



Our moonshot – digitally transform
Biopharmaceutical industry's ability
to assess risk as well as reduce time
and cost in clinical trials by delivering
data, analysis, insights....*faster*.

Paper DH13

Data Pipeline – Improving Cycle Time and Shifting Data Closer to Analytics

Greeshma Mundayat Clinical Solutions Owner
greeshma.mundayat@maxisit.com

Sarika Karnik Clinical Solutions Owner
sarika.karnik@maxisit.com

Feb 25 – 28 | PHUSE US Connect 2024 | Maryland



Agenda

Introduction

Industry Challenges, Solutions & Business Impact

Data Pipeline – Bringing data closer to analytics

On Demand data pipelines for Statistical Programming

Conclusion

Introduction

This paper discusses data-oriented challenges within statistical programming processes and presents data pipeline approach that offers faster access to reliable data in the hands of data consumers. Overall approach brings data closer to the analytics and analytics team with ability to initiate pipeline on demand alleviating dependencies on data suppliers. Additionally, it underscores the transformative impact of a fully automated data pipeline with CI/CD, crucial for efficient, high-quality data delivery in regulatory reporting.

Problem

- ❑ Lack of timely access to quality data
- ❑ Data changes impact regulatory reporting
- ❑ Increased manual interventions
- ❑ Negative impacts to regulatory reporting cycle-time due to data dependencies.

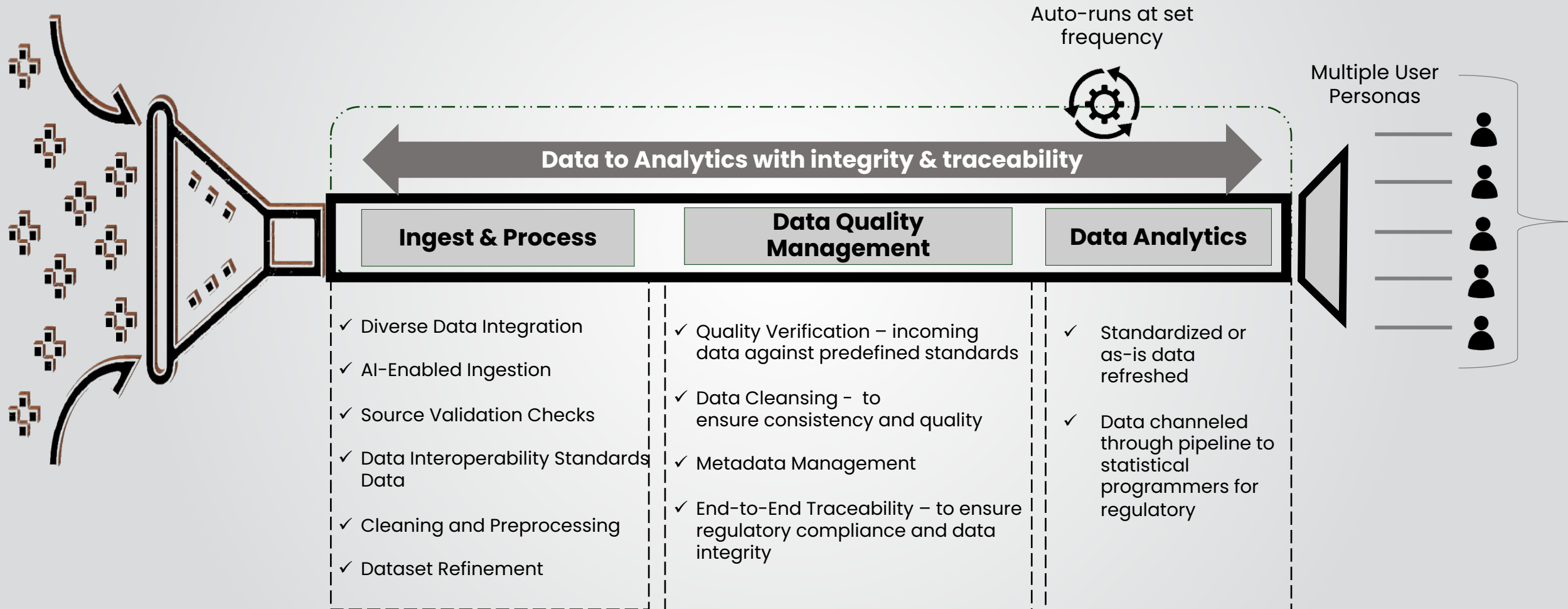
Solution

- ✓ Robust data ingestion and Validation process
- ✓ Fully automated data pipelines by source, by type, and by aggregation
- ✓ Built-in traceability on data changes, data quality and impact to downstream reporting

Business Impact

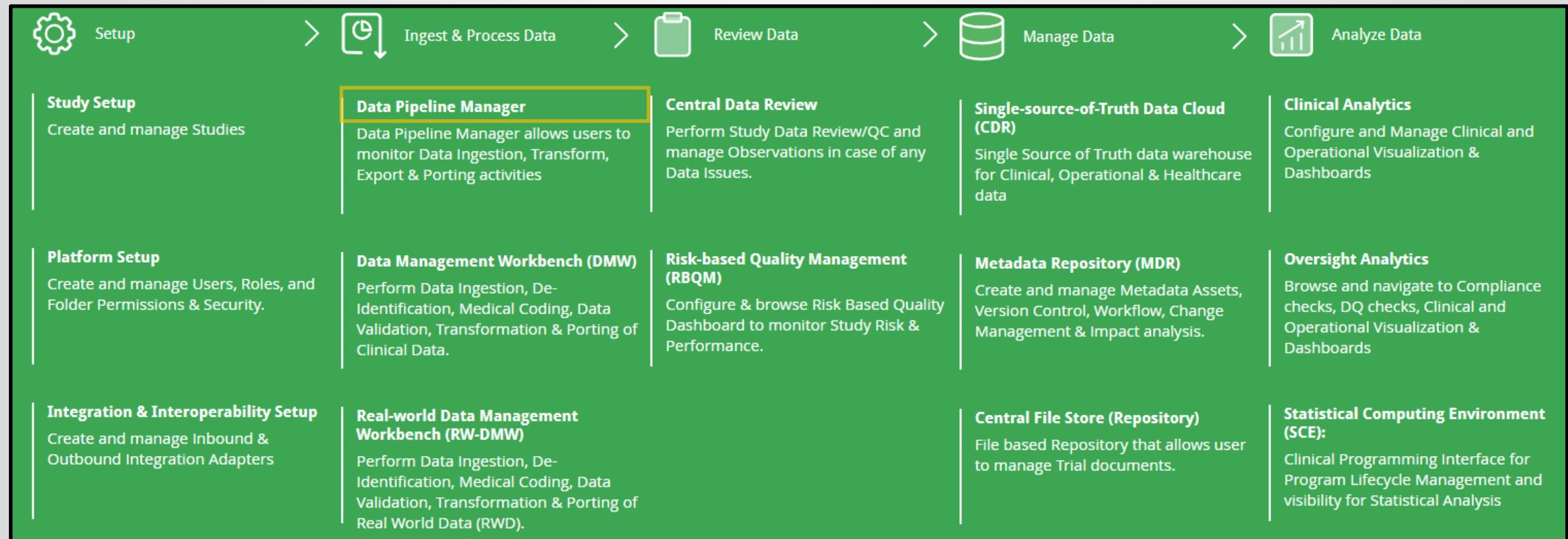
- ❑ Faster Access to Quality Data
- ❑ Improved cycle time on statistical programming
- ❑ End-to-end auditability and traceability
- ❑ Streamline management and reduce delays
- ❑ Enhances efficient communication between stakeholders

Data Pipeline – Bringing data closer to analytics

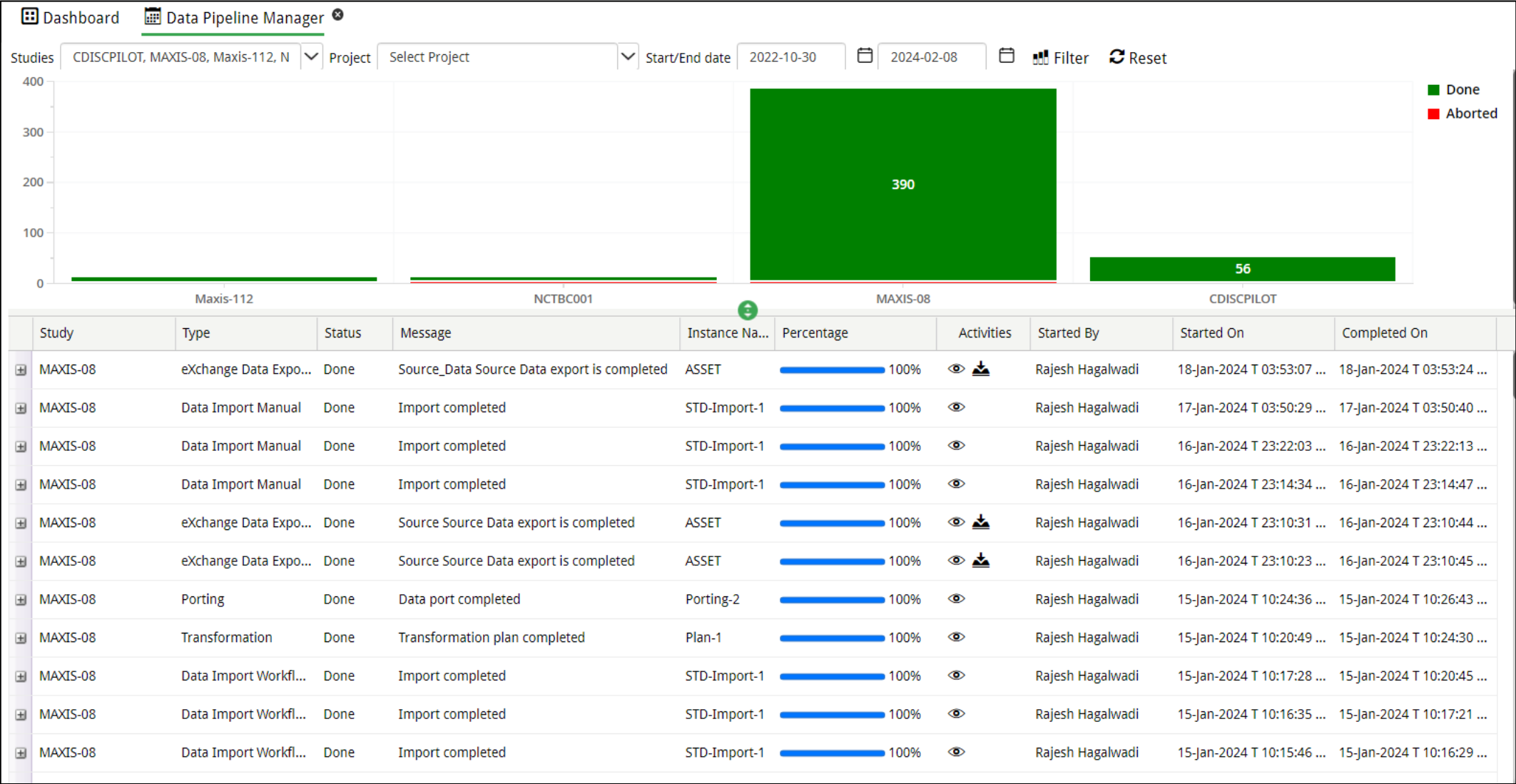


Process Navigator

The Process Navigator allows the user to view the different areas of the data processing steps within the platform



Data Pipeline Manager



Proprietary and Confidential Information

Data Pipeline (Automated Source to Target Data)



Data Pipeline Configuration

Workflow Activities

Activity

-- Select --

-- Select --

Data Import

Data Transformation

Data Porting to CDR

Data Porting to Repository

Validate and Port data to CDR

Validate and Port data to Repository

Activity Name	Project Name	Metadata Name	Data Mode	Plan Name	CDR Project Name	Source	Edit	Delete
Data Import	Clinical_Data_Pro...	Source_Data	Full Load			CRO1_SFTP_Connection(In)		
Data Import	Clinical_Data_Pro...	Source_Data	Full Load			ECG_SFTP_Connection(In)		
Data Import	Clinical_Data_Pro...	Source_Data	Full Load			Biomarker_SFTP_Connection(In)		
Data Import	Clinical_Data_Pro...	Source_Data	Full Load			Rave_EDC_Connection		
Data Transformation	Clinical_Data_Pro...		Full Load	Transformation_...				
Data Porting	Clinical_Data_Pro...		Full Load	Transformation_...	Clinical_Data_Pro...			

Recurring Schedule

Recurrence Pattern

☐ Hourly

☐ Daily

☒ Weekly

☐ Monthly

☐ Quarterly

☐ Continuous

Recur Every

1

Week On:

☐ Monday

☐ Tuesday

☐ Wednesday

☐ Thursday

☐ Friday

☐ Saturday

☐ Sunday

Range of Recurrence

☒ End After

1

occurrences

☐ End By

Active :

☐

Schedule :

08-Feb-2024 T 11:10:00 EST

Save

Data Pipeline Execution History

Workflow History

Schedule Name	Status	Reason for Abort	Start Date	End Date
Clinical_Data_Pipeline	Completed		05-Oct-2023 T 15:00:0...	05-Oct-2023 T 15:13:0...
Clinical_Data_Pipeline	Completed		06-Oct-2023 T 15:00:0...	06-Oct-2023 T 15:14:1...
Clinical_Data_Pipeline	Stopped	File not found	07-Oct-2023 T 15:00:0...	07-Oct-2023 T 15:00:0...
Clinical_Data_Pipeline	Stopped	File not found	08-Oct-2023 T 15:00:0...	08-Oct-2023 T 15:03:2...
Clinical_Data_Pipeline	Completed		09-Oct-2023 T 15:00:0...	09-Oct-2023 T 15:13:0...
Clinical_Data_Pipeline	Completed		10-Oct-2023 T 15:00:0...	10-Oct-2023 T 15:13:2...
Clinical_Data_Pipeline	Completed		11-Oct-2023 T 15:00:0...	11-Oct-2023 T 15:15:1...
Clinical_Data_Pipeline	Completed		12-Oct-2023 T 15:00:0...	12-Oct-2023 T 15:14:0...
Clinical_Data_Pipeline	Stopped	key is null.	11-Jan-2024 T 15:00:0...	11-Jan-2024 T 15:05:4...
Clinical_Data_Pipeline	Completed		12-Jan-2024 T 07:05:0...	12-Jan-2024 T 07:20:0...
Clinical_Data_Pipeline	Stopped	File not found	13-Jan-2024 T 07:05:0...	13-Jan-2024 T 07:06:2...
Clinical_Data_Pipeline	Stopped	User Stopped this schedule	14-Jan-2024 T 07:05:0...	14-Jan-2024 T 07:05:0...
Clinical_Data_Pipeline	Completed		15-Jan-2024 T 10:12:0...	15-Jan-2024 T 10:26:4...
Clinical_Data_Pipeline	Stopped	User Stopped this schedule	16-Jan-2024 T 10:12:0...	16-Jan-2024 T 10:12:0...

FILE

MESSAGE

Ignore

X

Reply

Reply All

Forward

Meeting

Biopharma News

Team Email

Reply & Delete

Delete

Respond

Quick Steps

Move

Rules +

OneNote

Assign Policy

Mark Unread

Categorize

Follow Up

Translate

Find

Related

Select

Zoom

Mon 1/10/2022 7:51 PM

ctrpro

Demo(maxisb) - Study 'MAXIS-08' Data ported successfully

To: Rajesh H

OnePlaceMail

Evernote

Trello

Mostra - Team Collaborati...

Get more apps

MAXIS®

Dear Demo User7,

Data Ported Successfully !

Data Port Details:

Study Name : MAXIS-08

Project Name : Data_Transformation_Project11021

Plan Name : Transformation_Plan_01

Job Id : d1a0d36a-7277-11ec-b185-0050568810c3

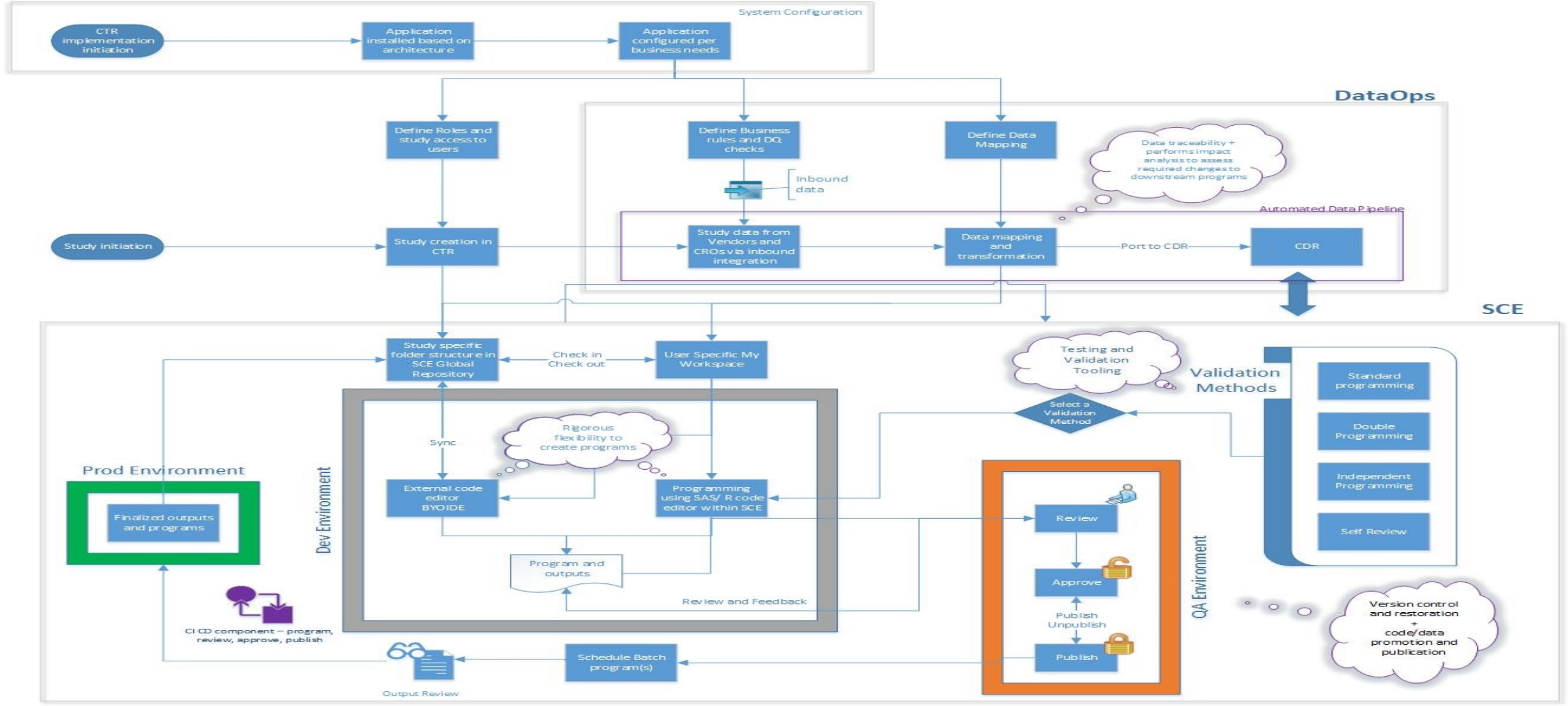
Mapping Specification name and Version : Mapping Specification_01 1.0

Porting Mode : Cumulative

Ported Date and Time : 10-01-2022 19:50:14 EST

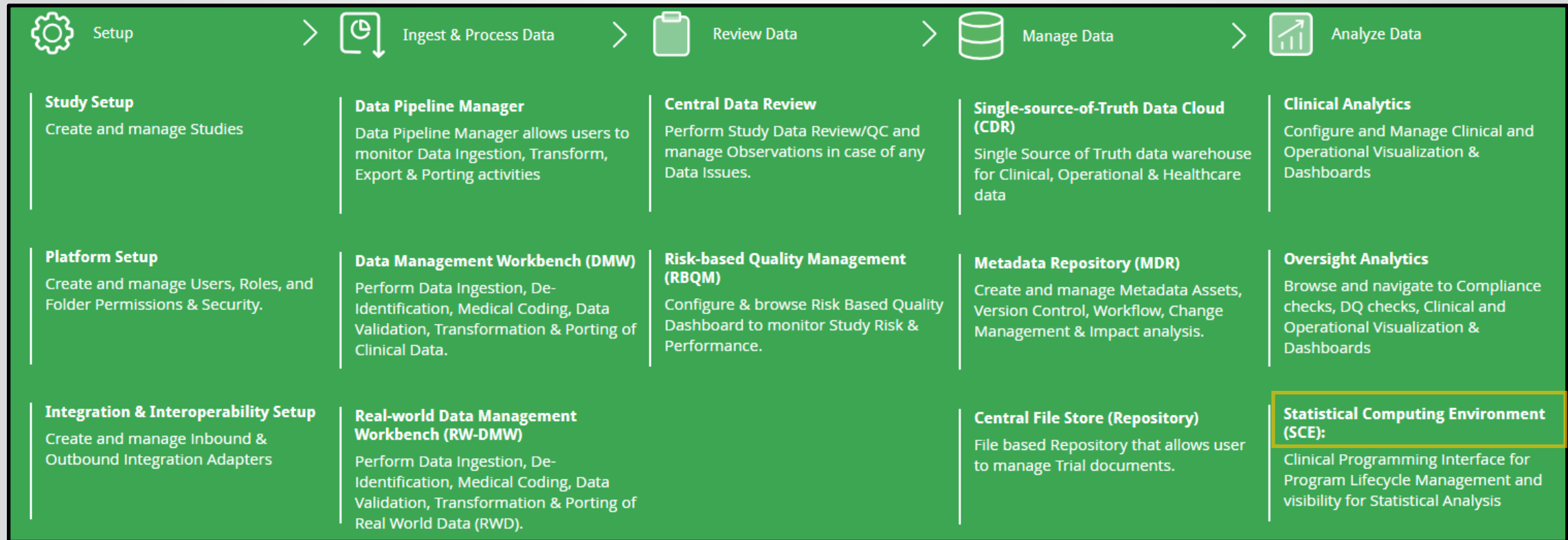
ctrpro: Demo(maxisb) - Data Import Error through SFTP-File not found

On-demand data pipeline enabled statistical computing environment



Process Navigator

The Process Navigator allows the user to view the different areas of the data processing steps within the platform



Output Traceability

- ✓ This will show programming Workplaces, BYIDE ,Environments ,SAS editor etc.

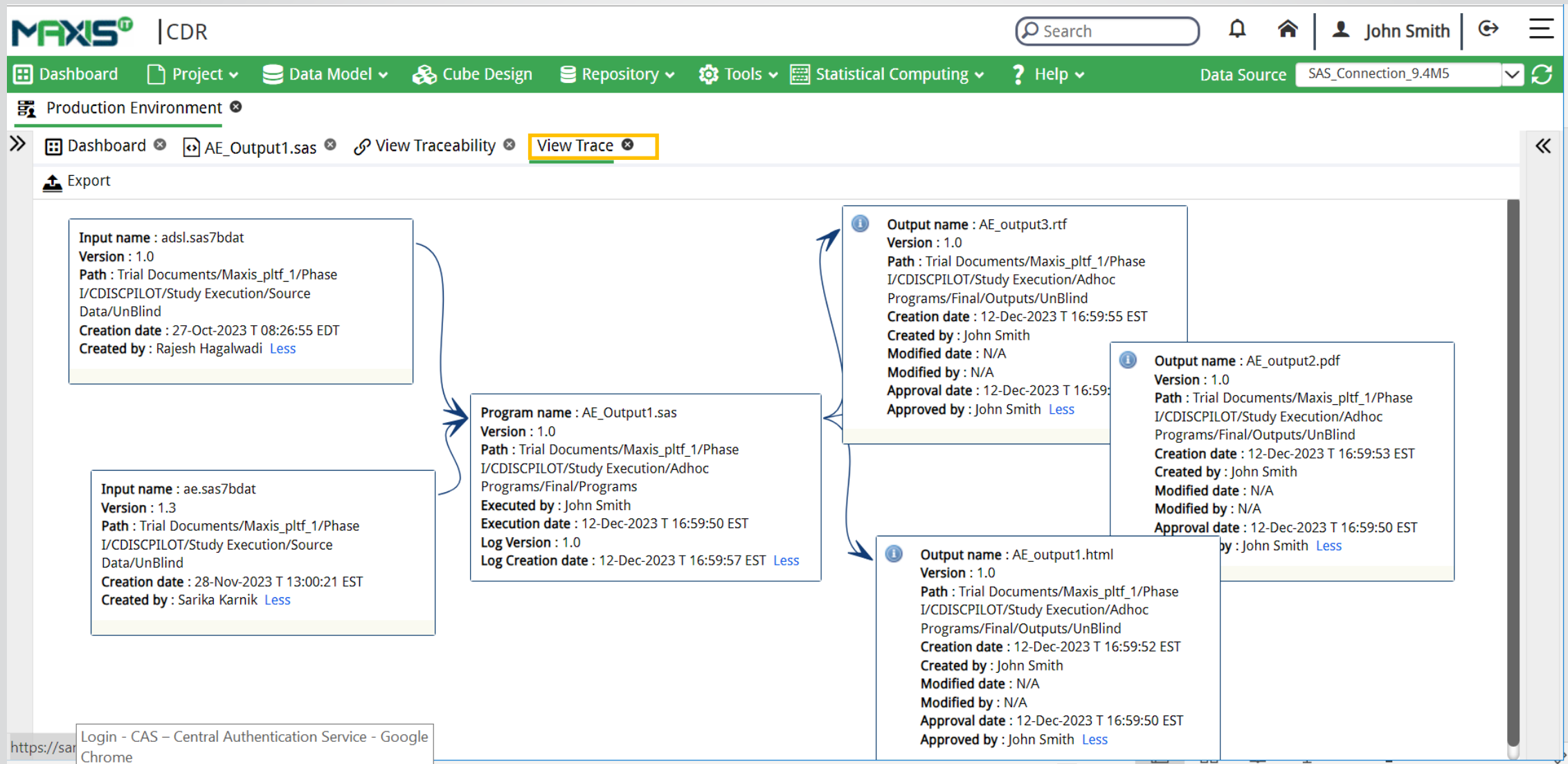
The screenshot displays the MAXIS CDR interface. The top navigation bar includes a search bar, user profile (Sarika Karnik), and various tool icons. The main workspace is divided into three panes:

- Left Pane (Workspace):** Shows a tree view of the project structure. The 'User Workspace' is expanded, showing folders like 'Study Management', 'MAXIS-08', 'Study Execution', 'Adhoc Programs', 'Draft', and 'Programs'. The 'Programs' folder is selected, showing a list of files including 'AE_Summary.sas', 'AE_report.sas', 'AE_OUTPUT.sas', 'AE_report1.sas', and 'ADSL_Program.sas'.
- Center Pane (Code Editor):** Displays the SAS program 'AE_Summary.sas'. The code includes comments, options, and a PROC SQL query. A context menu is open over the code, listing options: 'Development Environment', 'QC Environment', 'Production Environment', 'BYOIDE', 'BYOIDE - Admin', 'Manage SAS Connections', 'Global Share', 'Project Management', and 'Task Lov'.
- Right Pane (Details):** Shows a 'History' tab with a list of activity logs. Each log entry includes the user name, program name, activity type, environment, and timestamp.

The context menu is currently open, showing options for managing the environment and connections. The 'Development Environment' option is highlighted.

Output Traceability

- ✓ This will show end to end traceability for this Program (i.e., Input Data source ,program details and Outputs for the program.





Conclusion

- Increasing data diversity, siloed systems and manual interventions are potential risks to deriving synergies, maintaining data integrity and traceability in the last mile.
- Automated data pipeline prioritize timely access, quality, and end-to-end traceability, while streamlining the data-to-outcome/analysis process
- Maxis IT SCE brings together on demand data pipeline for statistical programming to overcome these challenges



THANK YOU

