



A Symbol of Excellence

TS04

Running
OpenCDISC from
SAS

Mark Crangle

- The OpenCDISC validator is a tool used to check the compliance of datasets with CDISC standards

- Open-source



Freely available and created by team of experts

- Metadata driven



Validation rules can be easily modified

- Java based



Usable on a variety of operating systems

- Can validate a range of CDISC datasets



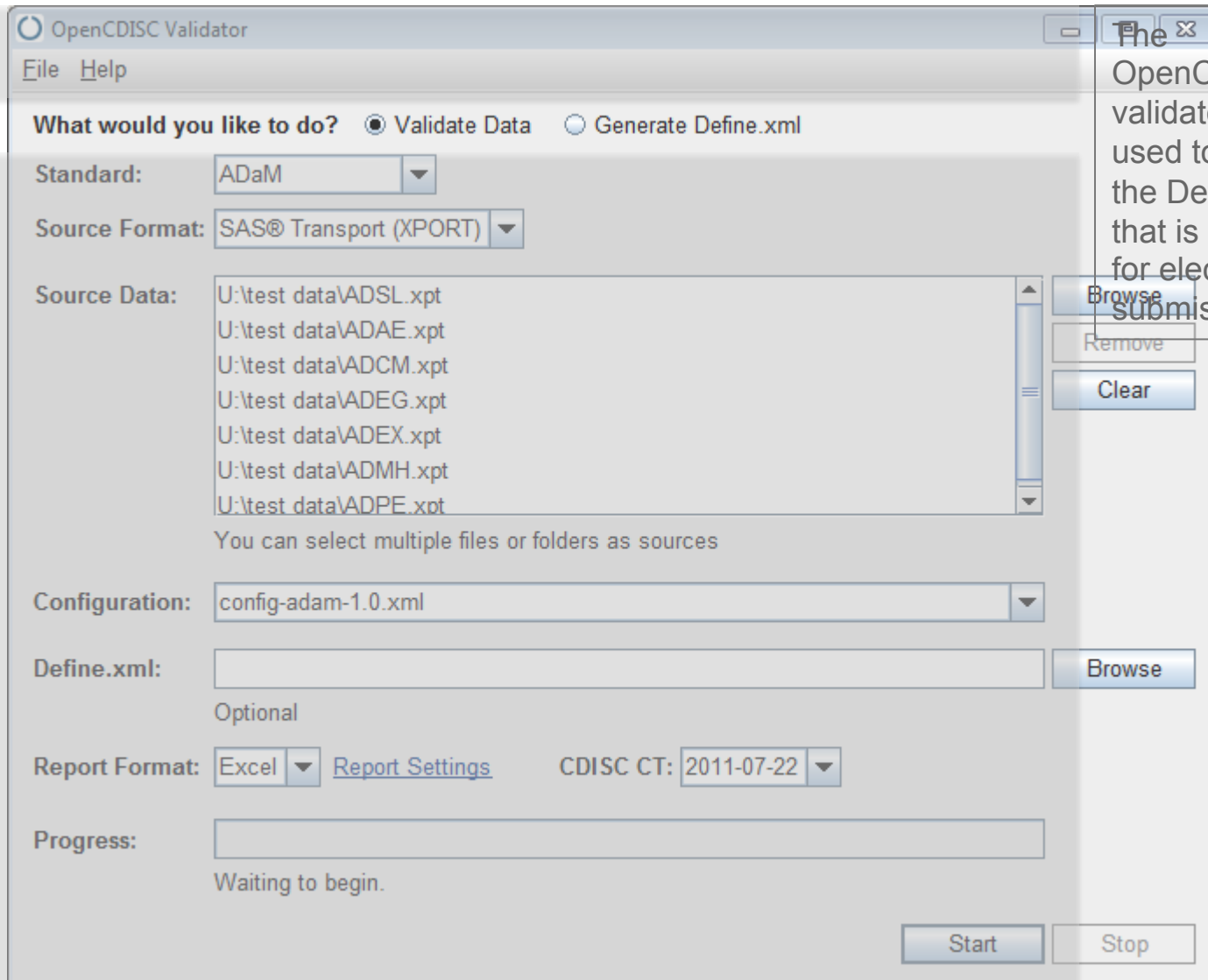
Solution can be applied to many types of dataset

- Tool is packaged with a graphical user interface but can also be run from a command line interface (CLI)

- Commonly a package of datasets is only checked for CDISC compliance once all are complete
- If any issues are identified then the dataset is updated and re-validated
 - Inefficient as can result in rework
 - Changes could affect other dataset/analysis programs that use the data
- More efficient solution would be to check CDISC compliance at the same time as other validation activities
- Issues can be identified and fixed early

- The OpenCDISC Validator can be configured to work more accurately for individual datasets
- The CLI can be used directly from a SAS session
 - Tool can be run alongside other validation activities
 - Results can be read back in to a SAS dataset, analysed and reported with other validation output

The OpenCDISC Validator



OpenCDISC Validator

File Help

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

Standard: ADaM

Source Format: SAS® Transport (XPORT)

Source Data:

- U:\test data\ADSL.xpt
- U:\test data\ADAE.xpt
- U:\test data\ADCM.xpt
- U:\test data\ADEG.xpt
- U:\test data\ADEX.xpt
- U:\test data\ADMH.xpt
- U:\test data\ADPE.xpt

You can select multiple files or folders as sources

Configuration: config-adam-1.0.xml

Define.xml: Browse

Optional

Report Format: Excel [Report Settings](#) CDISC CT: 2011-07-22

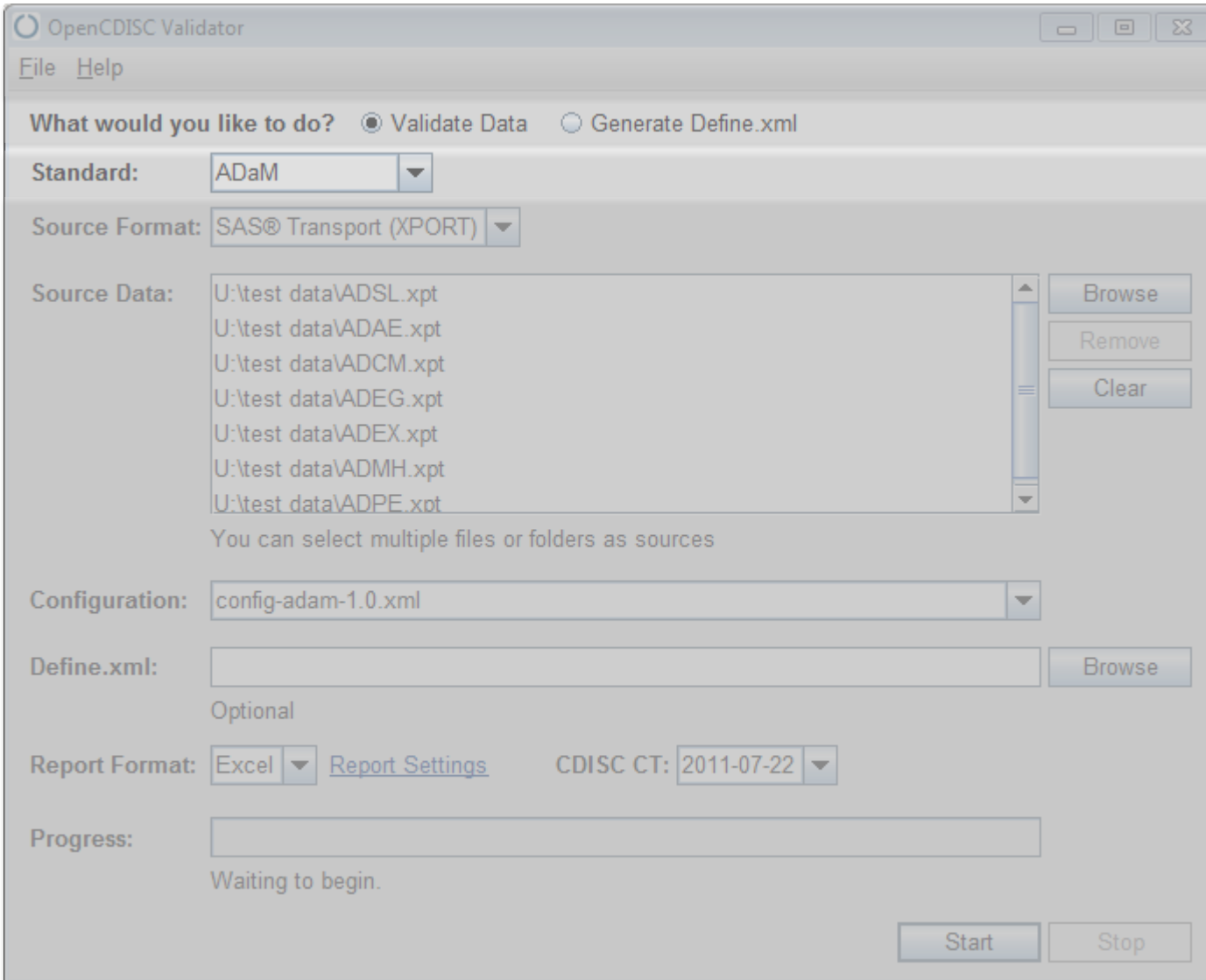
Progress:

Waiting to begin.

Start Stop

The OpenCDISC validator can be used to create the Define.xml that is required for electronic submission

The OpenCDISC Validator



The screenshot shows the OpenCDISC Validator application window. The title bar reads "OpenCDISC Validator". The menu bar contains "File" and "Help". Below the menu bar, there are two radio buttons: "Validate Data" (selected) and "Generate Define.xml".

The "Standard:" dropdown menu is set to "ADaM". The "Source Format:" dropdown menu is set to "SAS® Transport (XPORT)".

The "Source Data:" section contains a list of files in a scrollable area:

- U:\test data\ADSL.xpt
- U:\test data\ADAE.xpt
- U:\test data\ADCM.xpt
- U:\test data\ADEG.xpt
- U:\test data\ADEX.xpt
- U:\test data\ADMH.xpt
- U:\test data\ADPE.xpt

To the right of the list are three buttons: "Browse", "Remove", and "Clear". Below the list, it says "You can select multiple files or folders as sources".

The "Configuration:" dropdown menu is set to "config-adam-1.0.xml".

The "Define.xml:" section has an empty text box and a "Browse" button. Below it, it says "Optional".

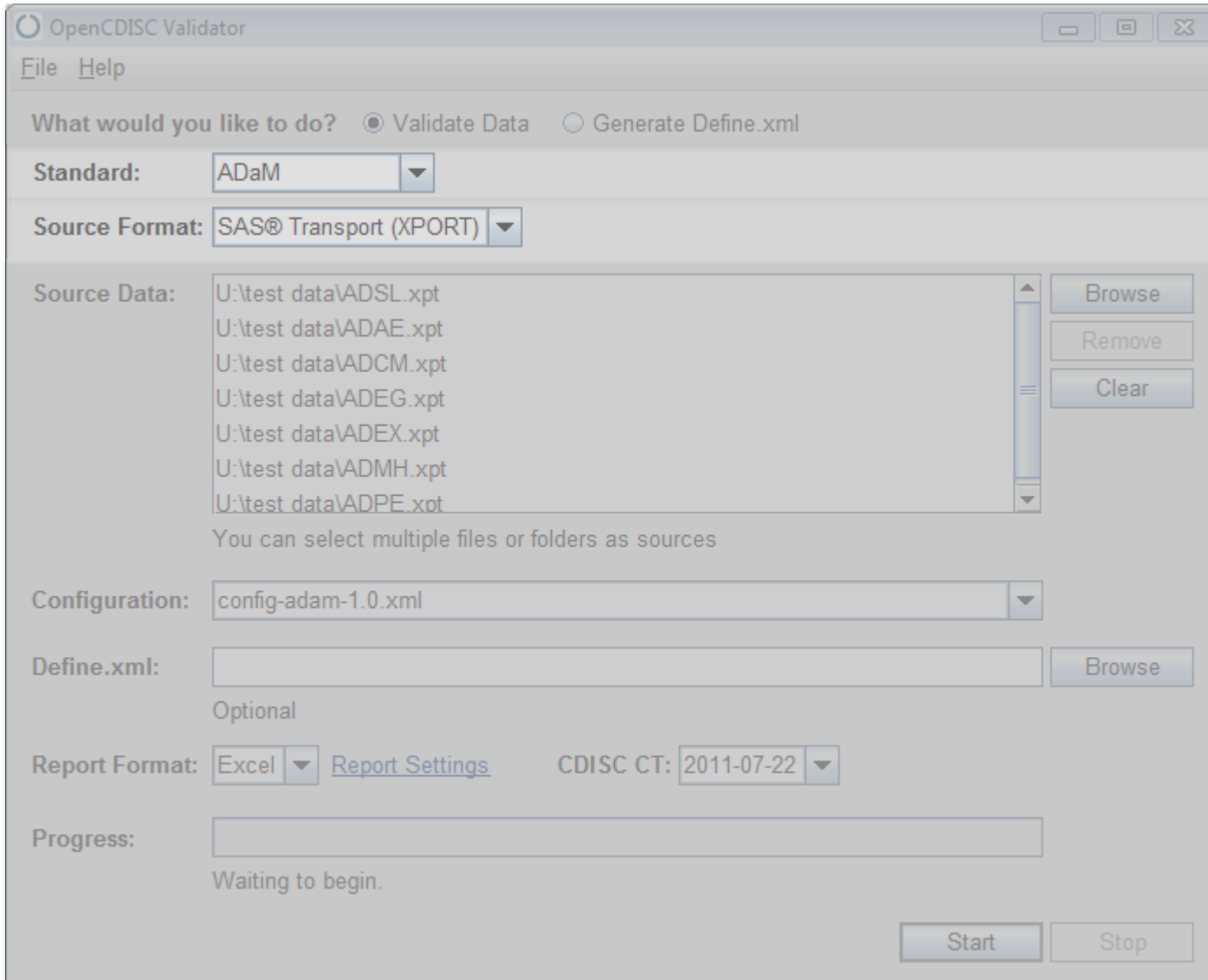
The "Report Format:" dropdown menu is set to "Excel", with a "Report Settings" link next to it. The "CDISC CT:" dropdown menu is set to "2011-07-22".

The "Progress:" section has an empty progress bar and the text "Waiting to begin.".

At the bottom right, there are two buttons: "Start" and "Stop".

The mentioned OpenCDISC tool validates and can be used to validate different types of CDISC standards for electronic submission.

The OpenCDISC Validator



The screenshot shows the OpenCDISC Validator application window. The title bar reads "OpenCDISC Validator". The menu bar contains "File" and "Help". Below the menu bar, there are two radio buttons: "Validate Data" (selected) and "Generate Define.xml".

The "Standard:" dropdown menu is set to "ADaM". The "Source Format:" dropdown menu is set to "SAS® Transport (XPORT)".

The "Source Data:" section contains a list of files: U:\test data\ADSL.xpt, U:\test data\ADAE.xpt, U:\test data\ADCM.xpt, U:\test data\ADEG.xpt, U:\test data\ADEX.xpt, U:\test data\ADMH.xpt, and U:\test data\ADPE.xpt. To the right of the list are three buttons: "Browse", "Remove", and "Clear". Below the list, it says "You can select multiple files or folders as sources".

The "Configuration:" dropdown menu is set to "config-adam-1.0.xml".

The "Define.xml:" section has an empty text box and a "Browse" button. Below it, it says "Optional".

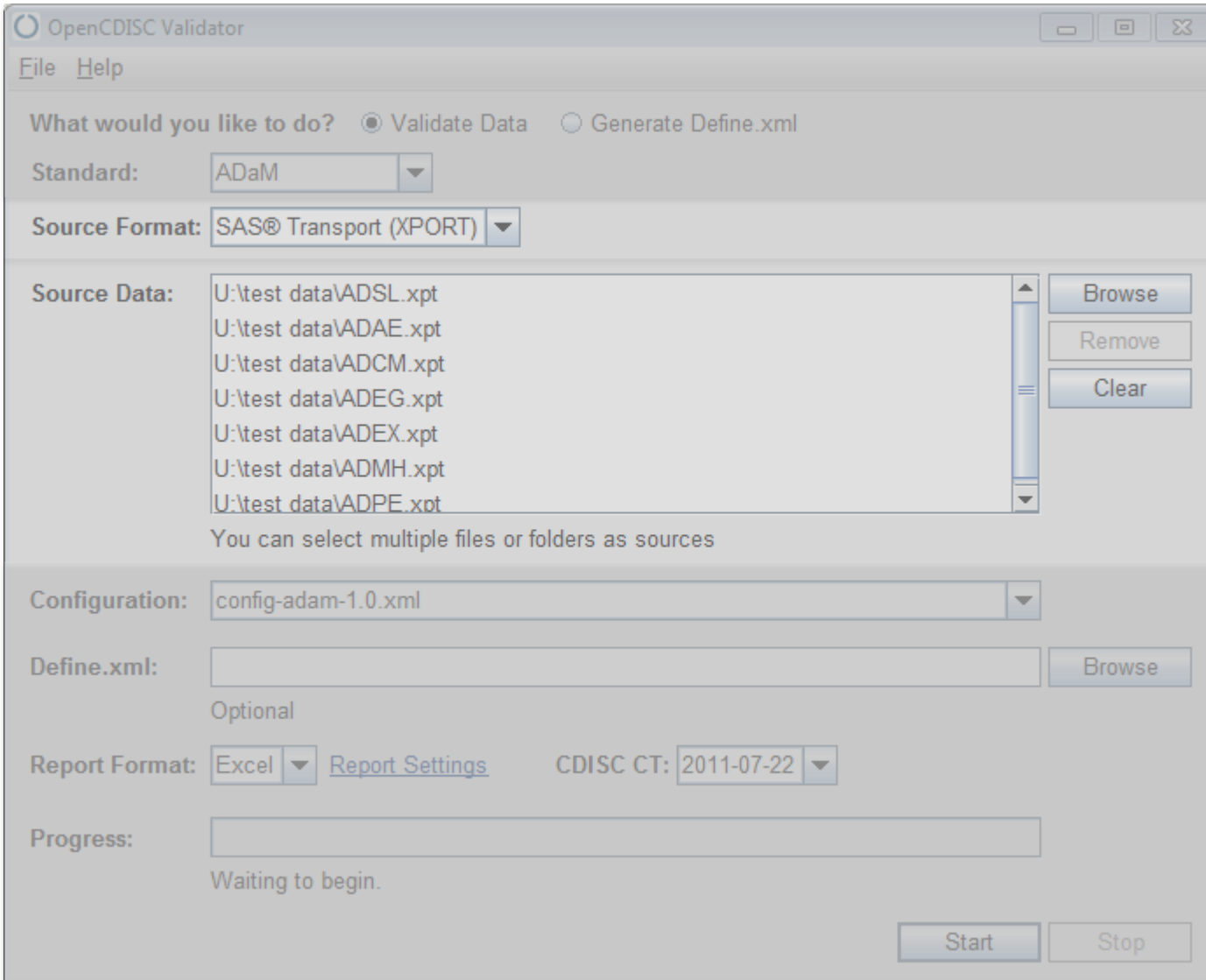
The "Report Format:" dropdown menu is set to "Excel", with a "Report Settings" link next to it. The "CDISC CT:" dropdown menu is set to "2011-07-22".

The "Progress:" section has an empty progress bar and the text "Waiting to begin.".

At the bottom right, there are two buttons: "Start" and "Stop".

As possible to introduce the assets in a complex manner to validated different types of CDISC standard items.

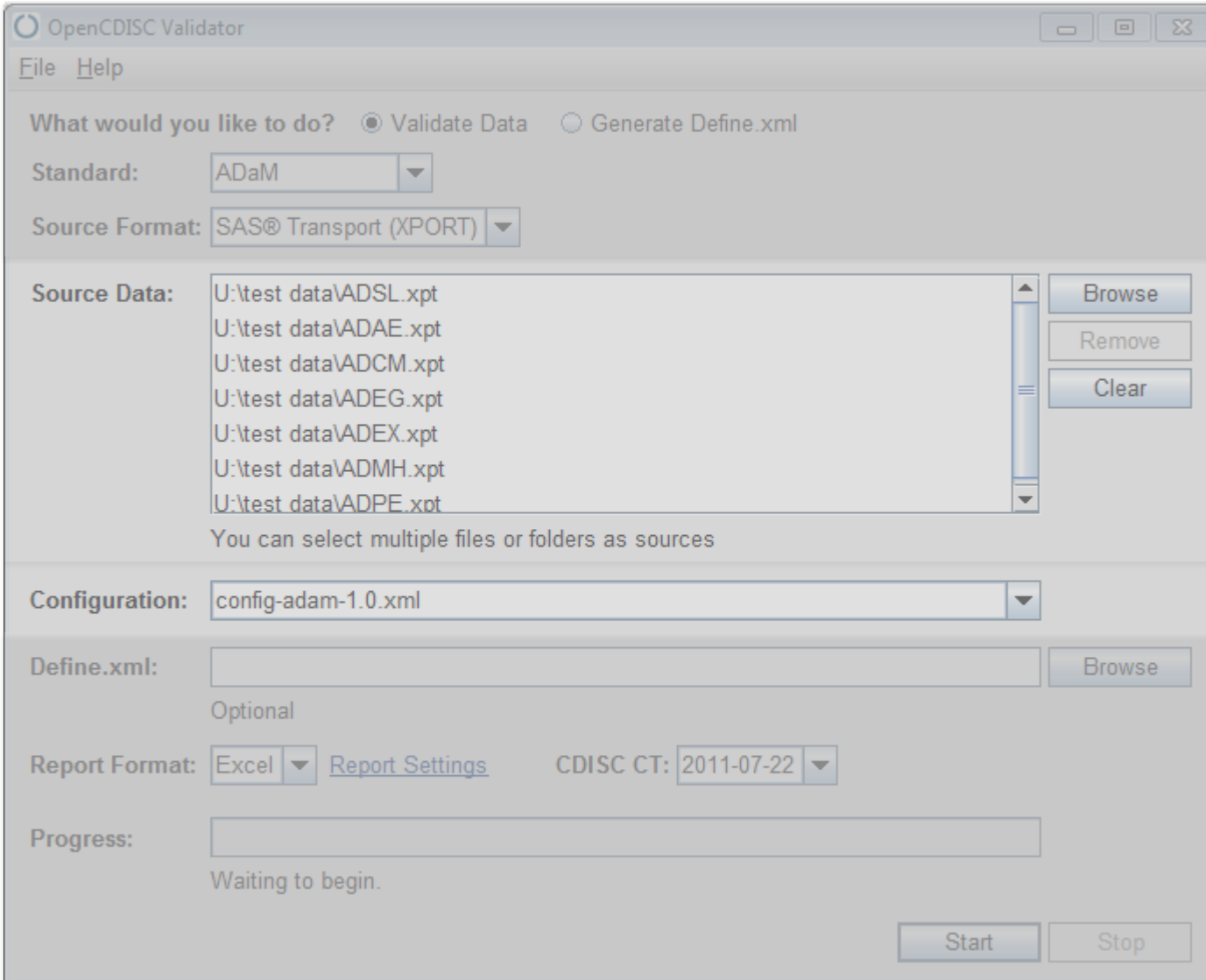
The OpenCDISC Validator



The screenshot shows the OpenCDISC Validator application window. At the top, there is a menu bar with 'File' and 'Help'. Below the menu bar, a section titled 'What would you like to do?' contains two radio buttons: 'Validate Data' (selected) and 'Generate Define.xml'. The 'Standard:' dropdown is set to 'ADaM'. The 'Source Format:' dropdown is set to 'SAS® Transport (XPORT)'. The 'Source Data:' list contains several files: 'U:\test data\ADSL.xpt', 'U:\test data\ADAE.xpt', 'U:\test data\ADCM.xpt', 'U:\test data\ADEG.xpt', 'U:\test data\ADEX.xpt', 'U:\test data\ADMH.xpt', and 'U:\test data\ADPE.xpt'. To the right of this list are 'Browse', 'Remove', and 'Clear' buttons. Below the list, a note states 'You can select multiple files or folders as sources'. The 'Configuration:' dropdown is set to 'config-adam-1.0.xml'. The 'Define.xml:' field is empty, with a 'Browse' button to its right. Below this, the text 'Optional' is displayed. The 'Report Format:' dropdown is set to 'Excel', with a 'Report Settings' link next to it. The 'CDISC CT:' dropdown is set to '2011-07-22'. The 'Progress:' field is empty, with the text 'Waiting to begin.' below it. At the bottom right, there are 'Start' and 'Stop' buttons.

Datasets to be checked are specified here in a delimited format

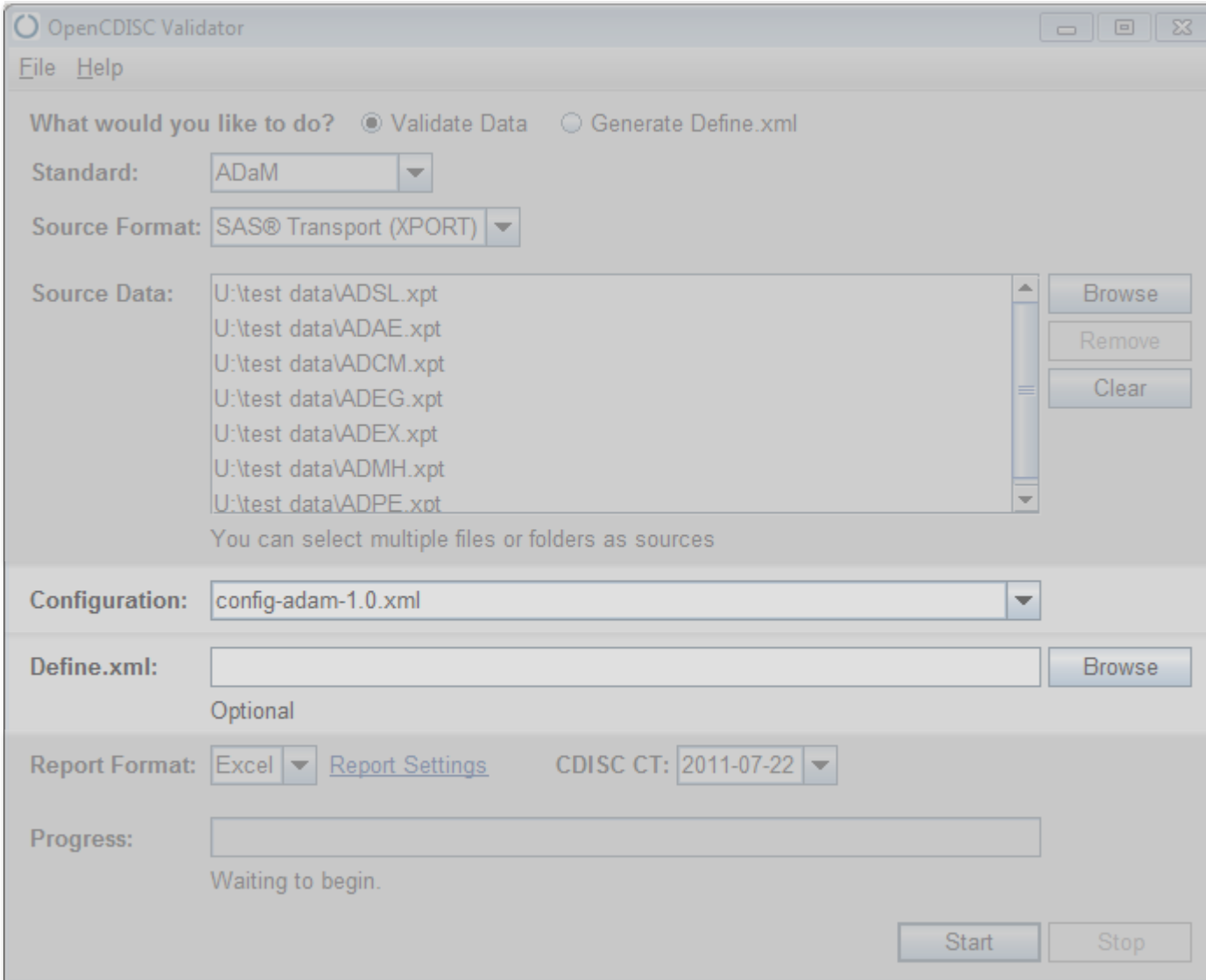
The OpenCDISC Validator



The screenshot shows the OpenCDISC Validator application window. The title bar reads "OpenCDISC Validator". The menu bar contains "File" and "Help". The main area has a section titled "What would you like to do?" with two radio buttons: "Validate Data" (selected) and "Generate Define.xml". Below this are two dropdown menus: "Standard:" set to "ADaM" and "Source Format:" set to "SAS® Transport (XPORT)". The "Source Data:" section features a list box containing several files: "U:\test data\ADSL.xpt", "U:\test data\ADAE.xpt", "U:\test data\ADCM.xpt", "U:\test data\ADEG.xpt", "U:\test data\ADEX.xpt", "U:\test data\ADMH.xpt", and "U:\test data\ADPE.xpt". To the right of the list box are three buttons: "Browse", "Remove", and "Clear". Below the list box is the text "You can select multiple files or folders as sources". The "Configuration:" section has a dropdown menu set to "config-adam-1.0.xml". The "Define.xml:" section has an empty text box and a "Browse" button, with the word "Optional" below it. The "Report Format:" section has a dropdown menu set to "Excel" and a "Report Settings" link. The "CDISC CT:" section has a dropdown menu set to "2011-07-22". The "Progress:" section has an empty progress bar and the text "Waiting to begin.". At the bottom right are "Start" and "Stop" buttons.

The datasets to be checked are specified in the configuration file. The validator is packaged with configuration files but these can be edited or new files created.

The OpenCDISC Validator

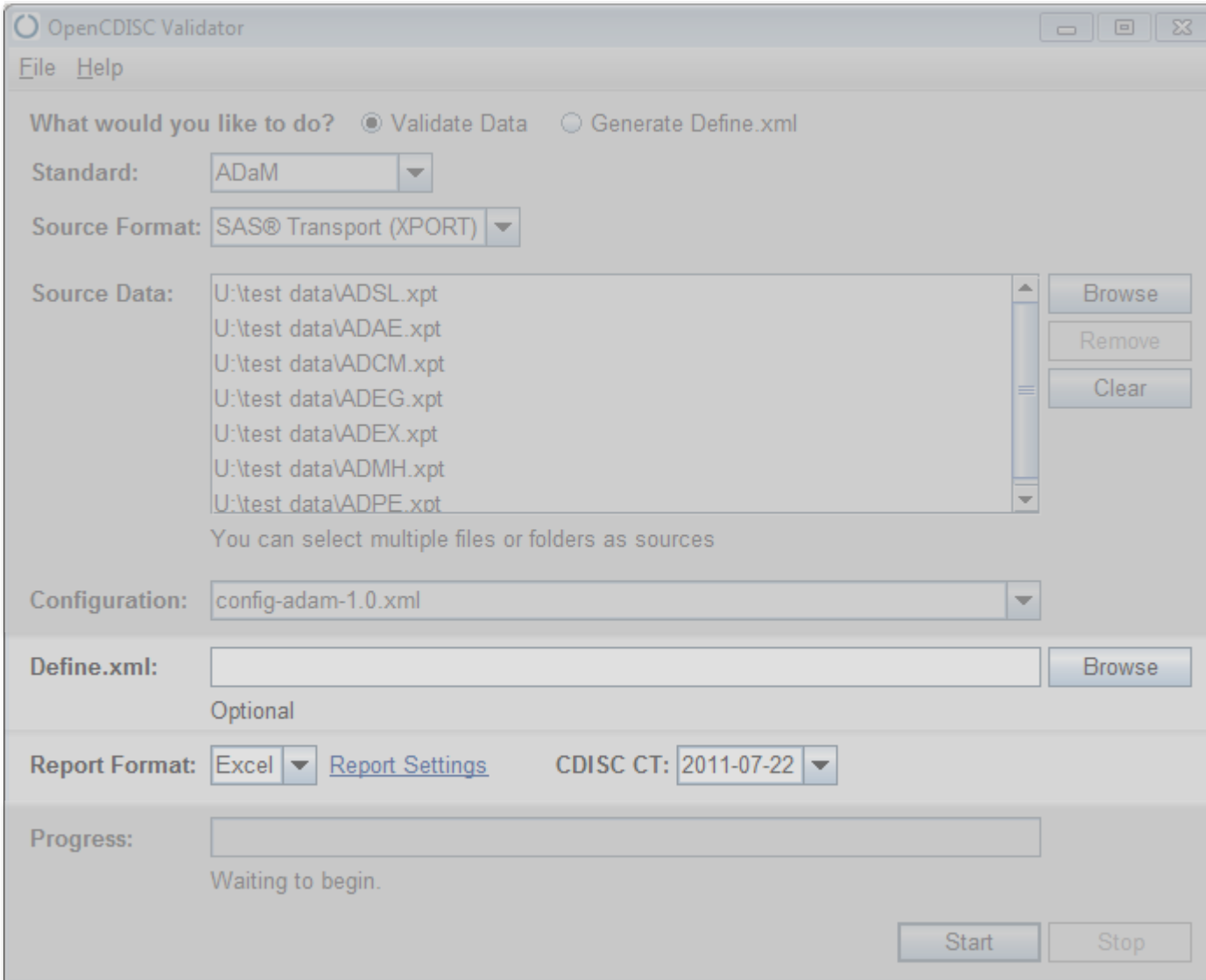


The screenshot shows the OpenCDISC Validator application window. At the top, there are window control buttons (minimize, maximize, close). Below the title bar is a menu bar with 'File' and 'Help'. The main area contains several sections:

- What would you like to do?** with radio buttons for 'Validate Data' (selected) and 'Generate Define.xml'.
- Standard:** a dropdown menu set to 'ADaM'.
- Source Format:** a dropdown menu set to 'SAS® Transport (XPORT)'.
- Source Data:** a list box containing several files: 'U:\test data\ADSL.xpt', 'U:\test data\ADAE.xpt', 'U:\test data\ADCM.xpt', 'U:\test data\ADEG.xpt', 'U:\test data\ADEX.xpt', 'U:\test data\ADMH.xpt', and 'U:\test data\ADPE.xpt'. To the right of the list are 'Browse', 'Remove', and 'Clear' buttons.
- A note below the list: 'You can select multiple files or folders as sources'.
- Configuration:** a dropdown menu set to 'config-adam-1.0.xml'.
- Define.xml:** an empty text box with a 'Browse' button to its right. Below it is the text 'Optional'.
- Report Format:** a dropdown menu set to 'Excel' and a 'Report Settings' link.
- CDISC CT:** a dropdown menu set to '2011-07-22'.
- Progress:** an empty progress bar with the text 'Waiting to begin.' below it.
- At the bottom right, there are 'Start' and 'Stop' buttons.

The Define.xml file for the dataset specified can be supplied that will be performed on the data checks metadata that these checks will be using. The validator is packaged with configuration files but these can be edited or new files created

The OpenCDISC Validator











The screenshot shows the OpenCDISC Validator application window. The title bar reads "OpenCDISC Validator". The menu bar contains "File" and "Help". The main area has a radio button group for "What would you like to do?" with "Validate Data" selected and "Generate Define.xml" unselected. Below this are dropdown menus for "Standard:" (set to "ADaM") and "Source Format:" (set to "SAS® Transport (XPORT)"). A list box for "Source Data:" contains several files: "U:\test data\ADSL.xpt", "U:\test data\ADAE.xpt", "U:\test data\ADCM.xpt", "U:\test data\ADEG.xpt", "U:\test data\ADEX.xpt", "U:\test data\ADMH.xpt", and "U:\test data\ADPE.xpt". To the right of the list are "Browse", "Remove", and "Clear" buttons. Below the list is the text "You can select multiple files or folders as sources". The "Configuration:" dropdown is set to "config-adam-1.0.xml". The "Define.xml:" field is empty with a "Browse" button and the text "Optional" below it. The "Report Format:" dropdown is set to "Excel" with a "Report Settings" link next to it. The "CDISC CT:" dropdown is set to "2011-07-22". At the bottom, there is a "Progress:" field with the text "Waiting to begin." and "Start" and "Stop" buttons.

The validation file for the dataset specified can be supplied to CSV. For the format is the easiest to read but the CSV file can be more easily read into a SAS dataset if required

The tool is packaged with different versions of the ADaM, SDTM and SEND Controlled Terminology which can be selected here.

- The rules the validator uses are specified in xml configuration files
- The tool is packaged with preset configuration files for each release of CDISC standards

 config-adam-1.0	XML Document	161 KB
 config-define-1.0	XML Document	19 KB
 config-define-2.0	XML Document	22 KB
 config-sdtm-3.1.1	XML Document	741 KB
 config-sdtm-3.1.2	XML Document	946 KB
 config-sdtm-3.1.3	XML Document	1,126 KB
 config-sdtm-3.2	XML Document	1,433 KB
 config-send-3.0	XML Document	787 KB

- Configuration file has section for global rules and then one for each dataset type
 - In ADaM configuration, one for ADSL and one for BDS
 - Each section is further divided into the variable metadata and rules specific to that datatype.

Validator Configuration

- The configuration file can be viewed in a web browser to show the validation rules in a user friendly format

Validation Rules					
ID	Description	Validator	Message	Severity	Active
AD0005	A variable with a suffix of FL must have value that is Y, N or null (exception 1: RFL, PFL , ABLFL, ANLzzFL. Exception 2: Population flags COMPLFL,FASFL,ITTFL,PPROTFL,SAFFL,RANDFL,ENRFL cannot be null and at least 1 must be included in ADSL)	Match	%Variable.1% FL value is not Y, N or null	Error	✓
AD0006	A variable with a suffix of FN has a value that is not Y, N or null (exception: RFL, PFL, ABLFN, ANLzzFN and population flags COMPLFN,FASFN,ITTFN,PPROTFN,SAFFN,RANDFN,ENRFLFN cannot be null and at least 1 must be included in ADSL)	Match	%Variable.1% FN value is not 0, 1 or null	Error	✓
AD0007	If a Flag Numeric (*FN) variable is present, its corresponding Flag Character (*FL) variable must be present	Find	*FL variable is not present but its corresponding *FN variable is present	Error	✓

Yes/No flag that defines if each Rule is applied by the Validator. Responds with a check flag for each dataset type where the condition is applicable by CD. Easily "switched off"

- Some of the checks are used to check consistency between datasets but these are not needed for checking individual datasets
 - Checking an ADaM package contains the required ADSL dataset. (AD0001)
 - Checking subjects present in a dataset are also included in ADSL/DM. (AD0053)
- To de-activate a specific check the xml file needs to be edited in a text editor
- Rules are applied to each dataset type by the <val:ValidationRuleRef> tag

```
<val:ValidationRuleRef RuleID="AD0001" Active="Yes"/>
```
- To de-activate a check set Active to No
- Once the file is saved the change can be easily seen in the web-browser view

Validation Rules						
ID	Description	Validator	Message	Severity	Active	
AD0001	ADaM Subject level (ADSL) dataset should be included in every submission.	Find	Missing ADSL dataset	 Error		

- The GUI is the most commonly used way to run the validator but it is also accessible using a command line interface.
- This is accessed using the file `validator-cli-version.jar` (eg. `validator-cli-1.5.jar`)
- Same options as the GUI are available as parameters:

Parameter	Valid Values	Description
-task	Validate, Generate	Validate data or generate a Define.xml
-type	SDTM, ADaM, SEND, Define, Custom	Data Standard/Structure to validate
-source		Path to the source data files
-config		Path to the xml configuration document specifying the rules/metadata to validate
-config:define		Path of the define.xml for the study
-config:cdisc		CDISC Controlled Terminology Version
-report		Path and filename where the validation report will be saved
-report:type	Excel, CSV	Report format

- The command line interface can be used through the command prompt

- Navigate to the folder where the CLI file is located
- Run the command

```
java -jar validator-cli-1.5.jar
```

adding the parameters needed

- For example

```
java -jar validator-cli-1.5.jar
```

```
-task=Validate
```

```
-type=ADAM
```

```
-source="U:/test data/*.xpt"
```

```
-config="U:/test data/opencdisc-validator/config/config-adam-1.0.xml"
```

```
-config:cdisc=2011-07-22
```

```
-report="U:/test data/opencdisc-validator/reports/report.xls"
```

```
-report:type=Excel
```

- Combine CLI with ability to run system commands from SAS to control the OpenCDISC validator from within SAS
- The edited configuration file is used to eliminate error messages that are caused by only submitting one dataset
- This is handled by a macro that can be called in a dataset validation program

Example

```
/*Code to create validation dataset*/

ODS RTF FILE = "outputfile.rtf"

PROC COMPARE BASE=dev COMPARE=val;
  ...
RUN;

%runOpenCDISC(version, dataset, etc.);

ODS RTF CLOSE;
```


- Validator outputs report to CSV file as easier to read into SAS

A	B	C	D	E	F	G	H
Name	Record	Variable	Value	Rule ID	Message	Category	Type
ADLB	25897	PARCAT3, PARAMCD	Secondary Laboratory, HEOSLES	AD0124	Inconsistent value for PARCAT3 within a unique PARAMCD	Consistency	Error
ADLB	25898	PARCAT3, PARAMCD	Secondary Laboratory, HHCTS	AD0124	Inconsistent value for PARCAT3 within a unique PARAMCD	Consistency	Error
ADLB	25899	PARCAT3, PARAMCD	Secondary Laboratory, HHGBS	AD0124	Inconsistent value for PARCAT3 within a unique PARAMCD	Consistency	Error
ADLB	25900	PARCAT3, PARAMCD	Secondary Laboratory, HLYMLES	AD0124	Inconsistent value for PARCAT3 within a unique PARAMCD	Consistency	Error

- Read CSV file into SAS dataset

```
PROC IMPORT OUT=test DATAFILE = "U:\OpenCDISC ADLB.csv" DBMS=CSV REPLACE;
  GETNAMES=YES;
  DATAROW=2;
  GUESSINGROWS=100000;
RUN;
```

- Summarise issues

```
PROC FREQ data=report NOPRINT;
  TABLE rule_id*message / OUT=numobs;
RUN;
```

Rule_ID	Message	COUNT
AD0124	Inconsistent value for PARCAT3 within a unique PARAMCD	790
AD0137	CRITyFL is populated and CRITy is not populated	656
AD0141	Inconsistent value for PARAM within a unique PARAMCD	56
AD0146	Inconsistent value for PARAM	56
AD0150	Inconsistent value for AVAL	159

- Report could be checked programmatically and further code run conditionally if common issues found
 - Common issues is that PARAM must have the same value within each unique value of PARAMCD (1-1 correspondence)
 - Difficult to see complete issue just by searching through report
 - Check if report contains this issue then run code to show all unique combinations of PARAM/PARAMCD so can easily see where the issue lies.
- Permanent copies of datasets containing OpenCDISC report could be kept to allow tracking of issues
 - Any warnings/notes that have been investigated and deemed acceptable could be flagged so that they are not reported each time
 - Tracking of common issues could be used to identify training needs across department

- CDISC compliance is something that must be considered at all stages of dataset design, development and validation
- Only checking compliance after all datasets are complete is inefficient and could require re-work
- Using the CLI, the OpenCDISC validator can be controlled from within a SAS session to check individual datasets
- This can be done inline with other validation activities
- OpenCDISC report can be processed by SAS to summarise issues
- Reporting can be done alongside output from other validation methods to create a complete record of dataset quality

ANY QUESTIONS?