PharmaSUG 2024 - Paper DS-205

A New Way to Automate Data Validation with Pinnacle 21 Enterprise CLI in LSAF

Crystal Cheng, SAS Institute;

ABSTRACT

Pinnacle 21 Enterprise is a software that provides checks on the data compliance with CDISC standards, control terminology and dictionaries when sponsors preparing clinical data submission to regulatory agencies. By validating clinical data early and frequently during the conduction of the clinical trial, it helps users discover data issues and address data issues in advance, ensuring the quality of submission data. There are different ways to execute validations in P21 Enterprise. Users can either manually run the validation via the P21 user interface or, for a more automated process, execute a process flow in SAS® Life Science Analytics Framework (LSAF) to invoke the Enterprise Command Line Interface(ECLI) from P21.

Integrating LSAF with P21 and setting up the validation process via process flow is timesaving for programmers and less prone to errors during packaging and uploading datasets for P21 validation. This paper will focus on the detailed steps to set up the automated process flow of the Pinnacle 21 Validation in SAS® Life Science Analytics Framework (LSAF) and explore the benefits of automating the validation process.

INTRODUCTION

Clinical Data validation is a critical process in clinical research. It is essential not only for sponsors to comply with the strict regulatory requirements for data submissions and review, but also because the data quality can significantly impact the time invested in the research. This paper talks about requirements and steps to successfully set up the P21 ECLI integration with SAS LSAF. Additionally, the paper also discusses the automation process of ECLI execution via LSAF scheduler or process flows, also known as workflow.

Pinnacle 21 ECLI 2.5.0 and LSAF latest release 5.4.1 version were utilized for the integration. LSAF is a cloud based Statistical Computing Environment(SCE) that enables users to streamline the clinical research development process. It allows users to retrieve the raw datasets from various sources, transform them into analysis datasets, perform data validation, and generate reports, all within a single platform.

INTEGRATION REQUIREMENTS

To successfully run the validation job from LSAF, there are few items required to be set up correctly on both LSAF and P21 sides. Failure in following the below set up steps can result common errors that will be discussed in the later section in this paper.

FIREWALLS AND PORTS

To set up the connection, ports TCP 443, TCP 20022, UDP 33001 need to be unblocked from LSAF(Johnston, 2024). Otherwise, the ‘unable to connect to server’ error can occur when executing the validation job. This is the most common error we’ve seen when implementing integrate P21 with LSAF.

API KEY GENERATION

P21 administrator user must create an API key from P21E GUI. The API key needs to be added to the configuration (.conf) file for authentication purposes. The API key functions like password and username pair, so it is important for users to store the key properly and securely. All the permissions control and security features from Pinnacle 21 are still enforced within ECLI.

USER GROUP CREATION AND ASSIGNMENT

A user group named enterprise_cli need to be created and assigned to correct project, study level. This user group is associate with the ECLI request and needs to be specify in the .conf file.

DOWNLOAD THE ECLI INTEGRATION CLIENTS

Before downloading the package zip file, user should verify their Pinnacle 21 Enterprise version. Base on the P21 documentation, if user Pinnacle 21 Enterprise version is 5.4.0, then ECLI 2.5.0 version is required. Older version of ECLI will not be compatible with the P21E 5.4.x (Mankus, 2024).

ECLI 2.5.0 Download link is available on the Pinnacle 21 help center. Within the zip file, user can find the validation job, corresponding SAS program, and macros folder including set of SAS macros and p21-client-2.5.0.jar. The downloaded zip file should be uploaded and expanded in LSAF designated area.

ECLI 2.5.0 download link: https://help.pinnacle21.net/en/articles/8209498-where-can-i-learn-about-ecli-2-5-0

CREATION OF PINNACLE21.CONF FILE

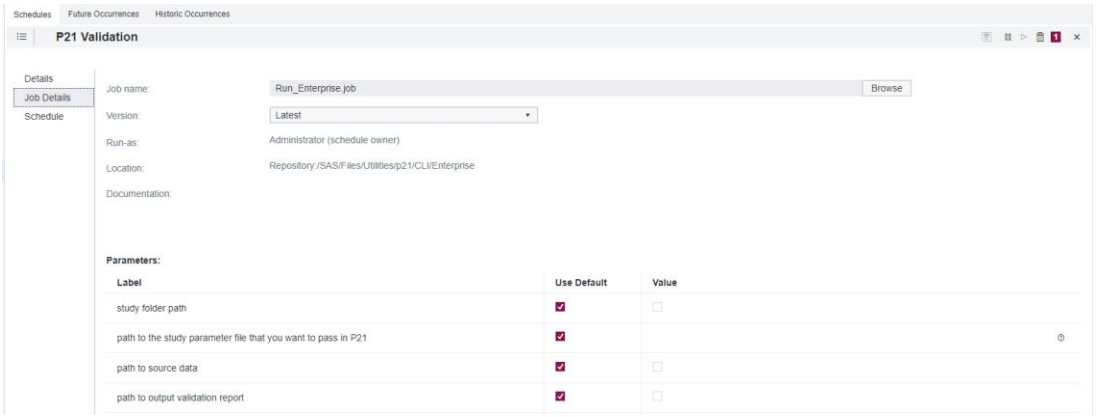
A pinnacle21.conf file is required for the request to send to P21 server. There are additional parameters user must specify in the pinnacle21.conf file. After the pinnacle21.conf file is created, user should upload the file to the 'macros' folder which includes the executable JAR file. Sample pinnacle21.conf file and parameters definition can be found on the Pinnacle 21 help center website (Strigunov, 2024).

VALIDATION EXECUTIONS

After all the set-up steps are completed, user can execute the enterprise validation job in LSAF. There are two different ways to invoke the validation job in LSAF, by scheduling the job in scheduler or setting up a workflow. These two methods both allow user to execute the Pinnacle 21 Validation automatically and then download the data validation report to LSAF designated location.

SCHEDULE JOBS

Validation job execution can be scheduled in the LSAF scheduler. Users have options to define different frequencies of the job execution. Once the job is run per schedule, a notification can be sent to the user. The validation results will be downloaded to designate folder in LSAF. User doesn't need to manually execute job and wait for the validation program to finish if the user would like to have the validation job run on a re-occurring basis.



Display 1. LSAF Schedules Jobs Window

Pro:

- Easy to setup in LSAF Scheduler.
- Minimize manual user interactions on multiple platforms to extract and prepping validation source datasets.
- Schedule job execution detail will be captured in the audit history.

Cons:

- Scheduled job will be executed automatically based on the selected frequency in scheduler in LSAF. Since the execution is solely based on the frequency, the validation report could possibly be duplicate if there are no updates on source data.

WORKFLOW

Workflow is a more automated way to execute the P21 validation job. Users can streamline the data validation process from the EDC source data extraction to report generation and report reviewing. The workflow process enables the user to run data validation dynamically based on the status of the source data feed. If there are no updates on source datasets, then the process will be terminated. However, if the source data extract is updated, then the process flow will move forward to regenerate datasets and the validation job will be triggered. The green flowchart in Figure 1. Displays the subprocess for the P21 validation job. After the validation job is completed, the validation reports are downloaded to an output folder in LSAF. Email notification or LSAF internal messages can be sent to user if the user subscribed the job run event.

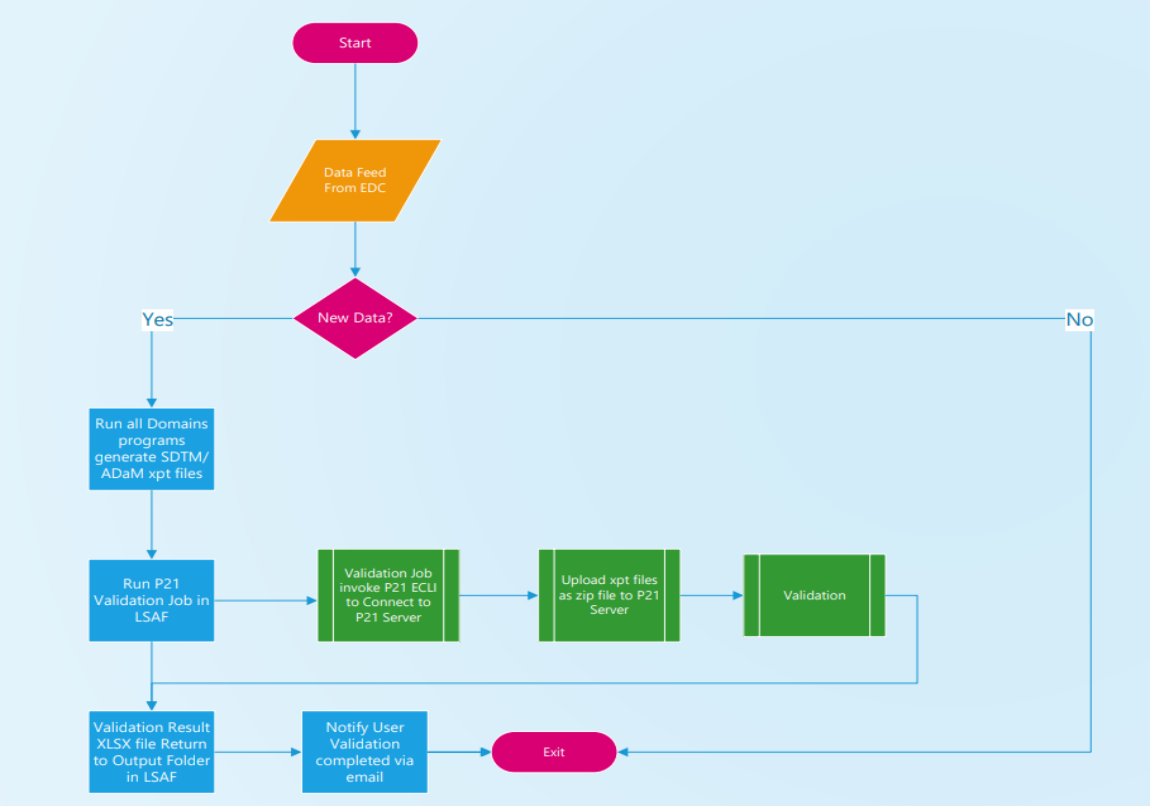


Figure 1. High-level LSAF process flow diagram for Pinnacle 21 validation

Pros:

- Dynamically runs P21 validation job base on the source data status.
- Minimize manual user interactions on GUI.

- Data validation will be run against the latest source data. No redundant report will be generated based on the execution condition it set up in process flow.
- Programmers and DM team users can collaborate easily and closely.
- Admin user/study lead can easily monitor the status of the study data.
- Workflow related execution or actions performed in LSAF repository will be recorded in audit history. Actions in repository are traceable.

Cons:

- BPMN file must be created for workflow definition in LSAF. To initiate the workflow, the workflow definition must be deployed first. Therefore, it required admin user have previous knowledge of BPMN file creation. Once the workflow deployed, other users can use it as it is.
- Extra testing and troubleshooting efforts might require when set up process flow, and the process flow must be validated first before deployment.

BPMN File Creation

To deploy a process flow in LSAF, a BPMN file (Business Process Model and Notation) is required. Users must upload the BPMN file to the workflow definition area to make it available for deployment. A BPMN file is a diagram which can be used to represent the business process. This paper use Eclipse to generate the demonstration BPMN file.

Figure 2 is a P21 validation job BPMN diagram which was created in the Eclipse. As you can see, the diagram consists with different shapes(flow elements), arrows, and descriptive text for each activity. Activities defined in the process flow will be implemented sequentially. Overall, the business process represented in the diagram is easy to follow by technical or non-technical users.

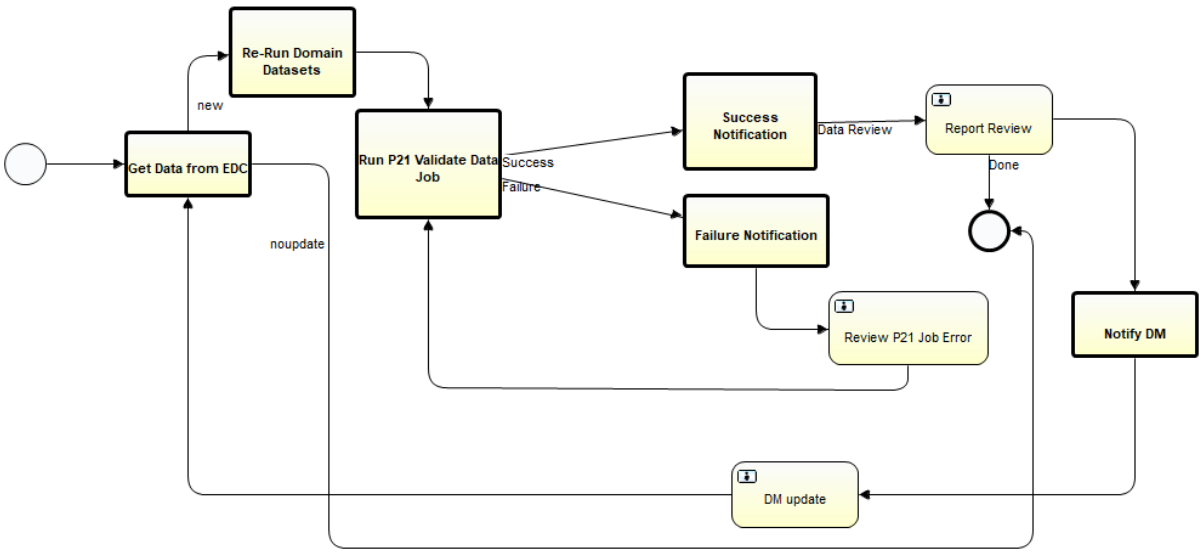
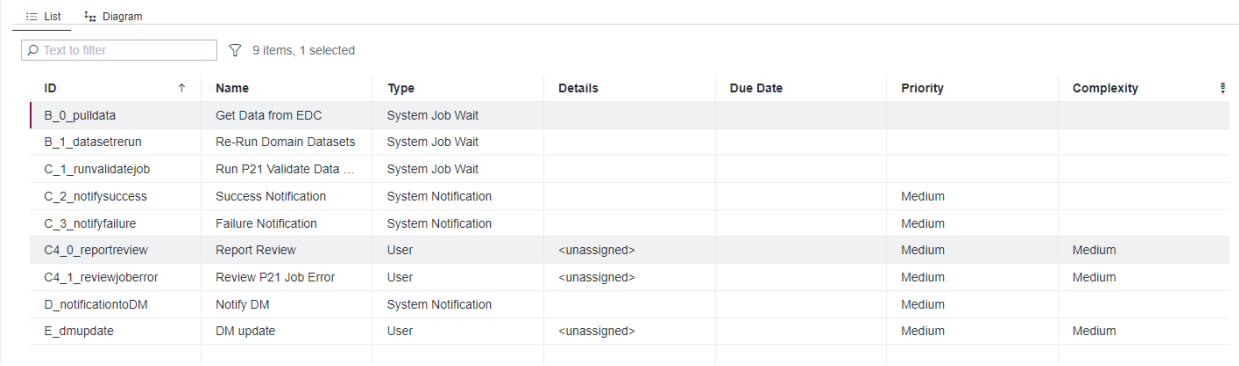


Figure2. BPMN diagram for Pinnacle 21 validation workflow

BPMN File Deployment

After the BPMN file was uploaded in LSAF, the process flow becomes available to deploy and implement. Display 2. shows the list of activities created for data validation in the process flow and users can identify each activity's BPMN type in LSAF.

The first activity in the P21 data validation process flow is extraction of raw datasets from EDC to LSAF. If there are new raw datasets, then the process moves on to pull the raw data to LSAF to a designated folder. New raw datasets will trigger re-run activity on domains. All the domains' programs will be re-run and domain data will be re-populated based off the latest updated raw source datasets. Then the P21 validation job kick off once the domain datasets are repopulated. The P21 Validation Enterprise job has four parameters which are the dynamic run-time values user can enter or preset to default values. Once the job execution is finished, success/failure notifications will be sent to users summarizing job completion status. For successful completion of the job, a P21 validation report will be downloaded to LSAF to the validation report folder. Then user can start the report reviewing process and notify DM team if there are any data issues. If the validation job is completed with errors or warnings, then user needs to review the error and determine if it is a job/P21 set up error or data issue. If the job error relates to set up, user can refer to common errors section in this paper below or refer to Pinnacle 21 help center page for troubleshooting. Once the job errors are fixed, the process will move to re-run the P21 enterprise validation job, and the validation process will be loop over again until user can generate the validation reports without job errors.



ID	Name	Type	Details	Due Date	Priority	Complexity
B_0_pulldata	Get Data from EDC	System Job Wait				
B_1_datasetrerun	Re-Run Domain Datasets	System Job Wait				
C_1_runvalidatejob	Run P21 Validate Data ...	System Job Wait				
C_2_notifysuccess	Success Notification	System Notification			Medium	
C_3_notifyfailure	Failure Notification	System Notification			Medium	
C4_0_reportreview	Report Review	User	<unassigned>		Medium	Medium
C4_1_reviewjoberror	Review P21 Job Error	User	<unassigned>		Medium	Medium
D_notificationtoDM	Notify DM	System Notification			Medium	
E_dmupdate	DM update	User	<unassigned>		Medium	Medium

Display 2. Workflow Details in LSAF

COMMON ERRORS

When the validation job is running, the job can complete successfully or complete with errors or warnings. Below are few common errors observed during the integration. There is a comprehensive list of error codes user can reference on Pinnacle 21 help center page(Johnston, 2024).

1. Unable to Connect to server. Check your connection and firewall settings.

   Troubleshooting Tip:

   Checking firewall setting and required ports has been unblocked on LSAF. User can simply run a proc http procedure in LSAF SAS session to check if the connection has been set up correctly. If the connection has not been set up correctly, then user will receive the error code indicate the connection failed as below example. Otherwise, with correct ports and firewall configured, "200OK" status code is expected.

```
filename hdrout "&_sasws_/users";
proc http
    headerout=hdrout
    url={pinnacle21 URL}
    method="GET"
    ct="application/x-www-form-urlencoded";
run;
ERROR: The tcpSockRead
```

```
call failed. The system error is 'The connection was reset by a peer.'.
ERROR: Call to tcpSockContinueSSL failed.


filename hdrout "&_sasws_/users";
 proc http
      headerout=hdrout
      url={pinnacle21 URL}
 method="GET"
      ct="application/x-www-form-urlencoded"
      proxyhost=
      proxyport=;
 run;
NOTE: 403 Forbidden
```

2. Service Response error [httpCode=403 action=action [httpCode=403 action=action=[method='GET', uri=]

   Troubleshooting tip:

   Pinnacle21 administrator should check if the user/user group has proper permissions granted. This error occurs when user or user group doesn't have sufficient permissions to complete the request via ECLI.

CONCLUSION

Data validation is an essential step in the clinical research. By seamlessly integrating Pinnacle 21 with Life Science Analytics Framework(LSAF) and leveraging the process flow within LSAF, user can streamline their workflow to save time. Rather than toggling between two applications and manually preparing validation datasets, users can perform all the necessary tasks within LSAF environment. Moreover, study team leaders can simply monitor data validation status via workflow status. Once the initial workflow is built, other users should be able to easily deploy the validation process workflow across studies . This streamlined approach not only accelerates the data validation process but also minimizes human error during the preparation of the validation datasets. Lastly, users can always access an up-to-date validation report base on the status of the source data extraction.

REFERENCES

Johnston, Philp.(2024) "What are the system requirements for ECLI?" https://help.pinnacle21.net/en/articles/5677484-what-are-the-system-requirements-for-ecli

Strigunov, Philipp.(2024) "How can I use pinnacle21.conf with ECLI?" *https://help.pinnacle21.net/en/articles/5589574-how-can-i-use-pinnacle21-conf-with-ecli*

Johnston, Philp.(2024) "What do ECLI error codes mean?" https://help.pinnacle21.net/en/articles/5674851-what-do-ecli-error-codes-mean

Mankus, Trevor.(2024) "Where can I learn about ECLI 2.5.0?" https://help.pinnacle21.net/en/articles/8209498-where-can-i-learn-about-ecli-2-5-0

ACKNOWLEDGMENTS

Special thanks to Ben Howell for helping with workflow BPMN file creation and troubleshooting. Laura Elliott and Ben Howell helped with editing; any errors or inaccuracies in this paper are solely the responsibility of the author.

RECOMMENDED READING

- Useful Pinnacle 21 help center documentations when implementing integration.

https://help.pinnacle21.net/en/articles/4357206-what-parameters-are-in-ecli

- BPMN files.

https://www.visual-paradigm.com/guide/bpmn/what-is-bpmn/

CONTACT INFORMATION

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

Crystal Cheng
SAS Institute
Crystal.cheng@sas.com