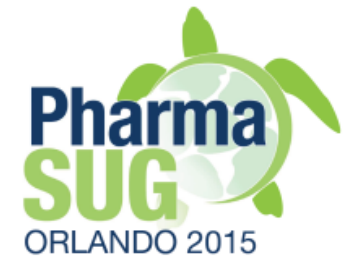
Source Data	File	Remove
	C:\InputData\ae.xpt	
	C:\InputData\cm.xpt	
	C:\InputData\da.xpt	
	C:\InputData\dm.xpt	
	C:\InputData\ds.xpt	
34 files Add more files		Remove all

Below the table, the "Configuration" is set to "SDTM 3.1.2 (FDA)". A "Create" button is visible.

The "Import Define.xml" section has an empty text field and a "Browse..." button.

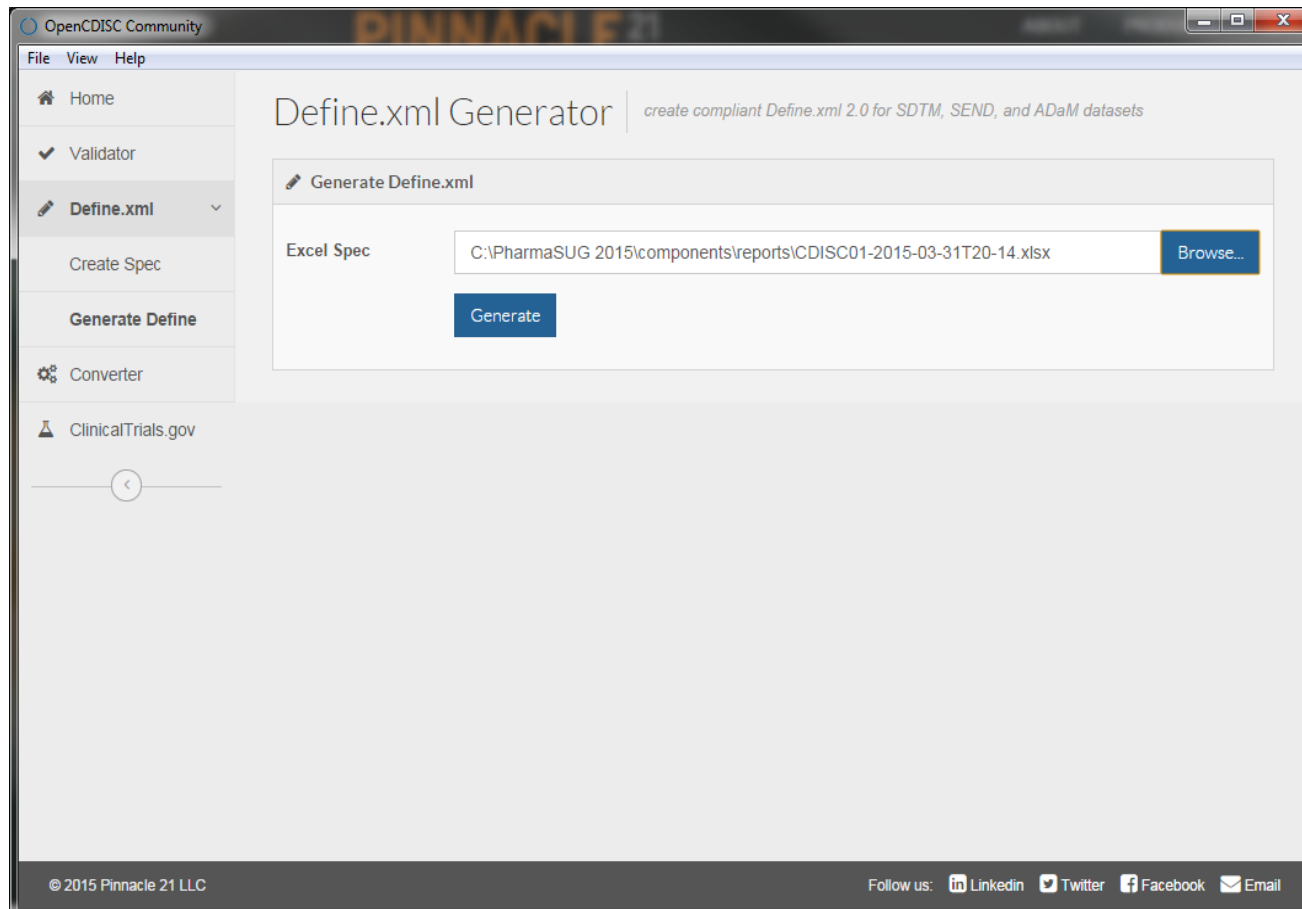
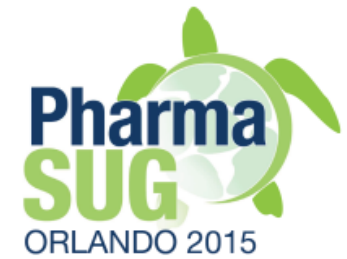
The "Define.xml" section has an empty text field and a "Create" button.

Excel specs



1	Order	Dataset	Variable	Label	Data Type	Length	Codelist	Origin	Pages	Method	Comm
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		

Generating Define.xml

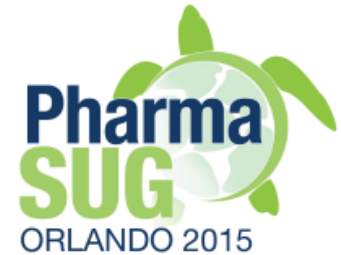


Elements and Attributes



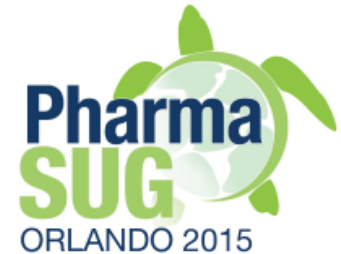
- ▶ There are dependencies between metadata
 - E.g., if Origin=CRF, then page should be populated
 - If Codelist is assigned then Codelist should be defined
 - See define.xml v2.0 standard for business rules and examples
 - We recommend using data package examples included into CDISC define.xml documentation. Run OpenCDISC to generate excel specs from high quality define.xml and use them as example
- ▶ Enterprise version scans CRFs, populates Codelists and Value Level from data scanning. It also includes versioning, comparison and other functionality.
- ▶ It would be helpful if you can replicate these features missing in Community version by SAS or other tools.

MS Excel ad data entry tool

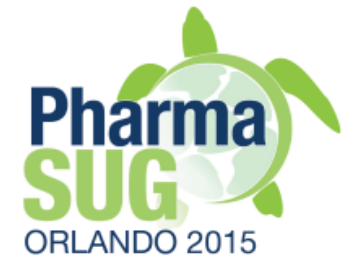


- ▶ Nice functionality
 - Filter, Sort, Hide column
- ▶ Useful functions
 - Concatenate, Vlookup, Exact, If
- ▶ Use Vlookup to merge external metadata.
 - E.g., CT, mapping specs, etc.
- ▶ Risks
 - Auto-correction. E.g., “ACN” -> “CAN”
 - Invisible trailing space characters in ID columns
 - Structural inconsistency in metadata may results in failing define.xml generation

Exercise #7

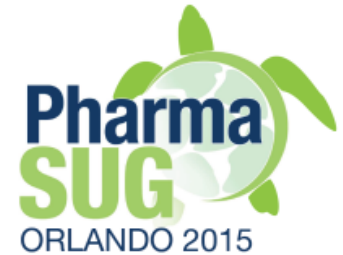


- ▶ Populate all expected metadata for DM domain
 - Scan data
 - Ensure that all codelists are present
 - Populate all codelist terms
 - Use SAS proc freq
 - Check CRF for missing terms
 - Populate NCI codes for standard terms
 - Populate Origin, Methods, Comments
 - Generate define.xml from updated specs
 - Validate define.xml and define.xml vs. data
 - Optional: populate metadata including Value Level for SUPPDM



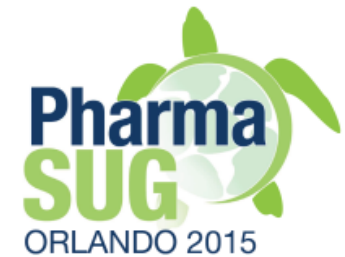
Summary

OpenCDISC Community

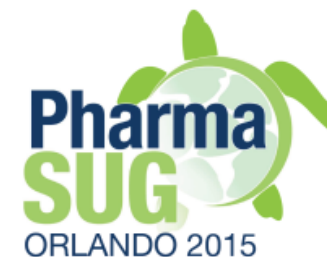


- ▶ Free and easy to use tool
- ▶ Utilized across industry including regulatory agencies
- ▶ Provides executable validation checks for FDA and CDISC official business rules
- ▶ Fully functioning basic Define.xml tool
- ▶ Data Converter
- ▶ May be automated
- ▶ Upgrade to Enterprise version is available

Useful Links



- ▶ <http://www.opencdisc.org>
- ▶ <http://www.pinnacle21.net>
- ▶ <http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm>
- ▶ <http://cdisc.org/standards-and-implementations>



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