

Archival Tool: Automation using PowerShell

Sumit Pratap Pradhan, Syneos Health

ABSTRACT

In clinical trials, patient data is organized into standard format using SDTM. Further, Statistical analysis is performed using Analysis dataset (ADaM) and reporting (table, listing and figure). Several deliveries (e.g. dry run, draft run etc.) are performed for each clinical study. For each delivery, Biostat programming team performs following tasks at minimum:

1. Datasets (SDTM and/or ADaM) and/or TFLs (tables, listings and figures) are sent to sponsor
2. Backup is created to save files related to each delivery. For example - Specification, Programs, Outputs (datasets and reports), log files etc.
3. Upload documents on Trial Master File (TMF) which is needed for effective monitoring and supervision (audit).

All above mentioned tasks can be performed manually but it will consume a lot of unnecessary time, and consistency can not be ensured. For example - Programmers might choose different sets of files as backup based on their judgement. Automation is the best Solution for this scenario.

In this paper, Archival Tool will be explained which has been developed using PowerShell. PowerShell is a cross-platform task automation and configuration management framework, consisting of a command-line shell and scripting language. GUI can be defined easily to take user input using PowerShell. Archival Tool has following features:

- Three primary modules - Delivery, Archive and TMF
- Package creation for each sponsor delivery
- Archive module creation for back-up of each delivery
- TMF module creates a collection of zip files and a supporting table of contents (TOC) that can be directly uploaded to the TMF

INTRODUCTION

The Archival Tool is designed to prepare files for Trial Master File (TMF) submissions, sponsor deliveries and archival for any clinical study where dataset (SDTM or ADaM) and reporting (table, listing, figure) are involved. Aim of the tools is to reduce the manual effort needed in copy and paste files from one location to another. Tool has three modules. A brief overview of these three modules is provided below:

1) Delivery

This module creates a package for each sponsor delivery.

2) Archive

This module creates a back-up for each delivery.

3) TMF

This module creates a collection of zip files and a supporting table of contents (TOC) that can be directly uploaded to the TMF.

Three PowerShell scripts (delivery.ps1, archive.ps1, tmf.ps1) have been written - one for each module. Main.ps1 is driver program and will call each module based on user selection. Below is the screenshot showing how to run the tool. See Figure 1

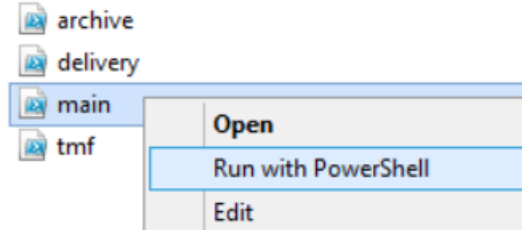


Figure 1. Running Archival Tool

Below GUI window will appear after running main.ps1 (See Figure 2). Once user selects a module, next GUI window will appear to ask specific details for selected module. Two options are provided for each module – Full Version and Define Only. If user selects Define Only, only Define XML files will be copied. If user selects Full Version, all files will be copied providing an option to either include or exclude Define XML files.

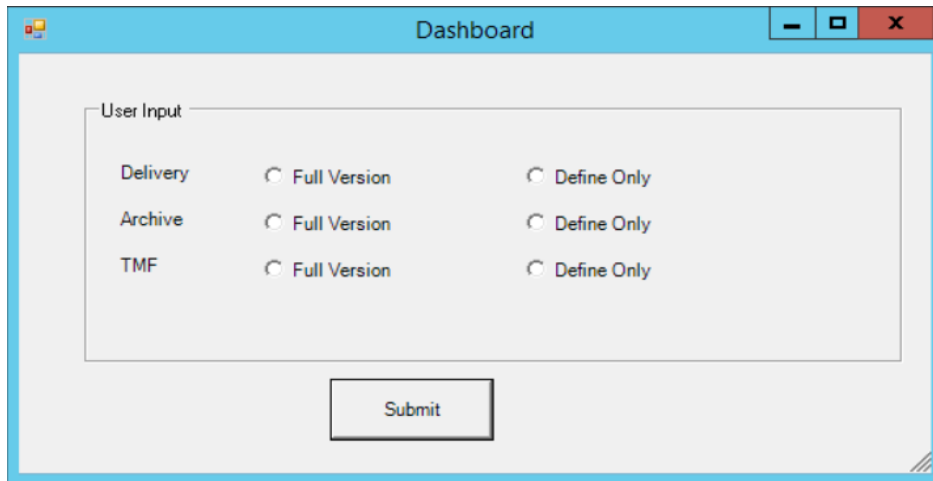


Figure 2. Dashboard for Archival Tool

Details of each module is mentioned in subsequent sections.

PREREQUISITE

Following are the Prerequisite/assumptions for tool:

- Each study must have a study-tracker (In excel format) where details like item name (dataset or TFLs), delivery date, program name and its corresponding output name should be present at minimum. Structure (i.e. columns to be included in study-tracker) and location of the tracker must be fixed across all studies. Tracker can have multiple tabs - one for each SDTM datasets, ADaM datasets and TLF. Below is the screenshot for TFL tab. See Figure 3.

Output Name	TFL Title	Delivery Date	Program Name
t-14-01-01-01	Subject Enrollment and Disposition by Dose Level and Overall	5-Apr-20	t-14-01-01-01.sas
t-14-01-02-01-01	Demographic and Baseline Characteristics by Dose Level and Overall	3-Apr-20	t-14-01-02-01.sas
t-14-01-02-01-02	Demographic and Baseline Characteristics by Dose Level and Overall	3-Apr-20	t-14-01-02-01.sas
t-14-01-02-01-03	Demographic and Baseline Characteristics by Dose Level and Overall	3-Apr-20	t-14-01-02-01.sas
t-14-01-02-01-04	Demographic and Baseline Characteristics by Dose Level and Overall	3-Apr-20	t-14-01-02-01.sas

Figure 3. Fields of study-tracker for TFL tab

Similar structure could be followed for SDTM and ADaM dataset.

Tool will read excel in background and use delivery date column for copying items. If delivery is scheduled on 05-Apr-20, only 1st item (e.g. t-14-01-01-01) will be copied.

If multiple outputs are created by single program, study-tracker should have one entry for each output. Refer to last 4 rows of Figure 3.

- Directory structure must be fixed across all studies. For example - table related files(program, log and output) can follow below mentioned structure:

.\\study-number\\biostat\\area-name\\tfl\\table\\output

.\\study-number\\biostat\\area-name\\tfl\\table\\log

.\\study-number\\biostat\\area-name\\tfl\\table\\program

Study-tracker can be saved under documents location:

.\\study-number\\biostat\\area-name\\documents\\tracker

DELIVERY MODULE

ASSUMPTIONS

Following are the assumptions for this module:

- Data will be posted in delivery folder under different subfolders. Example - consider the situation in which the tool is run on 10JUL2020(Dry Run) and 15SEP2020(Final Run). Data will be saved in the below folders:

.\\study-number\\biostat\\area-name\\delivery\\2020-07-10-t-12-12-06_Dry_Run

.\\study-number\\biostat\\area-name\\delivery\\2020-09-15-t-12-12-06_Final_Run

Time is included in folder name to make it unique even if someone runs the tools multiple times on the same day.

- ERROR.txt file will be generated - This file contains error messages (if found any) while copying files. Example - If t-14-01-02.rtf should be copied for delivery as per study-tracker but it is not present on table output location, following message will be printed

Does Not Exist: .\\study-number\\biostat\\area-name\\tfl\\table\\output\\t-14-01-02.rtf

RUNNING TOOL

Process will be explained in following steps:

- 1) Consider user selected Full Version option for delivery module (refer to Figure 2. Dashboard for Archival Tool). Next window will appear as shown below and user needs to provide details for mentioned fields. See Figure 4.

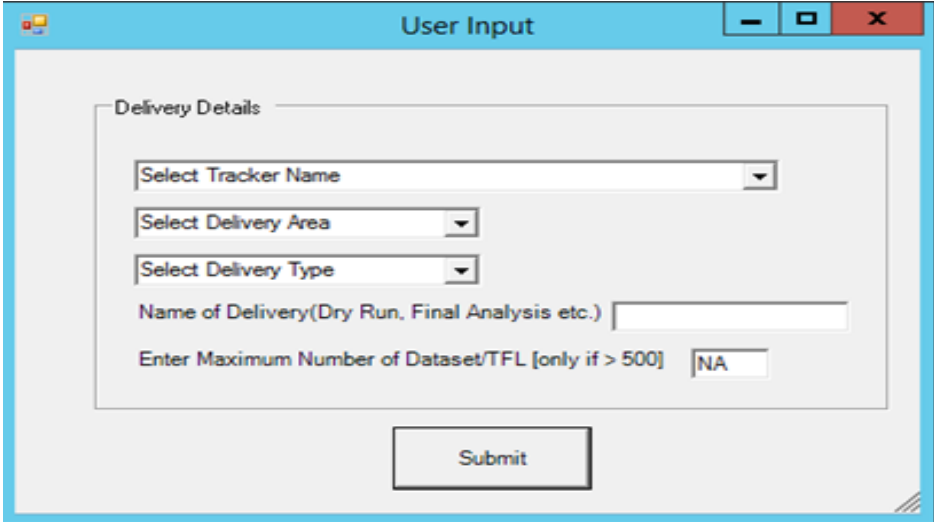


Figure 4. Fields of Delivery Module

Details of the fields are mentioned below:

Select Tracker Name - Dropdown will display list of all study-trackers available at predefined location for tracker. There might be multiple trackers available. Example – study-tracker-dry-run.xls, study-tracker-draft-run.xls etc. User needs to select one study-tracker based on delivery.

Select Delivery Area - Location from where files should be copied like 1-main, 2-drc, 3-dsmb etc. This will display all possible location for a study from where data might be present.

Select Delivery Type - Dropdown will show Component (SDTM, ADAM or ADAM-TFL) for which delivery has to be done. Files will be copied based on selected component. Example - If ADAM-TFL is selected, then ADaM datasets, tables, listings and figures will be copied. If ADAM is selected, then only ADaM datasets will be copied. If SDTM is selected, then only SDTM datasets will be copied.

Name of Delivery (Dry Run, Final Analysis etc.) - Name of the delivery must be entered like Dry Run, Final Analysis, etc. It is brief details of the delivery.

Enter Maximum Number of Dataset/TFL [only if > 500] - By default, 500 rows of study-tracker will be checked. If count of datasets or TFLs are more than 500, enter the count like 700, 800 etc. Ignore otherwise.

If any of the above fields has not been selected by user, Tool will display ERROR message on PowerShell window and processing will stop there (refer to Figure 5). Next window will appear only if all options are selected by user correctly.

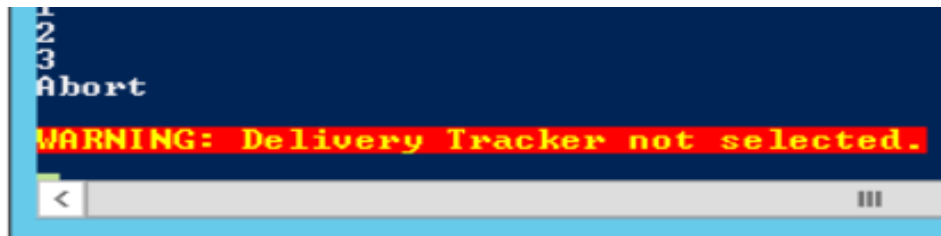


Figure 5. Error message: Study-tracker not selected by user

Example of options selected by user is shown below. See Figure 6

A screenshot of a Windows-style dialog box titled 'User Input'. The dialog has a light blue border and standard window controls (minimize, maximize, close) in the top right corner. Inside the dialog, there is a section titled 'Delivery Details' with a light gray background. This section contains five input fields: a dropdown menu with '00011-study\tracker.xlsx' selected, a dropdown menu with '4-dsmb' selected, a dropdown menu with 'ADaM-TFL' selected, a text input field with 'DSMB' entered, and another text input field with 'NA' entered. Below these fields is a 'Submit' button.

Figure 6. Options selected by User

- 2) Below window will be shown next if all options are selected correctly inConsider1)ConsiderConsiderConsiderConsiderConsiderConsiderConsiderConsiderConsider. See Figure 7.

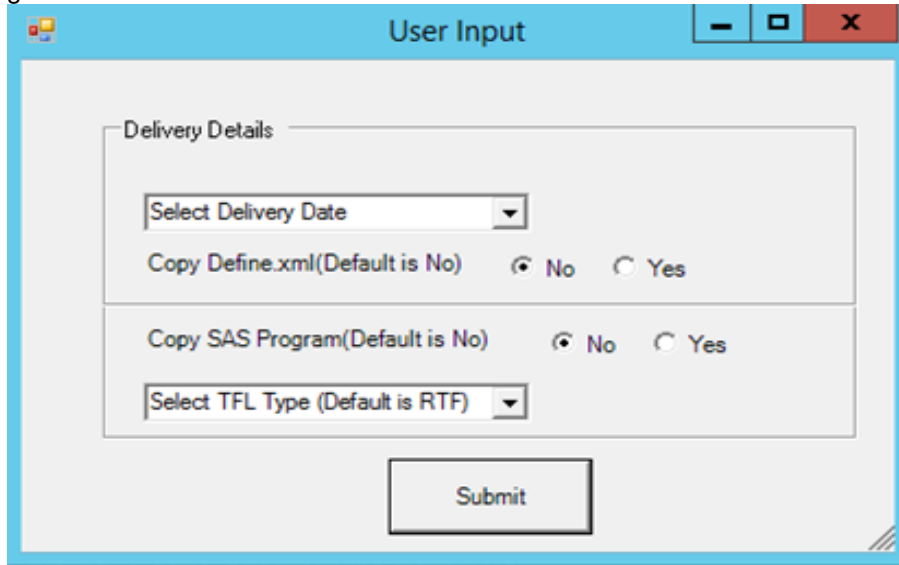


Figure 7. Fields of Delivery Module

Select Delivery Date - All delivery dates from study-tracker, for selected component (SDTM, ADAM or ADAM-TFL), will appear. User can select one date from dropdown list based on delivery. All items mentioned in study-tracker for selected delivery date will be copied.

Copy Define.xml (Default is No) - Define XML will be copied only if user selects Yes.

Copy SAS Program (Default is No) - SAS Programs will be copied only if user selects Yes.

Select TFL Type (Default is RTF) - User can use this to select the type of files to be copied. Options available are RTF, PDF and BOTH. If user does not select anything, RTF files will be copied by Default. PDF files will not be created by tool and will be copied only if it is present at study location.

If delivery date is not selected by user, tool will generate ERROR message on PowerShell window (similar to Figure 5). Rest of the other fields have Default value even if user doesn't select anything.

- 3) If user selects options correctly mentioned in Below , files will be copied to delivery folder like below. See Figure 8.

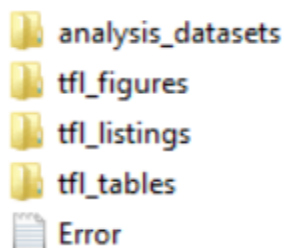


Figure 8. Files copied by Delivery Module

- 4) Before doing sponsor delivery these files must be checked manually to ensure all files are copied. If few files are specific to particular study/sponsor and are not copied by tool, this can be added by programmer after package is created.

- 5) For Define Only option mentioned in Figure 2. Dashboard for Archival Tool, only subset of fields (shown in Figure 4. Fields of Delivery Module) will be shown for user selection. Rest of the process will be same.

ARCHIVE MODULE

ASSUMPTIONS

Following are the assumptions for this module:

- Data will be posted in archive folder under different subfolders. Example - consider the situation in which the tool is run on 10JUL2020(Dry Run) and 15SEP2020(Final Run). Data will be saved in the below folders:
 .\study-number\biostat\area-name\archive\2020-07-10-t-14-12-06_Dry_Run
 .\study-number\biostat\area-name\archive\2020-09-15-t-15-12-06_Final_Run
 Time is included in folder name to make it unique even if someone runs the tools multiple times on the same day.
- Tool will copy fixed number of files. List of files to be included should be decided based feedback received from all key stakeholders (e.g. programmers, managers etc.) involved. Study-tracker has no role here. Files copied will depend upon the component (SDTM, ADAM, ADAM-TFL) selected by user
- ERROR.txt file will be generated – This file contains the error messages (if found any) while copying files. Example - If no content is available under data-issue-tracker location, following message will be printed

Empty Folder: .\study-number\biostat\issue-tracker

RUNNING TOOL

Process will be explained in following steps:

- 1) Consider user selected Full Version option for archive module in Figure 2. Dashboard for Archival Tool. Next window will appear as shown below and user needs to provide details for mentioned fields. See Figure 9.

Figure 9. Fields of Archive Module

Details of the fields are mentioned below:

Select Delivery Area - Location from where files should be copied like 1-main, 2-drc, 3-dsmb etc. This will display all possible location for a study from where data might be present.

Select archive Type - Dropdown will show Component (SDTM, ADAM or ADAM-TFL) for which delivery has to be done. Files will be copied based on selected component. Example - If ADAM-TFL is selected, then ADaM datasets, tables, listings and figures, documents related to ADAM-TFL will be copied, SDTM related files will not be copied in this case.

Copy Define.xml (Default is No) - Define XML will be copied only if user selects Yes.

Name of Delivery (Dry Run, Final Analysis etc.) - Name of the delivery must be entered like Dry Run, Final Analysis, etc. It is brief details of the delivery.

If any of the above fields has not been selected by user, tool will generate ERROR message on PowerShell window (similar to Figure 5) and processing will stop there.

- 2) If user selects options correctly mentioned in Figure 9, files will be copied to archive folder. All the files will be set to READ-ONLY mode and single ZIP file will be created. ZIP file is created to save the disk space. See Figure 10


Name	Date modified	Type	Size
 2020-09-20-t-15-29-46_Dry_RUN	9/20/2020 3:31 PM	Compressed (zipp...	55,198 KB

Figure 10. Zip file created by Archive Module

- 3) For **Define Only** option mentioned in Figure 2. Dashboard for Archival Tool, only subset of fields (shown in Figure 9. Fields of Archive Module) will be shown for user selection. Rest of the process will be same.

TMF MODULE

TMF module is almost similar to Delivery module. Files will be copied based on the delivery date mentioned in study-tracker and options selected by user. Below mentioned features are important where TMF module differs from Delivery module.

- Fixed number of ZIP files and corresponding TOC files will be created based on component selected (SDTM, ADAM, ADAM-TFL). See Figure 11 These files can uploaded to TMF portal directly.

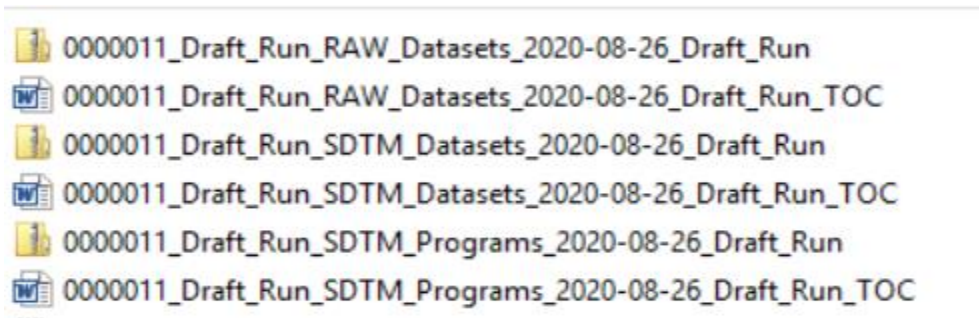


Figure 11. TMF package for SDTM - Zip and TOC files

- TOC is RTF file having details of content of corresponding ZIP file. For each ZIP file, there will be one TOC file. Example is shown below. See Figure 12.

File	Date	Size
ae.sas7bdat	31Aug2020 06:36:05	1048576 bytes
be.sas7bdat	31Aug2020 06:36:07	1048576 bytes
cm.sas7bdat	31Aug2020 06:36:07	2097152 bytes
co.sas7bdat	31Aug2020 06:36:08	2097152 bytes
dd.sas7bdat	31Aug2020 06:36:10	1048576 bytes
dm.sas7bdat	31Aug2020 06:36:10	1048576 bytes

Figure 12. Content of TOC file for SD TM datasets

ROLE OF POWERSHELL

Archival Tool has been developed using PowerShell. Below are pointers which will explain how PowerShell has helped to automate this process.

- Different forms are designed to take input from user. PowerShell script is used to design all forms
- PowerShell script is used to search a location and list specific files meeting criteria into drop-down field. Example - **Select Tracker Name** Dropdown field
- PowerShell script is used to apply all validation checks. Example - If user doesn't enter **Name of Delivery (Dry Run, Final Analysis etc.)** in delivery module, Error message (similar to Figure 5) will be shown on PowerShell window and processing will stop.
- PowerShell script is used to read the Excel file in background mode. Based on study-tracker selected in Figure 4, selected excel will be read in background and all available dates in delivery date column will be shown in Figure 7 under **Select Delivery Date** dropdown field.
- PowerShell script is used to create new folder and copy files from one location to another.
- PowerShell script is used to create text file. Example - ERROR.txt
- PowerShell script is used to create ZIP file for Archive module
- Content for TOC is saved in temporary file using PowerShell script. TOC file is created using SAS program which is submitted in batch mode using PowerShell.

CONCLUSION

Archival Tool is developed to reduce manual effort, minimize unnecessary time to perform repetitive task. Tool will ensure consistency and accuracy across studies especially for backup and TMF related activity. Since PowerShell is open source and pre-installed on Citrix environment (e.g. Windows Server 2012 R2), it requires no overhead for IT team to install any software.

REFERENCES

1. PowerShell Documentation
<https://docs.microsoft.com/en-us/powershell/>

ACKNOWLEDGMENTS

I would like to thank Rohit Srivastava, Martyn Walker, Sabine Laukamp, Gay Knudsen and Chelsea Bradshaw for their continuous guidance and invaluable feedback.

CONTACT INFORMATION

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

Sumit Pratap Pradhan

Syneos Health, Principal Statistical Programmer

Building No. 14, Tower B, DLF Cyber City, Gurgaon - 122002, Haryana, India

E-mail: sumit.pradhan@syneoshealth.com

LinkedIn: <https://www.linkedin.com/in/sumit-pradhan-71133345/>

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