

Know Your Sources – Decoding Inputs / Outputs of any SAS Program.

Sandeep Juneja, SAS Institute Inc, Cary, USA

ABSTRACT

One of the biggest challenges in SAS Programming is that there is no standard way to generate Traceability Matrix for SAS programs with their inputs and outputs. As time progresses programs are updated and there is no way to do impact analysis for these updates on existing programs. This paper talks about how to generate Traceability Matrix for any SAS program along with their inputs / outputs and their version numbers in SAS Drug Development Environment (SDD). It decodes the JobLog XML files (.jlgxml) and generates report which can help to carry out impact analysis for any updates to existing source code.

The Challenge

There is no standard way to tell what are the various input / output datasets and files in a given SAS program. The problem arises when one of the source program (for e.g. SAS macro) used as an input to other source programs is updated and it becomes challenge to find out what kind of impact these updates can have on other programs.

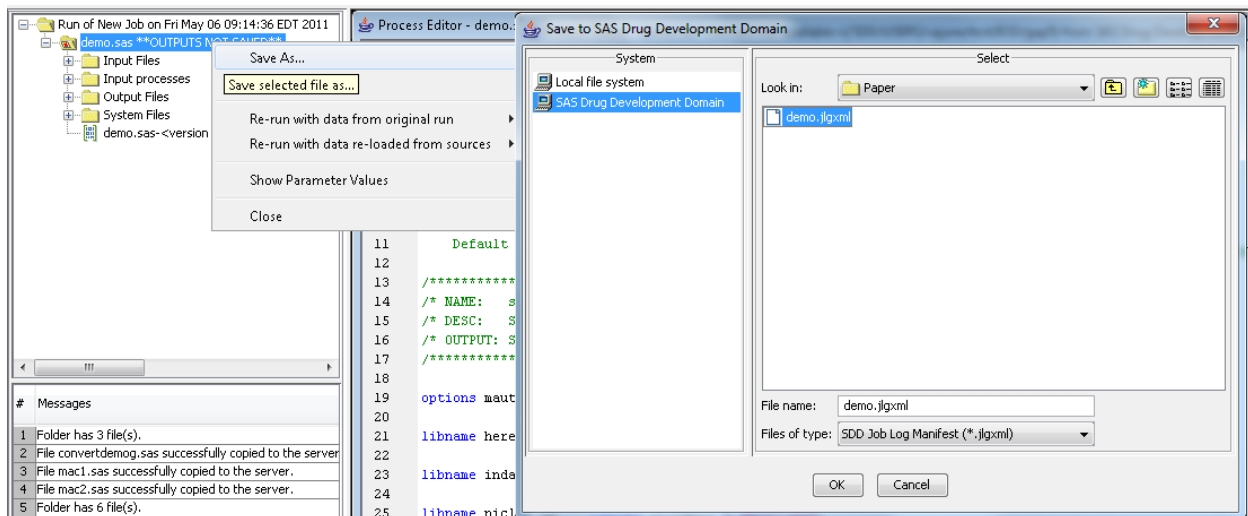
The Solution

If we can generate a Matrix with Source program name and various inputs / outputs and if we can also capture version numbers of these inputs/outputs it would be very easy to trace and analyze that if any update is done on one of the input program what kind of Impact this update will have on other source programs.

Luckily SAS does provide solution to above challenge in the form of SAS Drug Development (SDD). It's a control Environment System which lets you generate various versions of the files their by regulating various updates being done to these files. When SAS programs are run and saved in SDD they generate an XML file manifest which is called Job Log XML file. These files contain details about various inputs and outputs along with their version number and other details. The next challenge is how to decode these files to generate Traceability Matrix to carry out Impact Analysis.

JobLog XML (.jlgxml) files

When any SAS program is run in SDD it generates results with information about various Inputs/ Outputs. Right click

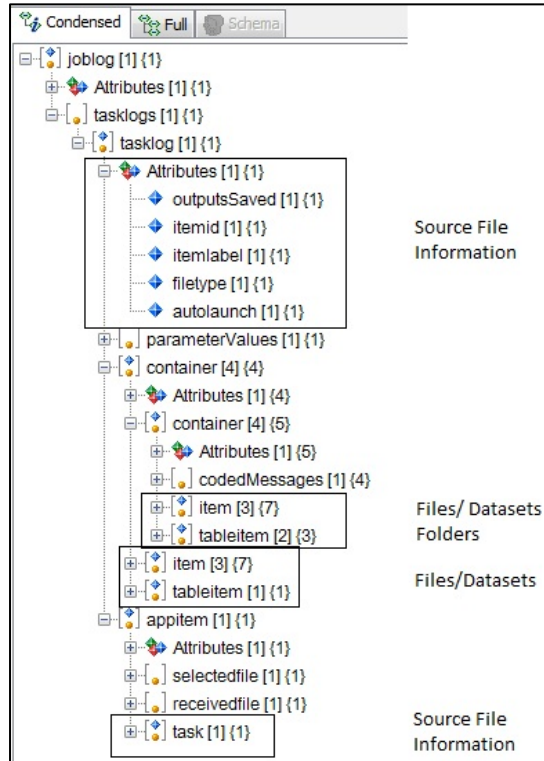


the job and click Save AS which should open another dialog box to save the results of the current program run in form of Joblog XML file.

Decode JobLog XML (.jlgxml) files

Step 1: Open the File in SAS XML Mapper

SAS XML Mapper is free utility which can be downloaded from SAS Website. It's a GUI tool which helps you convert XML file information into SAS dataset. It helps you generate XML Map (.map) file by using drag and drop elements feature and generates both the .map file and sas code which is used in SAS to read XML file and convert them into SAS dataset using XML libname engine.



Step 2: Understand the Basic Structure of joblog XML file

This diagram shows various Elements in XML file where important information is stored which when collected together help us generate Traceability matrix. Notice the information is stored in non-hierarchical way, so it's important to extract the information in multiple SAS datasets and then join them together using some key variables.

Source File Information

It could be possible multiple source program files can be included in one job log xml file so each source file is assigned unique id. The name and unique id for each file can be extracted from tasklogs/tasklog/Attributes/ itemid and item label elements respectively.

More Information about source file path and version is stored at the tasklogs/tasklog/appitem/task level.

Input/Output - Files/Datasets Information

Input/ Output can be further divided into two categories:

1. **Folder Level:** Initiating libname statement to a data folder can result in folder level input which means that all datasets in an existing folder are considered as input or including macros folder path in sasautos can result in all macros as input. As a Best Practice SDD does provide feature where you can explicitly select which files and datasets are to be used as an input to current source program simply by selecting those desired items even for the folder level inputs.

Folder level information for files and datasets is captured in joblog XML file under /container/container/item and tableitem elements respectively.

2. **File Level:** Explicitly including files/ datasets in source program.
File level information for files/datasets is captured in joblog XML file under /container/item and tableitem elements respectively.

Step 3: Extract the Information

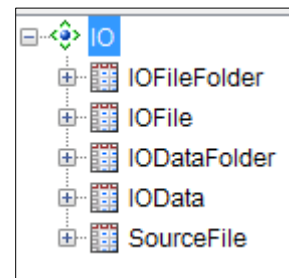
Using SAS XML Mapper the information extraction could be divided into five different datasets depending upon the information captured in the joblog.xml file.

- Folder File(s) Information
- Folder Dataset(s) Information
- File(s) Information
- Dataset(s) Information
- Source File Information

Folder Level Information

The information to be extracted can be broken down into various categories:

Source File Information: Multiple source files can be present in a single joblog file. To distinguish which Input/output corresponds to which source file, its important to extract source file information. The source file information can be extracted from /joblog/tasklogs/tasklog node.

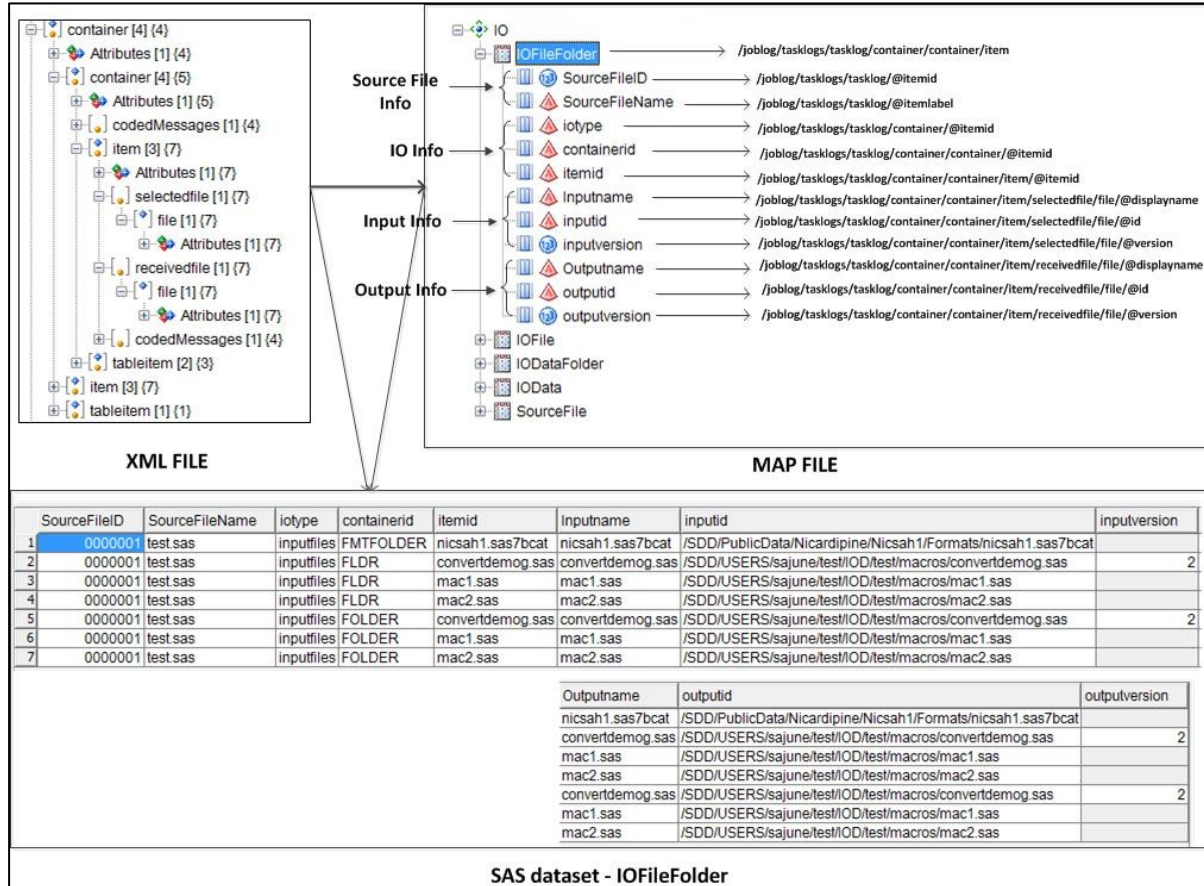


PhUSE 2011

IO Information: Input/Output Information lets you distinguish what is input and what is the output. This information can be extracted from /joblog/tasklogs/tasklog/container node.

Input File(s): The information about the input file(s) is stored under node - /joblog/tasklogs/tasklog/ container/ container/ item/selectedfile

Output File(s): The information about the output folder(s) is stored under node-joblog/tasklogs/tasklog/ container/ container/ item/receivedfile



Various attributes can be extracted both for the input and output files. Core among them are input/output filename, file path and file version.

File Level Information

Source File Information: As explained above, the source file information can be extracted from /joblog/tasklogs/tasklog node.

IO Information: Input/Output Information lets you distinguish what is input and what is the output. This information can be extracted from /joblog/tasklogs/tasklog/container node.

Input File(s): The information about the input file(s) is stored under node-joblog/tasklogs/tasklog/ container/ item/selectedfile

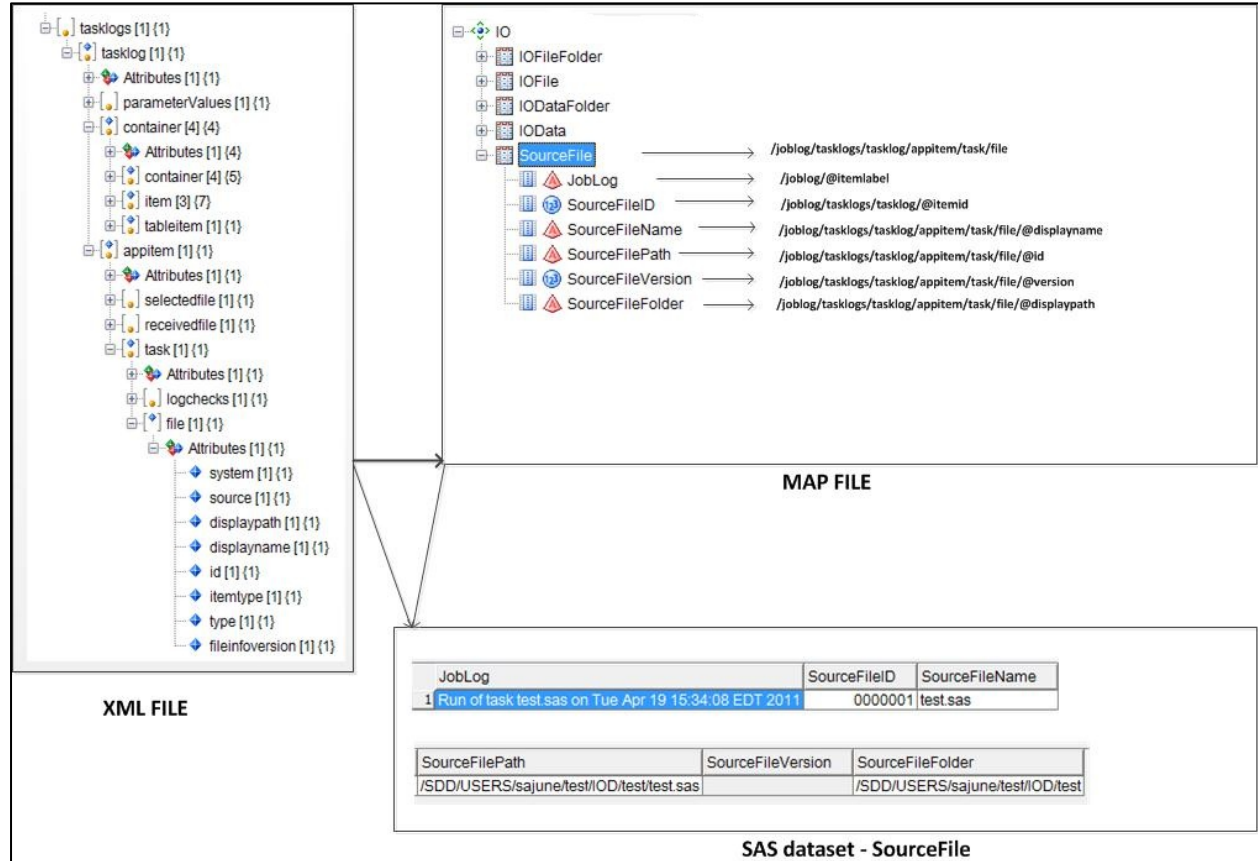
Output File(s): The information about the output folder(s) is stored under node-joblog/tasklogs/tasklog/ container/ item/receivedfile

Source File Information

The SourceID and name information is contained in other datasets. More descriptive information regarding Source files can be extracted and stored in a datasets for source file information.

PhUSE 2011

The information about the source file location and last Execution time for the joblog can be extracted and can be added to the source file dataset.



Step 4: Generate the report:

The information extracted in above step can be merged to create one master dataset and using ExcelXP tagset an

Input/Output Summary Report									
JobLog XML FileName	Source FileName	Source FilePath (Version)	Input/Output Type	Category	Display Name	Input File (Version)	Output File (Version)	Last Run Date/Time	
Multiple Jgml	test1.sas	/SDD/USERS/sajune/test/OD/gap5/test1.sas	INPUT	DATASET	demo.sas7bdat	/SDD/USERS/sajune/test/OD/gap5/data/demo.sas7bdat (1)		FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test1.sas	/SDD/USERS/sajune/test/OD/gap5/test1.sas	INPUT/OUTPUT	FILE	support.sas	/SDD/USERS/sajune/test/OD/gap5/support.sas	/SDD/USERS/sajune/test/OD/gap5/output/support.sas (1)	FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test1.sas	/SDD/USERS/sajune/test/OD/gap5/test1.sas	OUTPUT	DATASET	demo_newunits.sas7bdat		/SDD/USERS/sajune/test/OD/gap5/data/demo_newunits.sas7bdat (3)	FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test1.sas	/SDD/USERS/sajune/test/OD/gap5/test1.sas	OUTPUT	FILE	rfoutput.rtf		/SDD/USERS/sajune/test/OD/gap5/output/rfoutput.rtf	FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test2.sas	/SDD/USERS/sajune/test/OD/gap5/test2.sas (14)	INPUT	DATASET	demo.sas7bdat	/SDD/USERS/sajune/test/OD/gap5/data/demo.sas7bdat (1)		FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test2.sas	/SDD/USERS/sajune/test/OD/gap5/test2.sas (14)	INPUT	FILE	FormatCode.sas	/SDD/USERS/sajune/test/OD/gap5/format/FormatCode.sas		FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test2.sas	/SDD/USERS/sajune/test/OD/gap5/test2.sas (14)	INPUT/OUTPUT	DATASET	demo_newunits.sas7bdat	/SDD/USERS/sajune/test/OD/gap5/output/demo_newunits.sas7bdat (10)	/SDD/USERS/sajune/test/OD/gap5/output/demo_newunits.sas7bdat (10)	FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test2.sas	/SDD/USERS/sajune/test/OD/gap5/test2.sas (14)	INPUT/OUTPUT	FILE	car2.sas7bcat	/SDD/USERS/sajune/test/OD/gap5/output/car2.sas7bcat	/SDD/USERS/sajune/test/OD/gap5/output/car2.sas7bcat	FRI APR 29 10:13:54 EDT 2011	
Multiple Jgml	test2.sas	/SDD/USERS/sajune/test/OD/gap5/test2.sas (14)	OUTPUT	FILE	rfoutput.rtf		/SDD/USERS/sajune/test/OD/gap5/output/rfoutput.rtf	FRI APR 29 10:13:54 EDT 2011	

Excel report with auto filter option ON can be generated which can be used to carry out impact analysis for update to any of the input source file on the other programs.

CONCLUSION

The above report provides complete Traceability Matrix containing full details about various inputs/outputs; their locations and their version number which can help do Impact Analysis for any future update to existing code.

ACKNOWLEDGMENTS

I would like to thank Ben Bocchicchio and Frank Roediger who guided me in the development of this utility

PhUSE 2011

CONTACT INFORMATION

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

Sandeep Juneja

SAS Institute Inc

720 SAS Campus Drive

Cary, NC 27518

Phone:919-531-0541

Email: Sandeep.Juneja@sas.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