PharmaSUG 2012 - Paper HW07

SDTM, ADaM and define.xml with OpenCDISC®

Matt Becker, PharmaNet/i3, Cary, NC

ABSTRACT

Standards are an ongoing focus of the health care and life science industry. Common terms you will see and hear during industry conferences include: “SDTM,” “ADaM,” “ODM,” “LAB,” “SEND,” and “define(.xml/.pdf).” What do these terms mean? How do we create and validate these standards before submission to a client OR the FDA? Is there an easier way to ensure compliance? As a user, many of us have spent hours reading the SDTM/ADaM standards and implementation guides to generate “compliant” SAS data sets for our clinical studies. We have spent countless hours having another user QC our data structures...but is there an easier way?

Each of these terms and subsequent standards encompass a set of metadata. Metadata is “data about data.” For example, a Proc Contents of a data set produces a list of variables, variable type, variable length, variable format, etc. Each of these tidbits of information is metadata about that individual data set. SDTM defines standard metadata for “domains.” Similarly, ADaM defines standard metadata for analysis data sets. If that metadata is standard it should be feasible to construct an application that will check your data’s metadata against that standard’s metadata.

OpenCDISC is an open source community which is focusing on creating frameworks and tools for the implementation and advancement of CDISC Standards. OpenCDISC has created a CDISC Validator which will eliminate the need for individuals to develop their own custom processes in order to ensure that their CDISC models are compliant with CDISC standards. By taking common validation rules, OpenCDISC has developed an open-source tool which is freely available and of commercial-quality to ensure data compliance with CDISC models such as SDTM, ADaM, and Define.xml. The validation rules for each standard have been pooled into a CDISC Validation Rules Repository, providing users with a central listing. The listing is easy to use, contribute to, improve on and continue development.

In this Hands-On Workshop (HOW) we are going to briefly describe a few of the key terms (SDTM, ADaM, define) and investigate the use of OpenCDISC Validator to perform the following tasks:

- Validate SDTM 3.1.1 SAS data sets
- Validate SDTM 3.1.2 SAS data sets
- Validate ADaM 1.0 SAS data sets
- Generate define.xml

INTRODUCTION

Before going too much further with the discussion of OpenCDISC, a quick overview of CDISC (Clinical Data Interchange Standards Consortium) is merited. Over the past few years CDISC has become common terminology in our workplace and we have started to use CDISC standards in our work more and more. The CDISC standards provide data consistency across the spectrum and this standardization has helped streamline drug development.

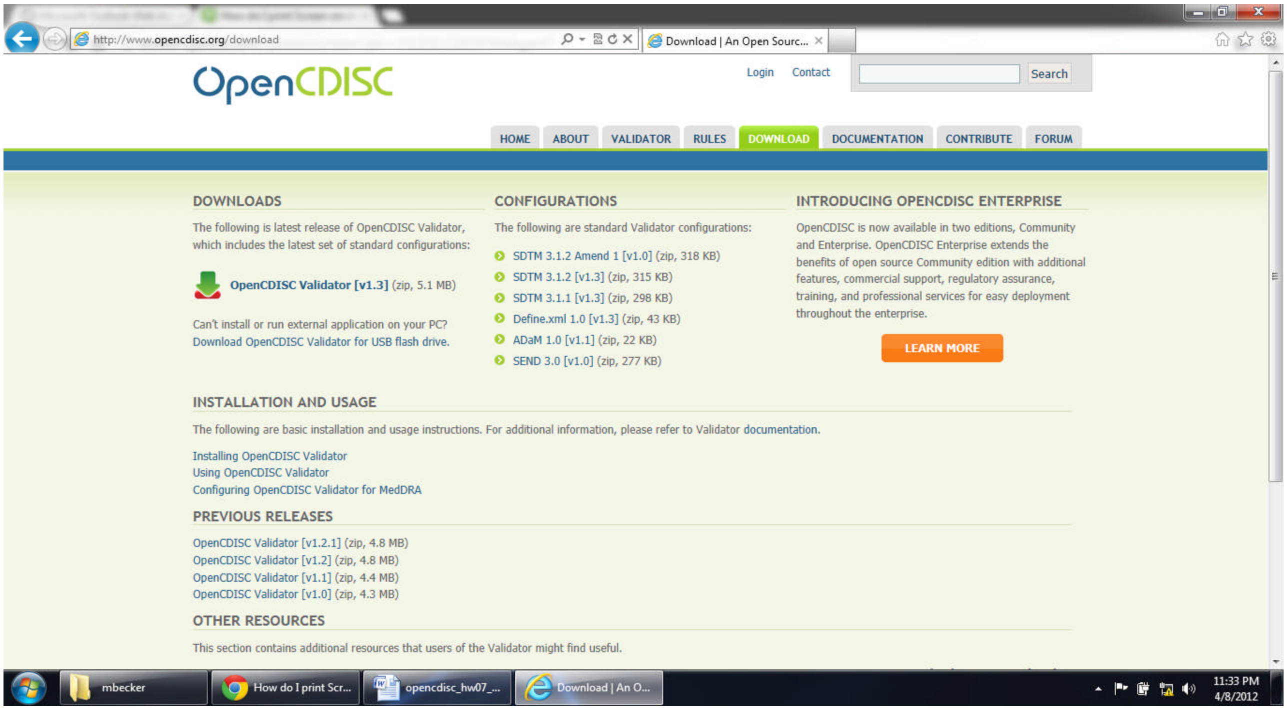
In this paper, we are going to concentrate on the SDTM, ADaM, and define.xml CDISC standards. SDTM (Study Data Tabulation Models) is the content standard of case report form data tabulations from clinical research studies. ADaM (Analysis Data Model) is the content standard of analysis datasets. Define.xml (Case Report Tabulation Data Definition Specification (CRTDDS)) is an XML-based content and format standard which contains the specifications for data definitions for CDISC SDTM datasets.

When we create SDTM files, ADaM files, and/or define.xml, we must make sure that they are compliant with CDISC standards. We must check our work. How is this done? Usually, by double programming (at least in the case of SDTM and ADaM files). The re-creation of the files by an independent programmer and comparing the two sets of results. This is no simple task. It requires a lot of time and a lot of reconciliation between the production programming and the validation programming in order to make sure there is compliance with the CDISC standards. And, once this process is complete, how can we guarantee 100% compliance? The individualized validation process for compliance with the CDISC standards is not a standardized task; each of us develops our own ways of validating our files.

Here is where OpenCDISC comes into the picture. OpenCDISC has created a CDISC Validator which will eliminate the need for individuals to develop their own custom processes. The OpenCDISC Validator ensures that your CDISC models are compliant with CDISC standards. OpenCDISC has taken common validation rules and pooled them into a CDISC Validation Rules Repository providing users with a central listing. The Validator is free and easy to use.

USING THE OPENCDISC VALIDATOR

The validator requires Java Runtime Environment (JRE) version 1.5 or higher and 2GB system RAM. Download the OpenCDISC validator from http://www.opencdisc.org, click on the OpenCDISC Validator [v1.3] link and unzip to your chosen directory. Detailed installation directions are provided on the website.



DOWNLOADS

The following is latest release of OpenCDISC Validator, which includes the latest set of standard configurations:

OpenCDISC Validator [v1.3] (zip, 5.1 MB)

Can't install or run external application on your PC? Download OpenCDISC Validator for USB flash drive.

CONFIGURATIONS

The following are standard Validator configurations:

- SDTM 3.1.2 Amend 1 [v1.0] (zip, 318 KB)
- SDTM 3.1.2 [v1.3] (zip, 315 KB)
- SDTM 3.1.1 [v1.3] (zip, 298 KB)
- Define.xml 1.0 [v1.3] (zip, 43 KB)
- ADaM 1.0 [v1.1] (zip, 22 KB)
- SEND 3.0 [v1.0] (zip, 277 KB)

INTRODUCING OPENCDISC ENTERPRISE

OpenCDISC is now available in two editions, Community and Enterprise. OpenCDISC Enterprise extends the benefits of open source Community edition with additional features, commercial support, regulatory assurance, training, and professional services for easy deployment throughout the enterprise.

INSTALLATION AND USAGE

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

Installing OpenCDISC Validator
Using OpenCDISC Validator
Configuring OpenCDISC Validator for MedDRA

PREVIOUS RELEASES

OpenCDISC Validator [v1.2.1] (zip, 4.8 MB)
OpenCDISC Validator [v1.2] (zip, 4.8 MB)
OpenCDISC Validator [v1.1] (zip, 4.4 MB)
OpenCDISC Validator [v1.0] (zip, 4.3 MB)

OTHER RESOURCES

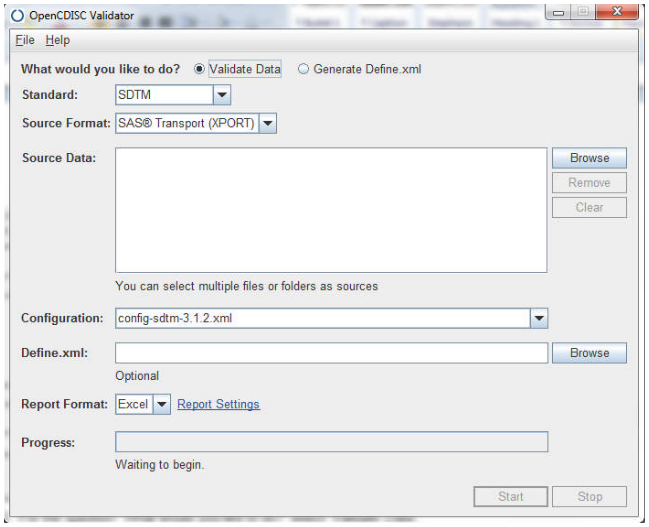
This section contains additional resources that users of the Validator might find useful.

Once the validator has been downloaded and unzipped, it is ready to use.

VALIDATING SDTM FILES

Step 1: Open the 'opendisc-validator' folder.

Step 2: Double click on the ‘client.bat’ file. This will bring up the OpenCDISC Validator window:



OpenCDISC Validator

What would you like to do? ● Validate Data ○ Generate Define.xml

Standard: SDTM

Source Format: SAS® Transport (XPORT)

Source Data:

You can select multiple files or folders as sources

Configuration: config-sdtm-3.1.2.xml

Define.xml:

Optional

Report Format: Excel

Progress:

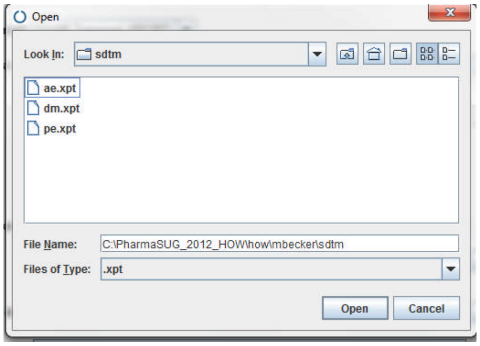
Waiting to begin.

Step 3: For the question “What would you like to do?’ select ‘Validate Data’.

Step 4: Choose the Standard (the default is SDTM). For this example, we chose SDTM.

Step 5: Choose the Format (the default is XPORT). Note that the SDTM files must be in SAS® Transport (XPORT) or a delimited file. The validator cannot process regular SAS datasets.

Step 6: Choose the source data by clicking on the Browse button on the right hand side. The following window will appear once you change the directory to a location that contains SAS XPT files:



Open

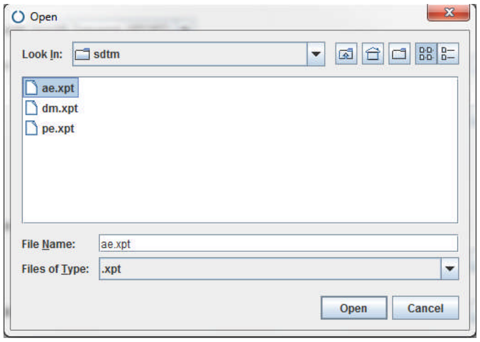
Look In: sdtm

ae.xpt
dm.xpt
pe.xpt

File Name: C:\PharmaSUG_2012_HOW\how\mbecker\sdtm

Files of Type: .xpt

Step 7: Highlight the SDTM file or files you want to check.



Open

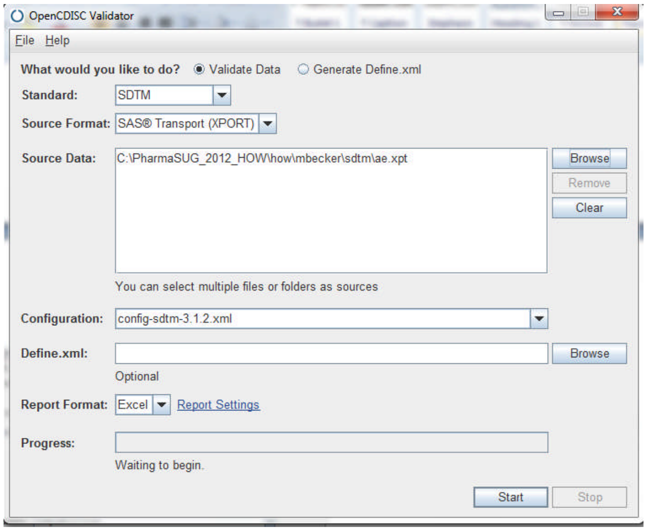
Look In: sdtm

ae.xpt
dm.xpt
pe.xpt

File Name: ae.xpt

Files of Type: .xpt

Step 8: Click Open. The OpenCDISC validator window will appear with the files or files you have selected in the Source Data field.



What would you like to do? ● Validate Data ○ Generate Define.xml

Standard: SDTM

Source Format: SAS® Transport (XPORT)

Source Data: C:\PharmaSUG_2012_HOW\how\mbecker\sdtm\ae.xpt

You can select multiple files or folders as sources

Configuration: config-sdtm-3.1.2.xml

Define.xml:

Optional

Report Format: Excel

Progress:

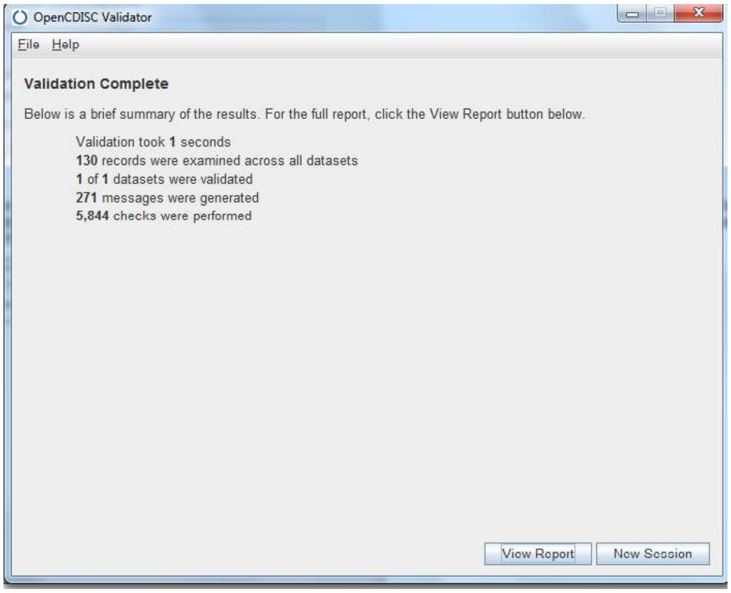
Waiting to begin.

Step 9: Choose the configuration. The default for SDTM files is sdtm-3.1.2.xml.

Step 10: Choose the report format. The default is Excel.

Step 11: In this example, we are now ready to start the validation of the AE SDTM file. Click the start button.

When the validation is complete, you will receive an information window providing you with how long it took for the validator to run, the number of records read, the number of datasets validated, the number of messages generated and the number of checks performed:



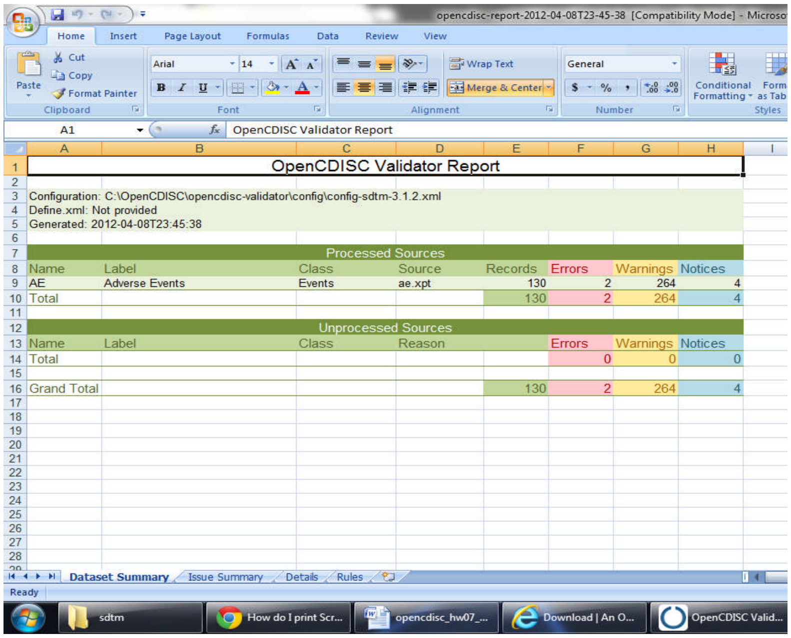
Validation Complete

Below is a brief summary of the results. For the full report, click the View Report button below.

Validation took **1** seconds
130 records were examined across all datasets
1 of **1** datasets were validated
271 messages were generated
5,844 checks were performed

Step 12: At this point, you may choose to view the report or start a new session. Choose ‘View Report’. The report will consist of 4 tabs within the Excel document: Dataset Summary, Issue Summary, Details and Rules.

The **Dataset Summary** tab provides a brief overview of what was encountered by the validator.



OpenCDISC Validator Report

Configuration: C:\OpenCDISC\opencdisc-validator\config\config-sdtm-3.1.2.xml
Define.xml: Not provided
Generated: 2012-04-08T23:45:38

Processed Sources

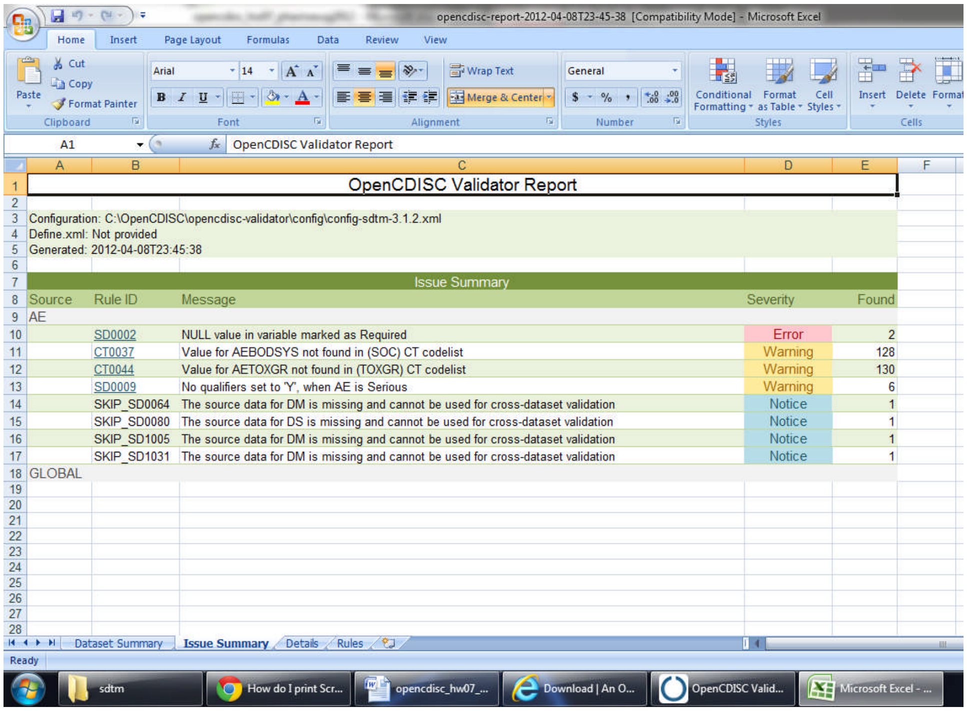
Name	Label	Class	Source	Records	Errors	Warnings	Notices
AE	Adverse Events	Events	ae.xpt	130	2	264	4
Total				130	2	264	4

Unprocessed Sources

Name	Label	Class	Reason		Errors	Warnings	Notices
Total					0	0	0
Grand Total				130	2	264	4

The report tells us that there are 130 records in the AE file that 2 errors were encountered, and 264 Warnings were reported.

The **Issues Summary** tab provides break down of the type of rules that have issues and how many have been reported.



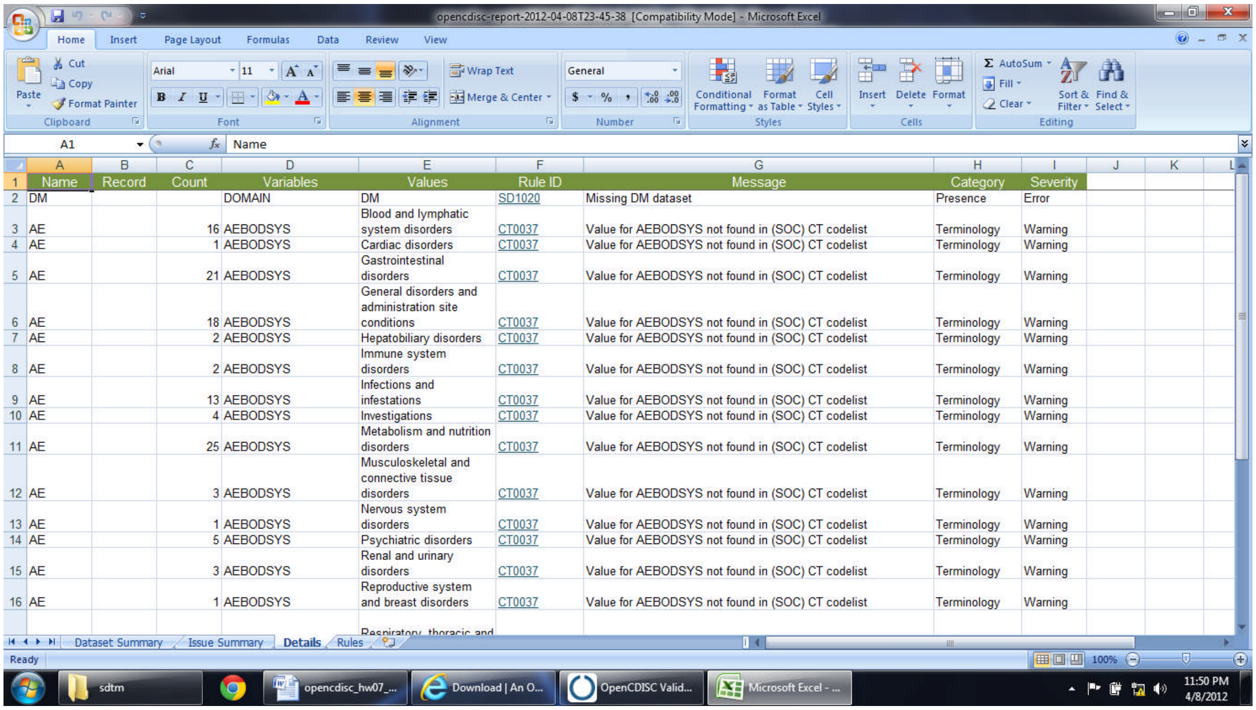
OpenCDISC Validator Report

Configuration: C:\OpenCDISC\opencdisc-validator\config\config-sdtm-3.1.2.xml
Define.xml: Not provided
Generated: 2012-04-08T23:45:38

Issue Summary

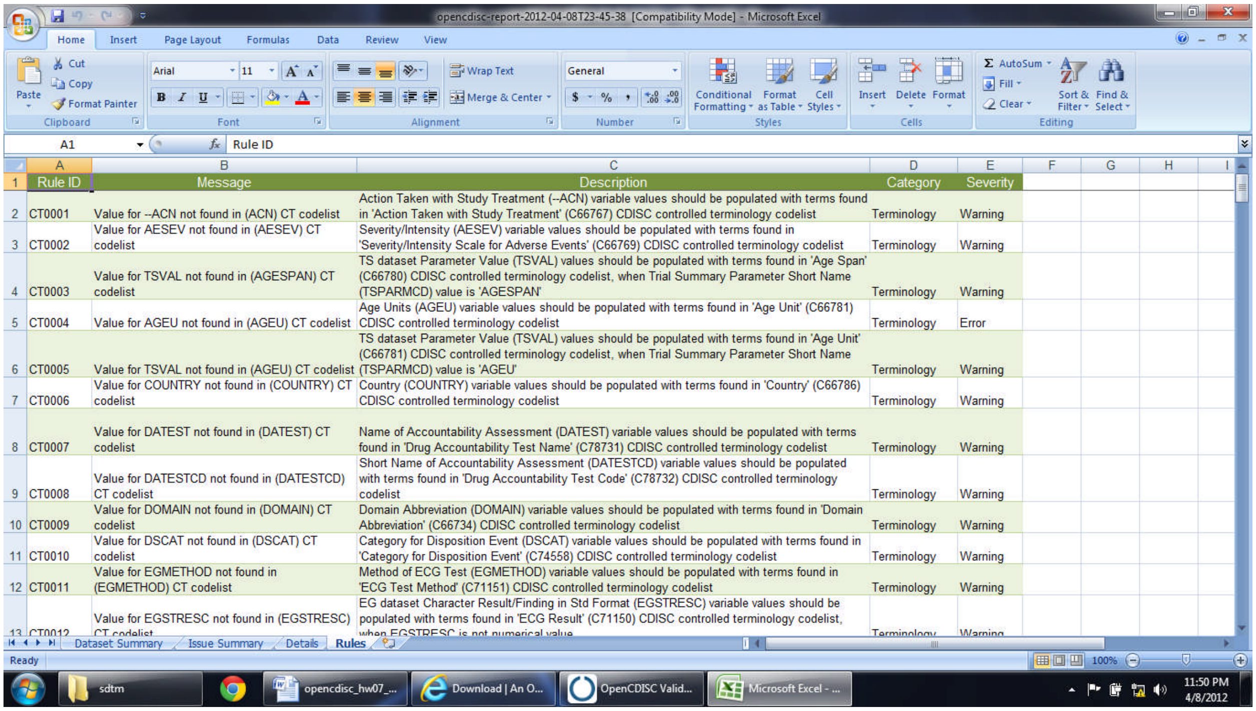
Source	Rule ID	Message	Severity	Found
AE				
	SD0002	NULL value in variable marked as Required	Error	2
	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Warning	128
	CT0044	Value for AETOXGR not found in (TOXGR) CT codelist	Warning	130
	SD0009	No qualifiers set to 'Y', when AE is Serious	Warning	6
	SKIP_SD0064	The source data for DM is missing and cannot be used for cross-dataset validation	Notice	1
	SKIP_SD0080	The source data for DS is missing and cannot be used for cross-dataset validation	Notice	1
	SKIP_SD1005	The source data for DM is missing and cannot be used for cross-dataset validation	Notice	1
	SKIP_SD1031	The source data for DM is missing and cannot be used for cross-dataset validation	Notice	1
GLOBAL				

The **Details** tab provides us just that…the details. Each error or warning message has been expanded. Below is a small sample of what this AE file's issues log generated:



Name	Record	Count	Variables	Values	Rule ID	Message	Category	Severity
DM			DOMAIN	DM	SD1020	Missing DM dataset	Presence	Error
AE		16	AEBODSYS	Blood and lymphatic system disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		1	AEBODSYS	Cardiac disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		21	AEBODSYS	Gastrointestinal disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		18	AEBODSYS	General disorders and administration site conditions	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		2	AEBODSYS	Hepatobiliary disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		2	AEBODSYS	Immune system disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		13	AEBODSYS	Infections and infestations	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		4	AEBODSYS	Investigations	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		25	AEBODSYS	Metabolism and nutrition disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		3	AEBODSYS	Musculoskeletal and connective tissue disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		1	AEBODSYS	Nervous system disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		5	AEBODSYS	Psychiatric disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		3	AEBODSYS	Renal and urinary disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		1	AEBODSYS	Reproductive system and breast disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
				Respiratory, thoracic and				

And finally, the **Rules** tab shows us the standards. Here is a small sample:



Rule ID	Message	Description	Category	Severity
CT0001	Value for --ACN not found in (ACN) CT codelist	Action Taken with Study Treatment (--ACN) variable values should be populated with terms found in 'Action Taken with Study Treatment' (C66767) CDISC controlled terminology codelist	Terminology	Warning
CT0002	Value for AESEV not found in (AESEV) CT codelist	Severity/Intensity (AESEV) variable values should be populated with terms found in 'Severity/Intensity Scale for Adverse Events' (C66769) CDISC controlled terminology codelist	Terminology	Warning
CT0003	Value for TSVAL not found in (AGESPAN) CT codelist	TS dataset Parameter Value (TSVAL) values should be populated with terms found in 'Age Span' (C66780) CDISC controlled terminology codelist, when Trial Summary Parameter Short Name (TSPARMCD) value is 'AGESPAN'	Terminology	Warning
CT0004	Value for AGEU not found in (AGEU) CT codelist	Age Units (AGEU) variable values should be populated with terms found in 'Age Unit' (C66781) CDISC controlled terminology codelist	Terminology	Error
CT0005	Value for TSVAL not found in (AGEU) CT codelist	TS dataset Parameter Value (TSVAL) values should be populated with terms found in 'Age Unit' (C66781) CDISC controlled terminology codelist, when Trial Summary Parameter Short Name (TSPARMCD) value is 'AGEU'	Terminology	Warning
CT0006	Value for COUNTRY not found in (COUNTRY) CT codelist	Country (COUNTRY) variable values should be populated with terms found in 'Country' (C66786) CDISC controlled terminology codelist	Terminology	Warning
CT0007	Value for DATEST not found in (DATEST) CT codelist	Name of Accountability Assessment (DATEST) variable values should be populated with terms found in 'Drug Accountability Test Name' (C78731) CDISC controlled terminology codelist	Terminology	Warning
CT0008	Value for DATESTCD not found in (DATESTCD) CT codelist	Short Name of Accountability Assessment (DATESTCD) variable values should be populated with terms found in 'Drug Accountability Test Code' (C78732) CDISC controlled terminology codelist	Terminology	Warning
CT0009	Value for DOMAIN not found in (DOMAIN) CT codelist	Domain Abbreviation (DOMAIN) variable values should be populated with terms found in 'Domain Abbreviation' (C66734) CDISC controlled terminology codelist	Terminology	Warning
CT0010	Value for DSCAT not found in (DSCAT) CT codelist	Category for Disposition Event (DSCAT) variable values should be populated with terms found in 'Category for Disposition Event' (C74558) CDISC controlled terminology codelist	Terminology	Warning
CT0011	Value for EGMETHOD not found in (EGMETHOD) CT codelist	Method of ECG Test (EGMETHOD) variable values should be populated with terms found in 'ECG Test Method' (C71151) CDISC controlled terminology codelist	Terminology	Warning
CT0012	Value for EGSTRESC not found in (EGSTRESC) CT codelist	EG dataset Character Result/Finding in Std Format (EGSTRESC) variable values should be populated with terms found in 'ECG Result' (C71150) CDISC controlled terminology codelist, when EGSTRESC is not numerical value	Terminology	Warning

With the information obtained from the OpenCDISC validator, the user can now go back to the production SDTM file and correct any issues the validator has flagged.

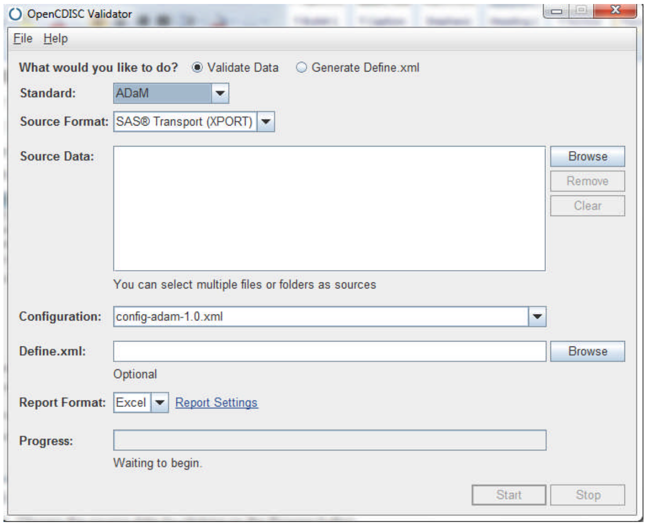
VALIDATING ADAM FILES

Step 1: Open the ‘opendisc-validator’ folder.

Step 2: Double click on the ‘client.bat’ file. This will bring up the OpenCDISC Validator window.

Step 3: For the question ‘What would you like to do?’ select ‘Validate Data’.

Step 4: Choose the Standard (the default is SDTM). For this example, we chose **ADaM**.



What would you like to do? ● Validate Data ○ Generate Define.xml

Standard: ADaM

Source Format: SAS® Transport (XPORT)

Source Data:

You can select multiple files or folders as sources

Configuration: config-adam-1.0.xml

Define.xml:

Optional

Report Format: Excel

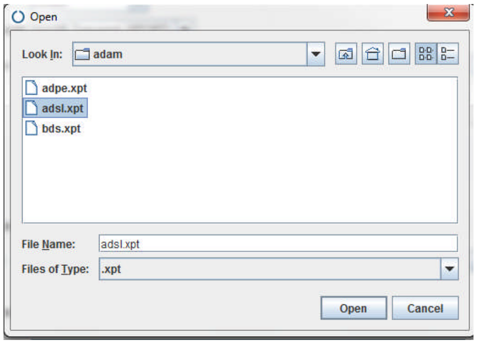
Progress:

Waiting to begin.

Step 5: Choose the Format (the default is XPORT).

Step 6: Choose the source data by clicking on the Browse button.

Step 7: Highlight the ADaM file or files you want to check.



Look In: adam

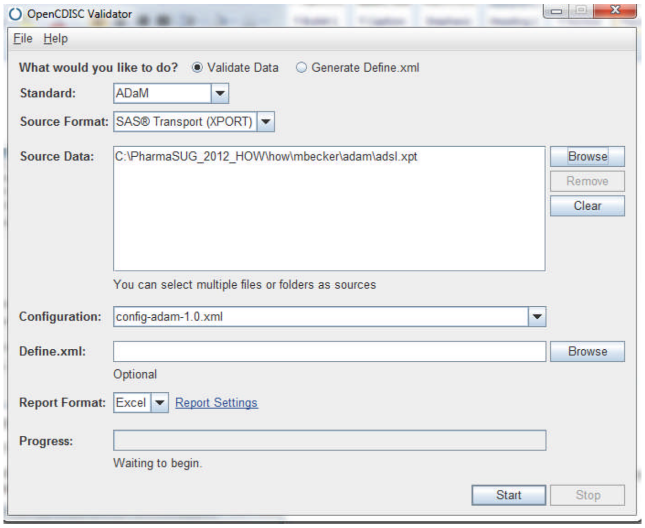
adpe.xpt
adsl.xpt
bds.xpt

File Name: adsl.xpt

Files of Type: .xpt

Step 8: Click Open. The OpenCDISC validator window will appear with the files or files you have selected in the Source Data field.

Step 9: Choose the configuration. Select ‘config-adam-1.0.xml’. This is the default for ADaM files, and currently your only choice.



What would you like to do? ● Validate Data ○ Generate Define.xml

Standard: ADaM

Source Format: SAS® Transport (XPORT)

Source Data: C:\PharmaSUG_2012_HOW\how\mbecker\adam\adsl.xpt

You can select multiple files or folders as sources

Configuration: config-adam-1.0.xml

Define.xml:

Optional

Report Format: Excel

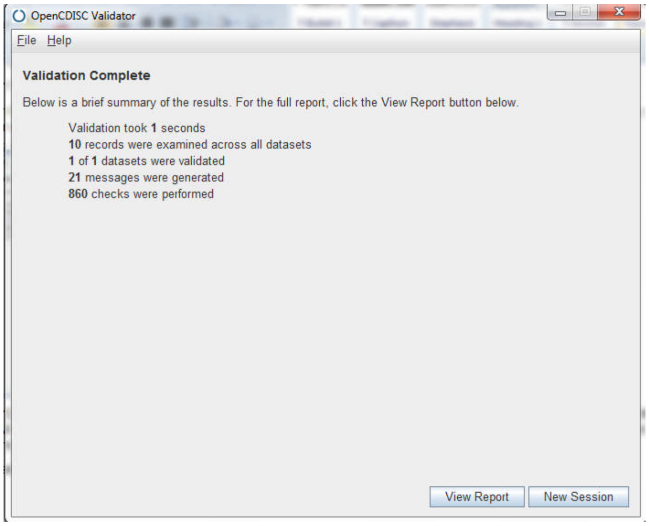
Progress:

Waiting to begin.

Step 10: Choose the report format. The default is Excel.

Step 11: In this example, we are now ready to start the validation of the ADSL ADaM file. Click the start button.

When the validation is complete, you will receive the information window providing you the same type of summary information as when we ran the STDM file validation.



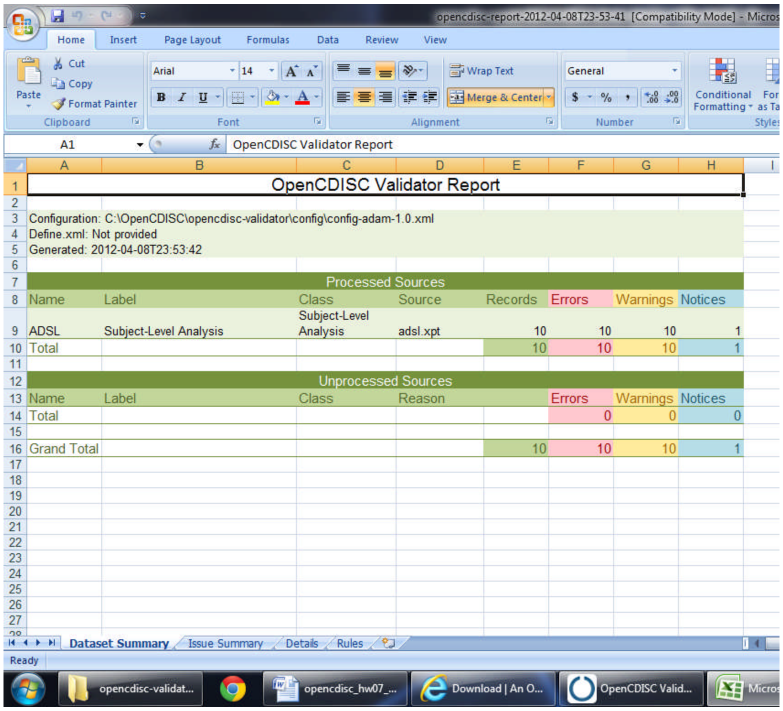
Validation Complete

Below is a brief summary of the results. For the full report, click the View Report button below.

Validation took **1** seconds
10 records were examined across all datasets
1 of **1** datasets were validated
21 messages were generated
860 checks were performed

Step 12: At this point, you may choose to view the report or start a new session. Choose ‘View Report’. As with the SDTM validation, the ADaM validation provides a similar OpenCDISC Validator Report containing 4 tabs: Dataset Summary, Issue Summary, Details and Rules.

Dataset Summary:



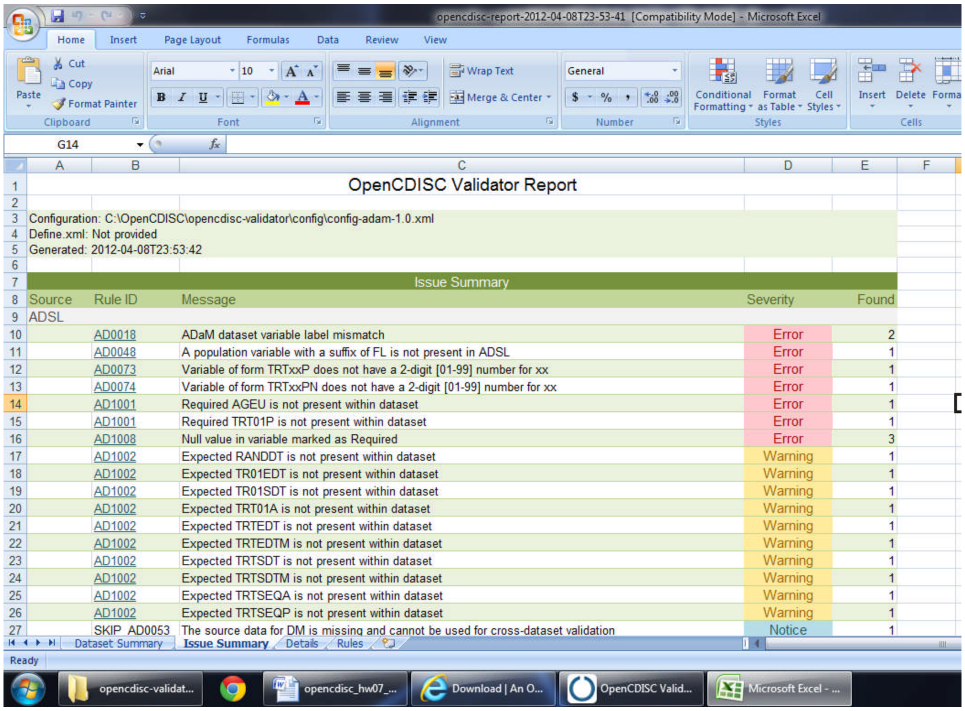
OpenCDISC Validator Report

Configuration: C:\OpenCDISC\opencdisc-validator\config\config-adam-1.0.xml
Define.xml: Not provided
Generated: 2012-04-08T23:53:42

Processed Sources							
Name	Label	Class	Source	Records	Errors	Warnings	Notices
ADSL	Subject-Level Analysis	Subject-Level Analysis	adsl.xpt	10	10	10	1
Total				10	10	10	1

Unprocessed Sources							
Name	Label	Class	Reason		Errors	Warnings	Notices
Total					0	0	0
Grand Total				10	10	10	1

Issue Summary:

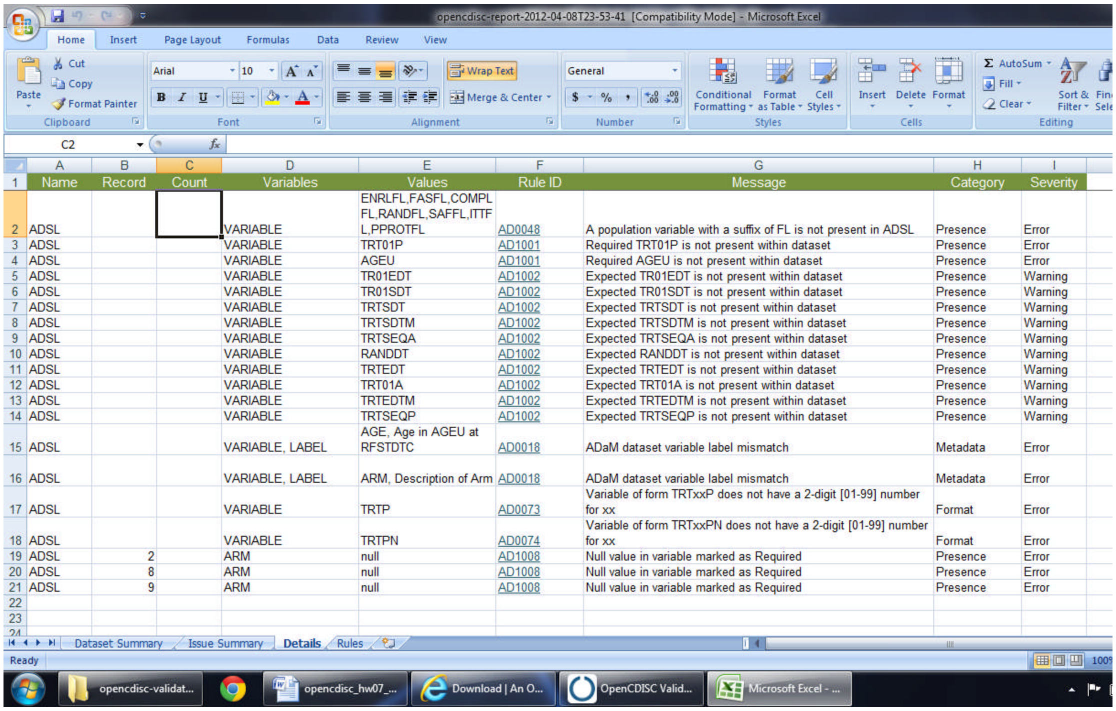


OpenCDISC Validator Report

Configuration: C:\OpenCDISC\opencdisc-validator\config\config-adam-1.0.xml
Define.xml: Not provided
Generated: 2012-04-08T23:53:42

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADSL				
	AD0018	ADaM dataset variable label mismatch	Error	2
	AD0048	A population variable with a suffix of FL is not present in ADSL	Error	1
	AD0073	Variable of form TRTxxP does not have a 2-digit [01-99] number for xx	Error	1
	AD0074	Variable of form TRTxxPN does not have a 2-digit [01-99] number for xx	Error	1
	AD1001	Required AGEU is not present within dataset	Error	1
	AD1001	Required TRT01P is not present within dataset	Error	1
	AD1008	Null value in variable marked as Required	Error	3
	AD1002	Expected RANDDT is not present within dataset	Warning	1
	AD1002	Expected TR01EDT is not present within dataset	Warning	1
	AD1002	Expected TR01SDT is not present within dataset	Warning	1
	AD1002	Expected TRT01A is not present within dataset	Warning	1
	AD1002	Expected TRTEDT is not present within dataset	Warning	1
	AD1002	Expected TRTEDTM is not present within dataset	Warning	1
	AD1002	Expected TRTSDT is not present within dataset	Warning	1
	AD1002	Expected TRTSDTM is not present within dataset	Warning	1
	AD1002	Expected TRTSEQA is not present within dataset	Warning	1
	AD1002	Expected TRTSEQP is not present within dataset	Warning	1
	SKIP AD0053	The source data for DM is missing and cannot be used for cross-dataset validation	Notice	1

Details:

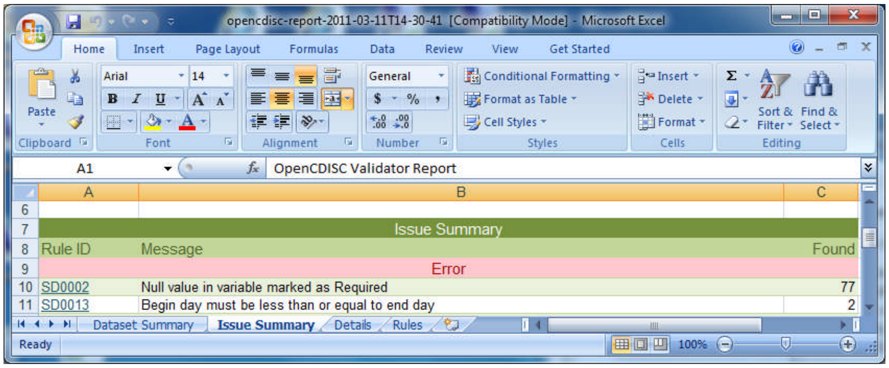

opencdisc-report-2012-04-08T23-53-41 [Compatibility Mode] - Microsoft Excel
C2
Dataset Summary
Issue Summary
Details
Rules
Ready

Name	Record	Count	Variables	Values	Rule ID	Message	Category	Severity
ADSL			VARIABLE	ENRLFL,FASFL,COMPLFL,RANDFL,SAFFL,ITTFL,PPROTFL	AD0048	A population variable with a suffix of FL is not present in ADSL	Presence	Error
ADSL			VARIABLE	TRT01P	AD1001	Required TRT01P is not present within dataset	Presence	Error
ADSL			VARIABLE	AGEU	AD1001	Required AGEU is not present within dataset	Presence	Error
ADSL			VARIABLE	TR01EDT	AD1002	Expected TR01EDT is not present within dataset	Presence	Warning
ADSL			VARIABLE	TR01SDT	AD1002	Expected TR01SDT is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRTSDT	AD1002	Expected TRTSDT is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRTSDTM	AD1002	Expected TRTSDTM is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRTSEQA	AD1002	Expected TRTSEQA is not present within dataset	Presence	Warning
ADSL			VARIABLE	RANDDT	AD1002	Expected RANDDT is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRTEDT	AD1002	Expected TRTEDT is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRT01A	AD1002	Expected TRT01A is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRTEDTM	AD1002	Expected TRTEDTM is not present within dataset	Presence	Warning
ADSL			VARIABLE	TRTSEQP	AD1002	Expected TRTSEQP is not present within dataset	Presence	Warning
ADSL			VARIABLE, LABEL	AGE, Age in AGEU at RFSTDTC	AD0018	ADaM dataset variable label mismatch	Metadata	Error
ADSL			VARIABLE, LABEL	ARM, Description of Arm	AD0018	ADaM dataset variable label mismatch	Metadata	Error
ADSL			VARIABLE	TRTP	AD0073	Variable of form TRTxxP does not have a 2-digit [01-99] number for xx	Format	Error
ADSL			VARIABLE	TRTPN	AD0074	Variable of form TRTxxPN does not have a 2-digit [01-99] number for xx	Format	Error
ADSL	2		ARM	null	AD1008	Null value in variable marked as Required	Presence	Error
ADSL	8		ARM	null	AD1008	Null value in variable marked as Required	Presence	Error
ADSL	9		ARM	null	AD1008	Null value in variable marked as Required	Presence	Error

INTERPRETING THE OPENCDISC VALIDATOR OUTPUT

As we have seen, for both SDTM and ADaM files, the OpenCDISC Validator generates the same type of report. In the examples above, we requested that our report be put in Excel format. Within the Excel spreadsheet, the validator generated 4 information tabs, providing different levels of information on each tab. However, prior to viewing the report, the validator provides some feedback as to what it has encountered. For example, with a different SDTM AE file it reported that 3,606 records were read, 10,899 messages were generated, and 122,674 checks were performed. The number of messages generated is a bit overwhelming. So, let's take a closer look and see if we can figure out why there are so many issues with this file.

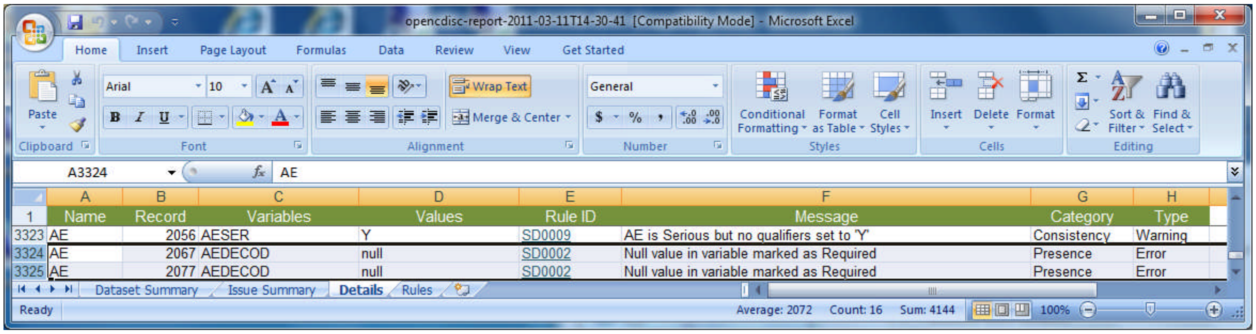
The Dataset Summary tab gives an overview of the issues encountered in the validation. It reported that encountered 79 errors and 10816 warnings. 79 errors are less daunting than the 10,899 messages. But we need to take a closer look. The Issue Summary tab provides a bit more detail. Breaking down the Error messages and Warning messages by type.


opencdisc-report-2011-03-11T14-30-41 [Compatibility Mode] - Microsoft Excel
A1
OpenCDISC Validator Report
Dataset Summary
Issue Summary
Details
Rules
Ready
100%

Issue Summary

Rule ID	Message	Found
Error		
SD0002	Null value in variable marked as Required	77
SD0013	Begin day must be less than or equal to end day	2

Looking at the first error message (Rule ID: SD0002), we see that the validator found 77 occurrences for a variable that is NULL but is a REQUIRED variable. Now we are starting to zoom in on the problem. The Details tab does just that. Searching for Rule ID SD0002 we quickly determine that NULL values are for the variable AEDECOD.


opencdisc-report-2011-03-11T14-30-41 [Compatibility Mode] - Microsoft Excel
Home Insert Page Layout Formulas Data Review View Get Started
Paste Clipboard Font Alignment Number Styles Cells Editing
Wrap Text Merge & Center General Conditional Formatting Format as Table Cell Styles Insert Delete Format Sort & Filter Find & Select
Dataset Summary Issue Summary Details Rules
Ready Average: 2072 Count: 16 Sum: 4144 100%

A3324 AE

	Name	Record	Variables	Values	Rule ID	Message	Category	Type
3323	AE	2056	AESER	Y	SD0009	AE is Serious but no qualifiers set to 'Y'	Consistency	Warning
3324	AE	2067	AEDECOD	null	SD0002	Null value in variable marked as Required	Presence	Error
3325	AE	2077	AEDECOD	null	SD0002	Null value in variable marked as Required	Presence	Error

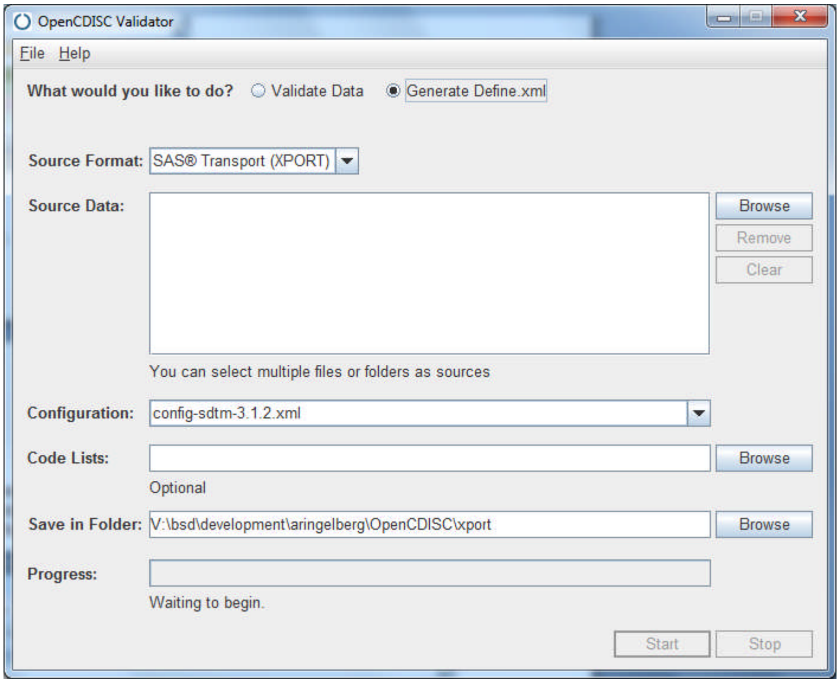
It looks like we have some adverse events that have not yet been coded. Knowing what the problem is, we can go back to the production dataset and make the correction.

The Rules tab is a for-your-information tab. It reports all of the validator rules by ID and provides the message, description, category and type for each rule.

CREATING DEFINE.XML

The OpenCDISC Validator can also generate DEFINE.XML. From the OpenCDISC Validator window perform the following steps.

Step 1: From the 'What would you like to do?' prompt, choose 'Generate Define.xml'.



What would you like to do? ○ Validate Data ● Generate Define.xml

Source Format: SAS® Transport (XPORT)

Source Data:

You can select multiple files or folders as sources

Configuration: config-sdtm-3.1.2.xml

Code Lists:

Optional

Save in Folder: V:\bsd\development\aringelberg\OpenCDISC\xport

Progress:

Waiting to begin.

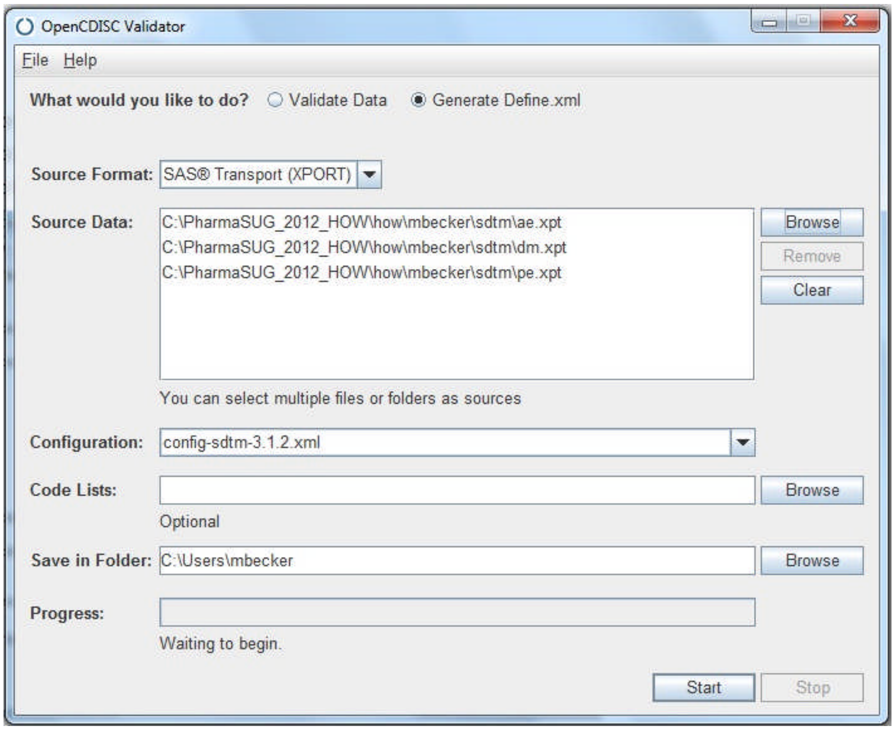
Step 2: Choose the Source Format. For this example we are using the default, XPORT.

Step 3: Choose the file(s) you want to create define.xml for using the Browse button.

Step 4: Choose the Configuration. We are using 'config-sdtm-3.1.2.xml' for this example.

Step 5: Select the folder where you want to save the define.xml file.

Step 6: Click on Start.



What would you like to do? ○ Validate Data ● Generate Define.xml

Source Format: SAS® Transport (XPORT)

Source Data:
C:\PharmaSUG_2012_HOW\how\mbecker\sdtm\ae.xpt
C:\PharmaSUG_2012_HOW\how\mbecker\sdtm\dm.xpt
C:\PharmaSUG_2012_HOW\how\mbecker\sdtm\pe.xpt

You can select multiple files or folders as sources

Configuration: config-sdtm-3.1.2.xml

Code Lists:

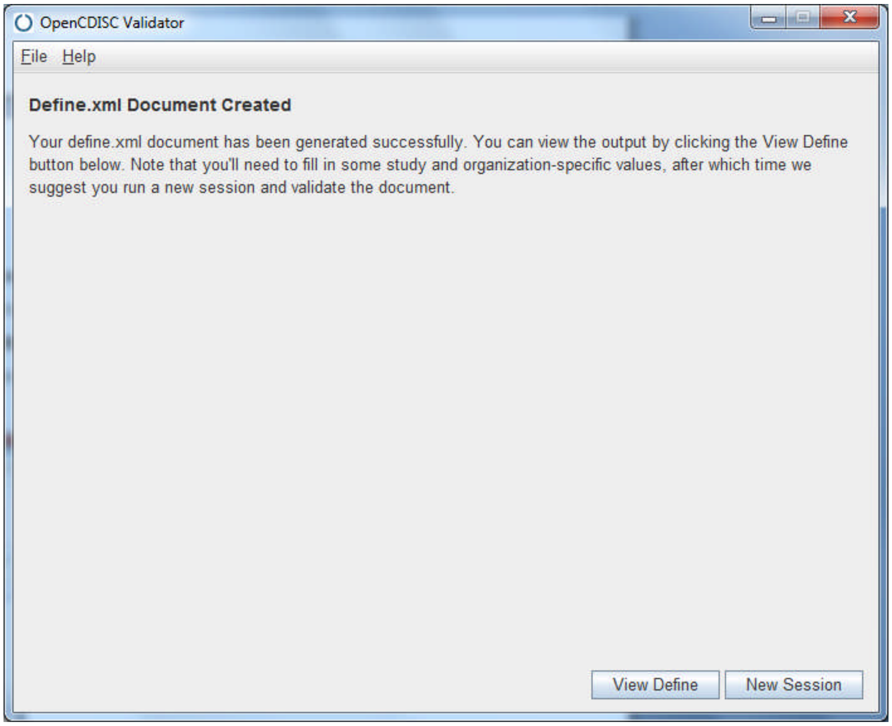
Optional

Save in Folder: C:\Users\mbecker

Progress:

Waiting to begin.

When the define.xml file has been generated, the following information screen will appear:

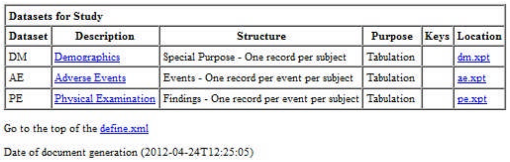


Define.xml Document Created

Your define.xml document has been generated successfully. You can view the output by clicking the View Define button below. Note that you'll need to fill in some study and organization-specific values, after which time we suggest you run a new session and validate the document.

Step 7: Choose 'View Define' to see the newly created define.xml file.

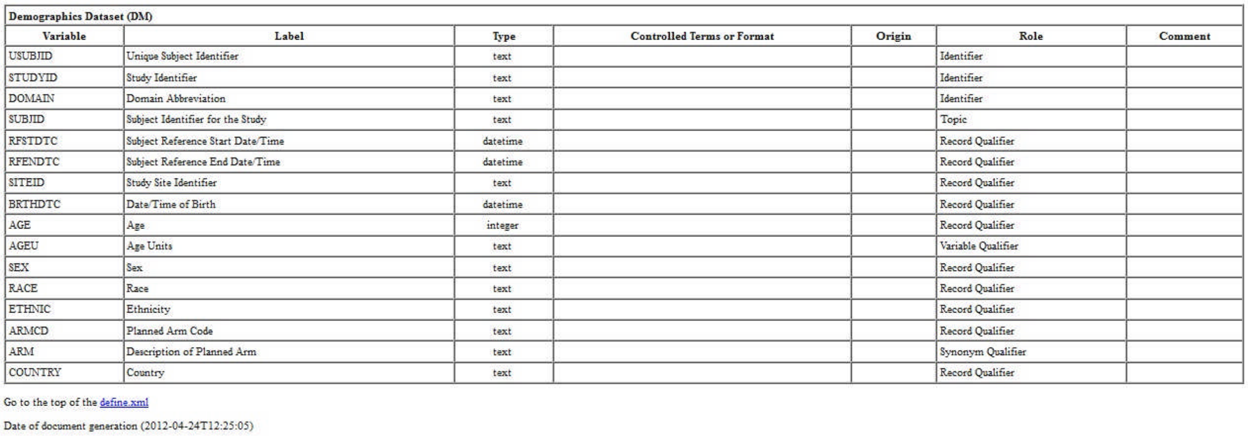
In this example, 3 SDTM files were selected: DM, AE and PE. Below is some of the XML output. Note that some study and organization-specific values still need to be filled in.



Datasets for Study					
Dataset	**Description**	**Structure**	**Purpose**	**Keys**	**Location**
DM	Demographics	Special Purpose - One record per subject	Tabulation		dm.xpt
AE	Adverse Events	Events - One record per event per subject	Tabulation		ae.xpt
PE	Physical Examination	Findings - One record per event per subject	Tabulation		pe.xpt

Go to the top of the define.xml

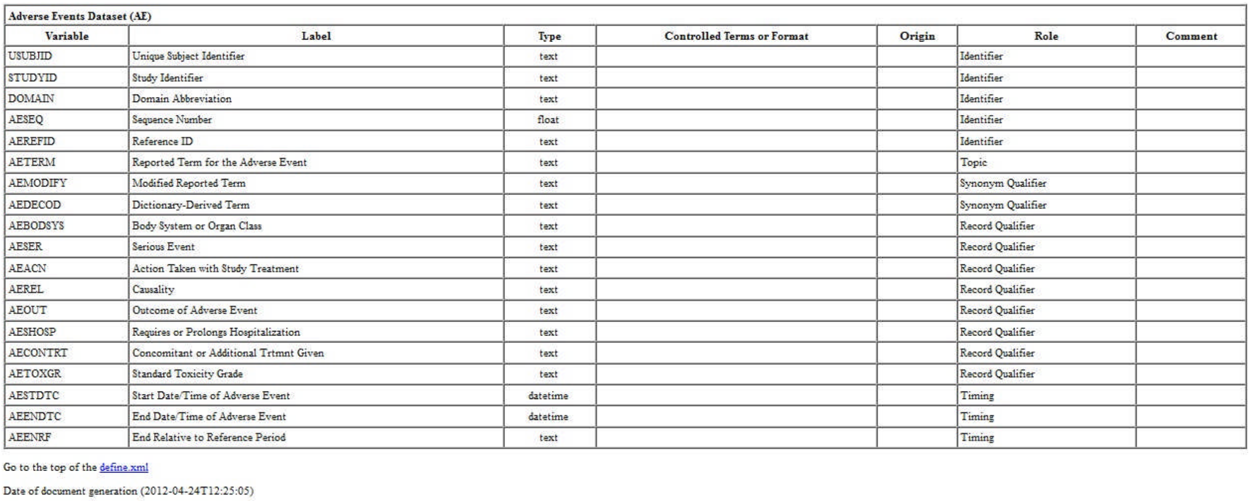
Date of document generation (2012-04-24T12:25:05)



Demographics Dataset (DM)						
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Origin**	**Role**	**Comment**
USUBJID	Unique Subject Identifier	text			Identifier	
STUDYID	Study Identifier	text			Identifier	
DOMAIN	Domain Abbreviation	text			Identifier	
SUBJID	Subject Identifier for the Study	text			Topic	
RFSTDTC	Subject Reference Start Date/Time	datetime			Record Qualifier	
RFENDTC	Subject Reference End Date/Time	datetime			Record Qualifier	
SITEID	Study Site Identifier	text			Record Qualifier	
BRTHDTC	Date/Time of Birth	datetime			Record Qualifier	
AGE	Age	integer			Record Qualifier	
AGEU	Age Units	text			Variable Qualifier	
SEX	Sex	text			Record Qualifier	
RACE	Race	text			Record Qualifier	
ETHNIC	Ethnicity	text			Record Qualifier	
ARMCD	Planned Arm Code	text			Record Qualifier	
ARM	Description of Planned Arm	text			Synonym Qualifier	
COUNTRY	Country	text			Record Qualifier	

Go to the top of the define.xml

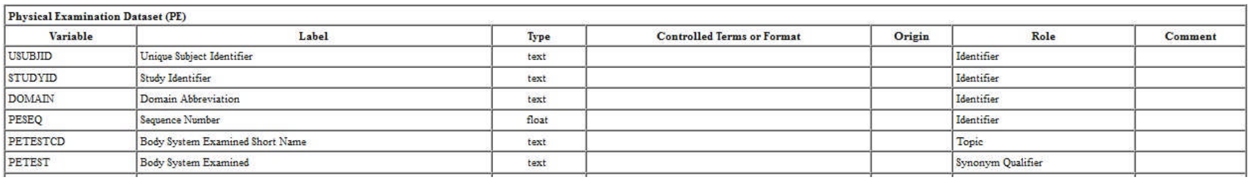
Date of document generation (2012-04-24T12:25:05)



Adverse Events Dataset (AE)						
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Origin**	**Role**	**Comment**
USUBJID	Unique Subject Identifier	text			Identifier	
STUDYID	Study Identifier	text			Identifier	
DOMAIN	Domain Abbreviation	text			Identifier	
AESEQ	Sequence Number	float			Identifier	
AEREFID	Reference ID	text			Identifier	
AETERM	Reported Term for the Adverse Event	text			Topic	
AEMODIFY	Modified Reported Term	text			Synonym Qualifier	
AEDECOD	Dictionary-Derived Term	text			Synonym Qualifier	
AEBODSYS	Body System or Organ Class	text			Record Qualifier	
AESER	Serious Event	text			Record Qualifier	
AEACN	Action Taken with Study Treatment	text			Record Qualifier	
AEREL	Causality	text			Record Qualifier	
AEOUT	Outcome of Adverse Event	text			Record Qualifier	
AESHOSP	Requires or Prolongs Hospitalization	text			Record Qualifier	
AECONTRT	Concomitant or Additional Trtmnt Given	text			Record Qualifier	
AETOXGR	Standard Toxicity Grade	text			Record Qualifier	
AESTDTC	Start Date/Time of Adverse Event	datetime			Timing	
AEENDTC	End Date/Time of Adverse Event	datetime			Timing	
AEENRF	End Relative to Reference Period	text			Timing	

Go to the top of the define.xml

Date of document generation (2012-04-24T12:25:05)



Physical Examination Dataset (PE)						
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Origin**	**Role**	**Comment**
USUBJID	Unique Subject Identifier	text			Identifier	
STUDYID	Study Identifier	text			Identifier	
DOMAIN	Domain Abbreviation	text			Identifier	
PESEQ	Sequence Number	float			Identifier	
PETESTCD	Body System Examined Short Name	text			Topic	
PETEST	Body System Examined	text			Synonym Qualifier	

CONCLUSION

The development of the open-source tool, the OpenCDISC Validator, is helping to ensure data compliance with CDISC models such as SDTM, ADaM and Define.xml. The tool is of commercial-quality, freely available and EASY to use. The tool eliminates the need for individualized QC that makes sure SDTM and ADaM files are CDISC compliant. The information generated by the validator is easy to interpret, allowing the user to make corrections in the affected SDTM or ADaM file - bringing that file up to CDISC standards. The validator also has the ability to create define.xml. It is as simple as bringing up the OpenCDISC Validator, choosing the option of ‘Create define.xml’ and instructing the validator which SDTM files need to be included. The OpenCDISC Validator saves a lot time and effort when it comes to the validation of SDTM and ADaM files and the creation of define.xml.

REFERENCES

HTTP://WWW.CDISC.ORG

HTTP://WWW.OPENCDISC.ORG

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Matt Becker
Address: 1000 CentreGreen Way, Suite 300
City, State ZIP: Cary, NC 27513
Work Phone: 919-388-4031
E-mail: mbecker@pharmanet-i3.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.