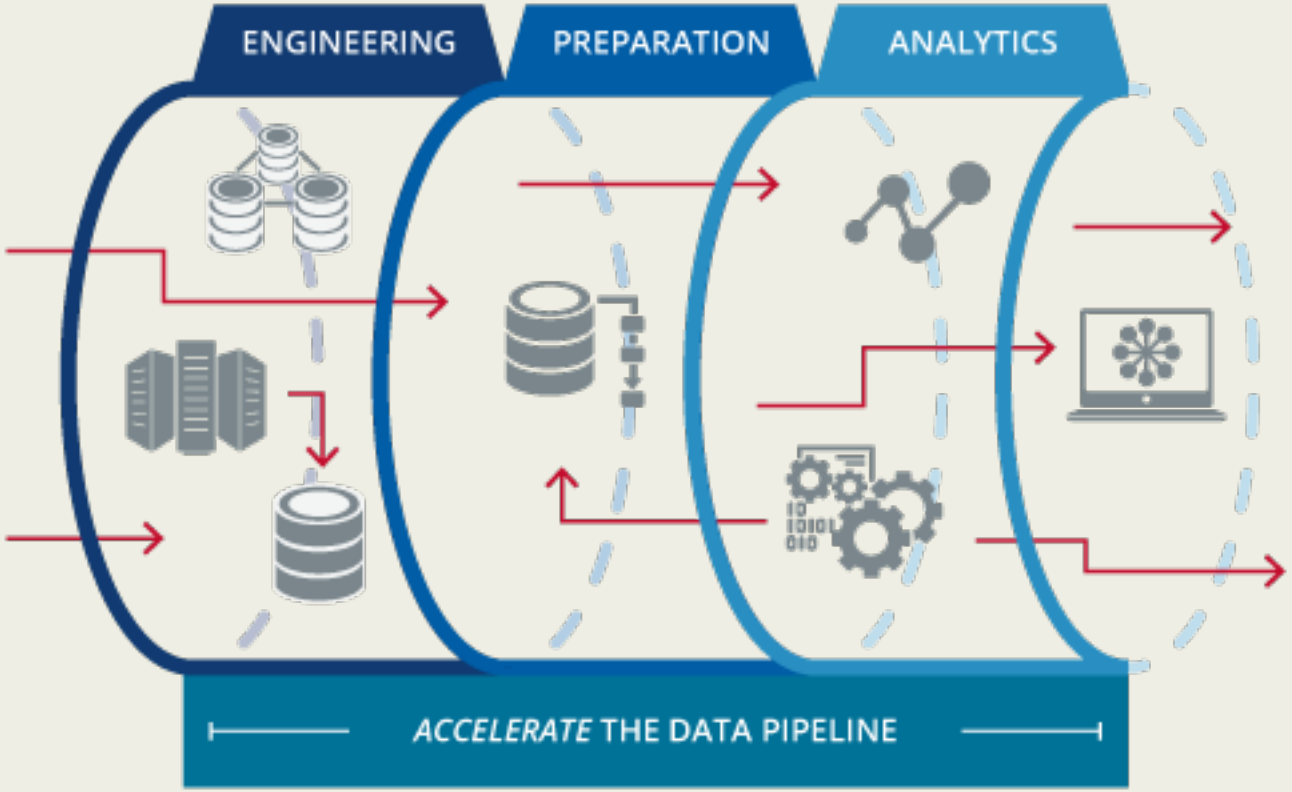
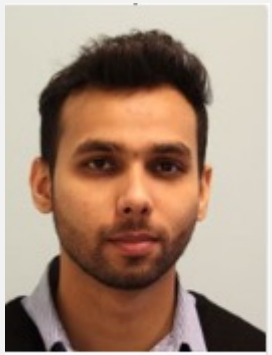


Data Pipeline Automation

Data Pipeline Automation for Streamlined Clinical and Operational Data Review

RINDRA ANDRIANANTENAINA & MUHAMMAD UZAIR KHALID – 10 OCT 2023 – PHUSE SDE 2023

Rindra Andrianantenaina

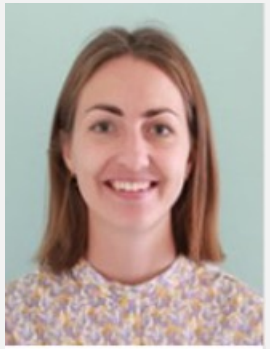


Muhammad Uzair Khalid

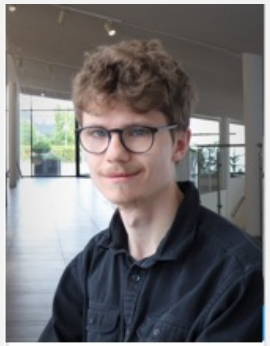
Fei Xie

Sean Bronée

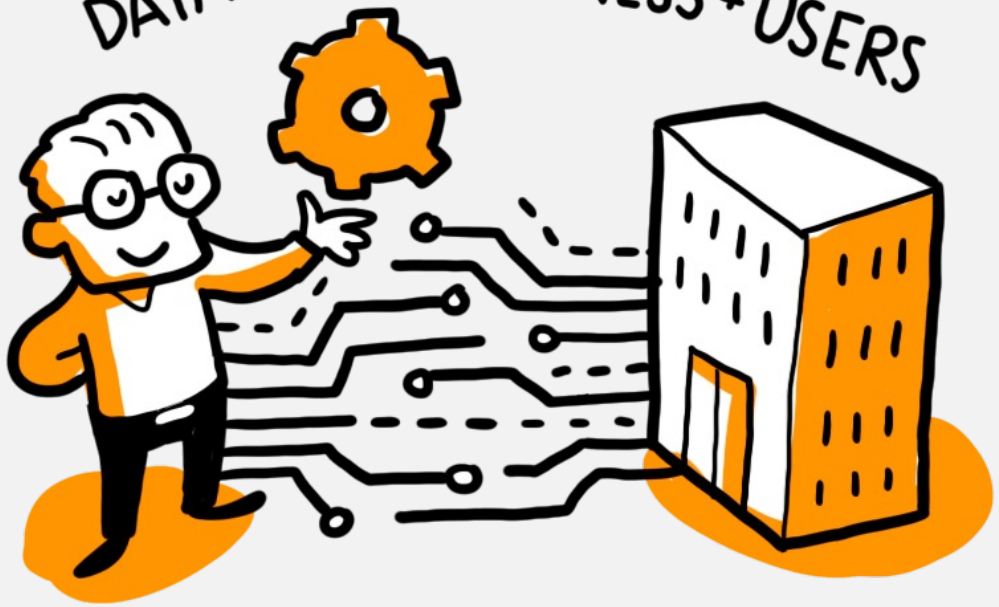
Disha Singh



Tamara Tvistholm



Thorkild Dam-Dalgeir



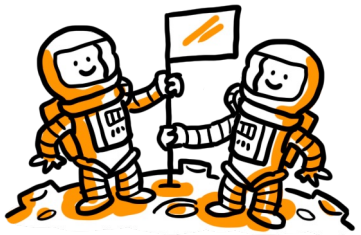
Digital Solutions Biometrics

Challenges and opportunities

- **Data Inconsistency**: Automated pipelines ensure that data is consistently processed in the same manner, reducing the risk of human error and inconsistency in data handling.
- **Time and Resource Intensive Manual Processes**: Automation can significantly reduce the time and resources required for data processing and review, leading to increased efficiency.
- **Real-time Data Processing**: Automated pipelines can process and analyze data in real-time, which is crucial for time-sensitive decisions in clinical settings.
- **Scalability**: As data volume grows, manual processes can become overwhelmed. Automated pipelines are designed to handle large volumes of data, ensuring scalability.
- **Data Quality**: Automated pipelines include data validation steps to ensure the quality and accuracy of the data, which is critical for reliable clinical and operational decisions.
- **Data Integration**: Automated pipelines can integrate data from various sources, formats, and standards, which is often a challenge in clinical settings due to the heterogeneity of clinical data.
- **Regulatory Compliance**: Automated pipelines can be designed to adhere to regulatory requirements, ensuring that data handling processes are compliant with relevant laws and regulations.

HANDLE TECHNICAL INFRASTRUCTURE AND DATA INTEGRATION

OPERATIONAL EXCELLENCE



TAKE OWNERSHIP

SOURCES

- medidata Rave
- Lundbeck EasyShare
- POSSUM
- CRO CTMS
- PLUTO

PROCESSING

- EntimICE
- sas
- R
- python
- ORACLE SQL
- PINNACLE 21 BY CERTARA
- box

STORAGE

- EntimICE
- GitHub
- ORACLE CDL
- ORACLE ACROSS
- ORACLE POSSUM
- box

REPORTING

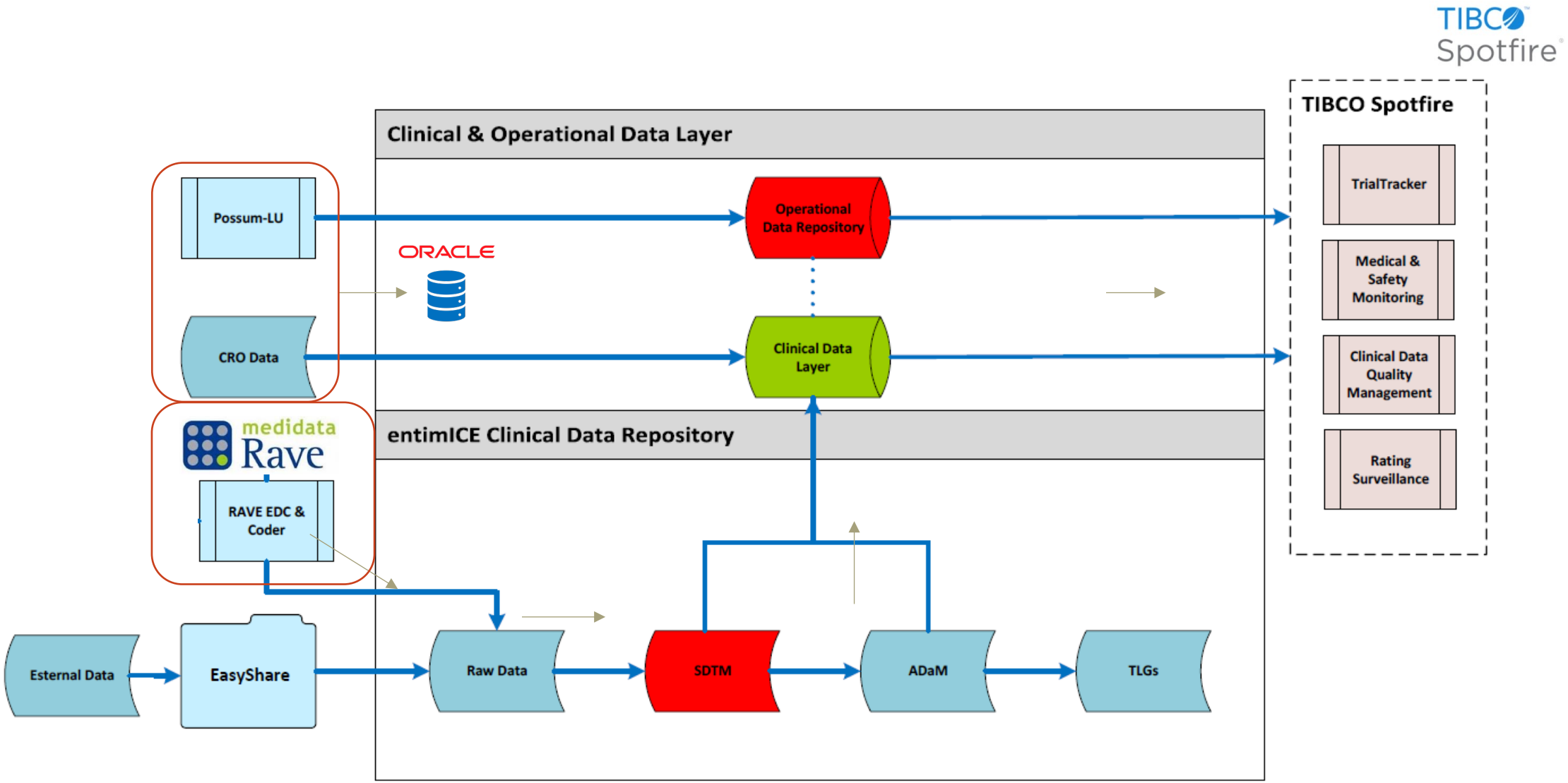
- TIBCO Spotfire JavaScript
- Pluto
- TrialScope DISCLOSE

DEVELOP TOOLS TO STANDARDIZE ANALYTICS

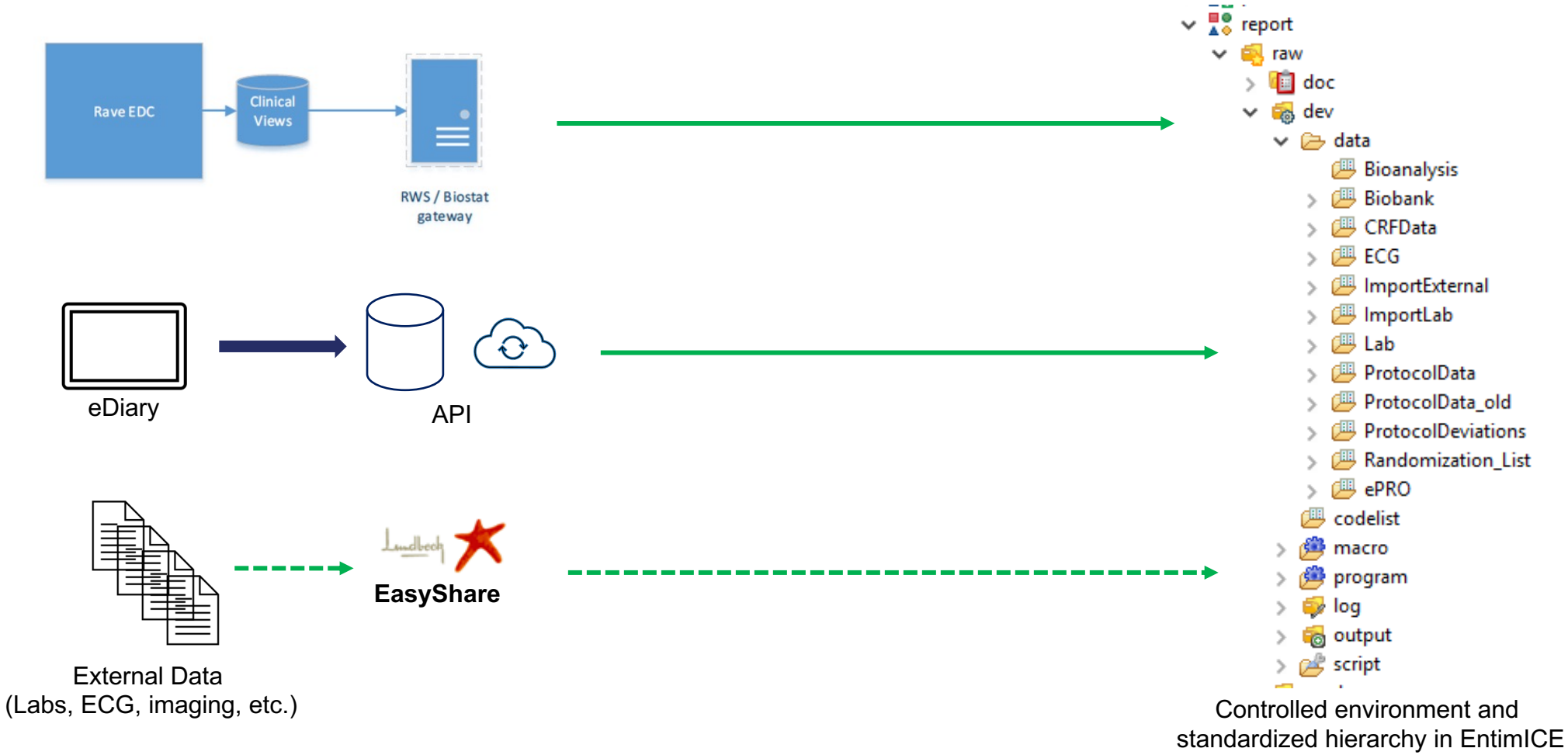
INCREASE GOVERNANCE AND REUSABILITY ACROSS PROJECTS

GxP COMPLIANCE

Process overview



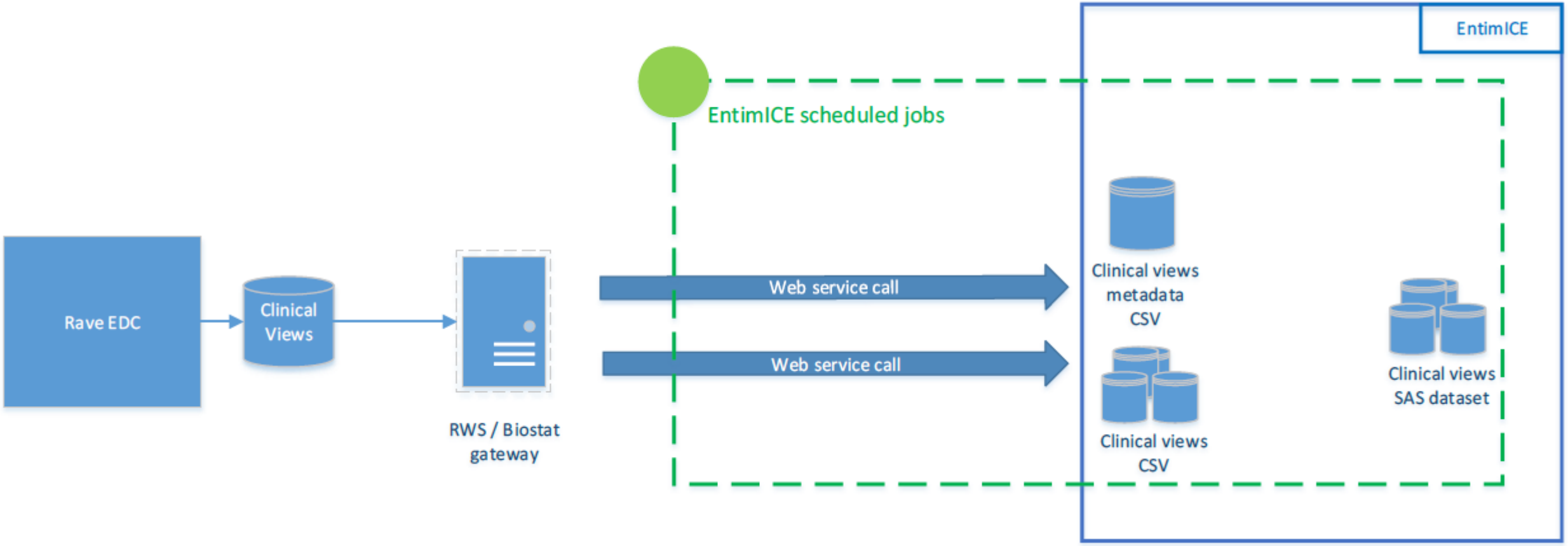
Getting the raw data into the Clinical Data Repository
EDC & External Data



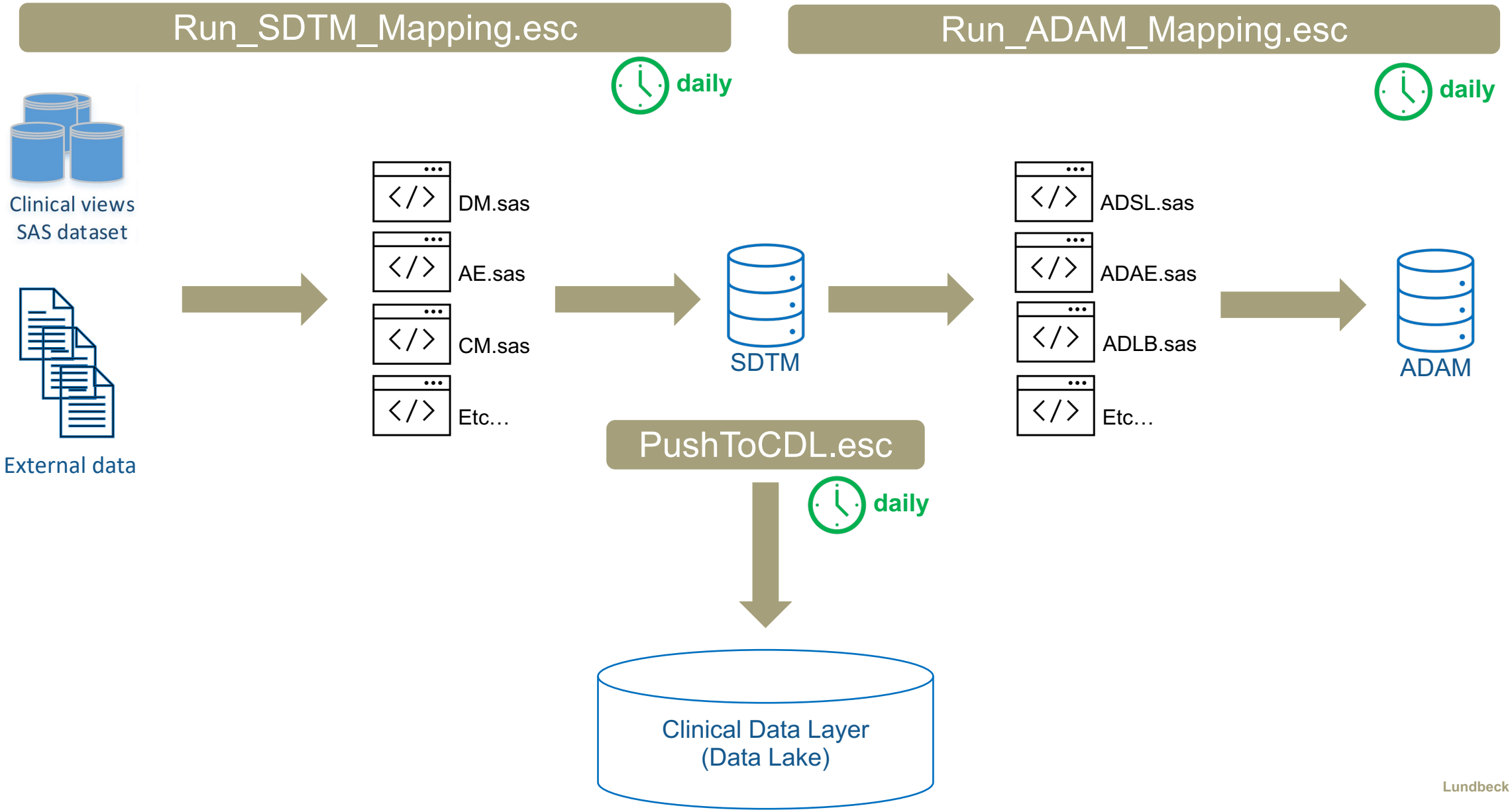
Automated interface

Manual interface

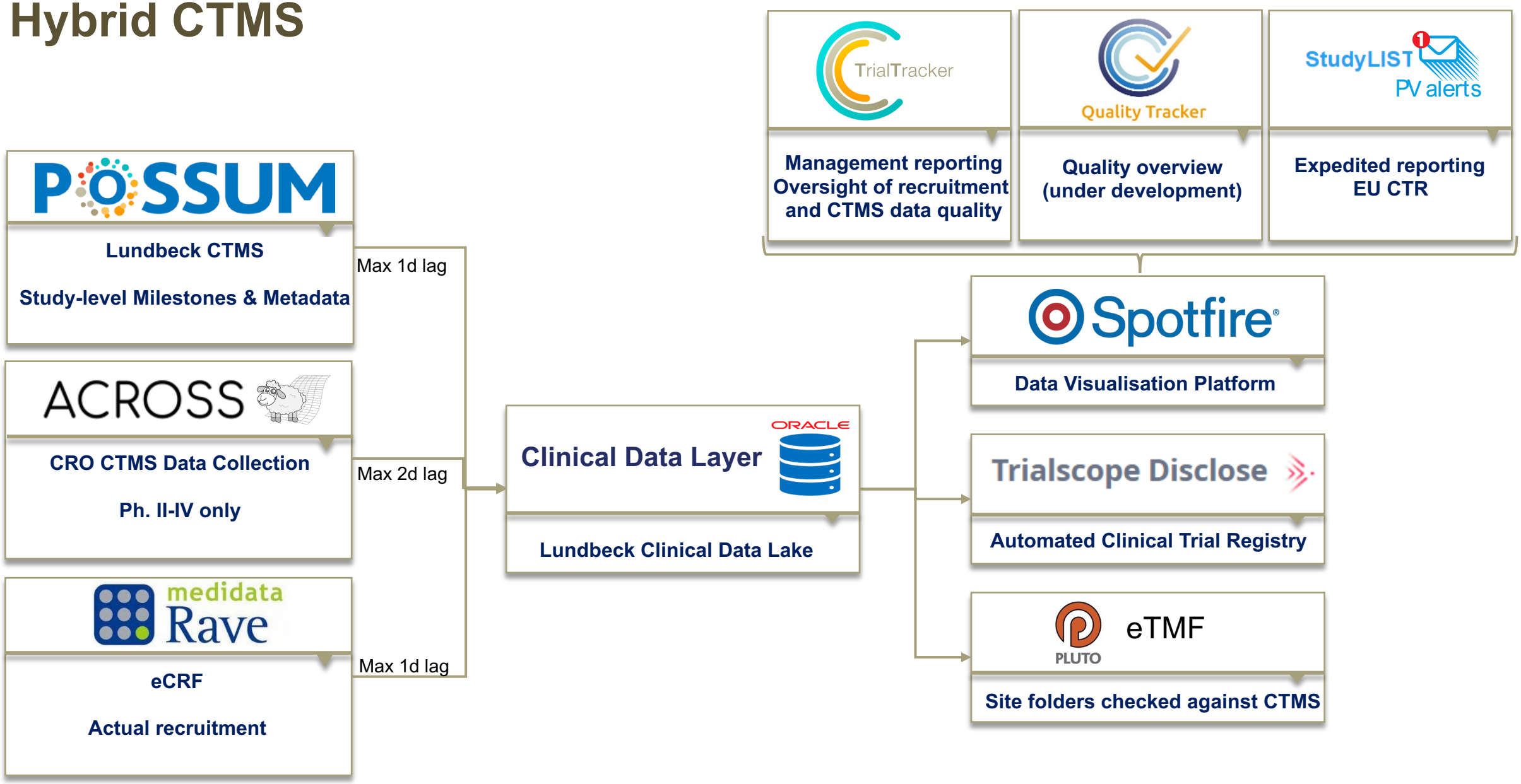
Rave EDC interface with EntimICE



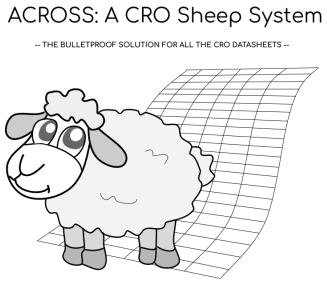
Data Processing (Raw – SDTM – ADAM)

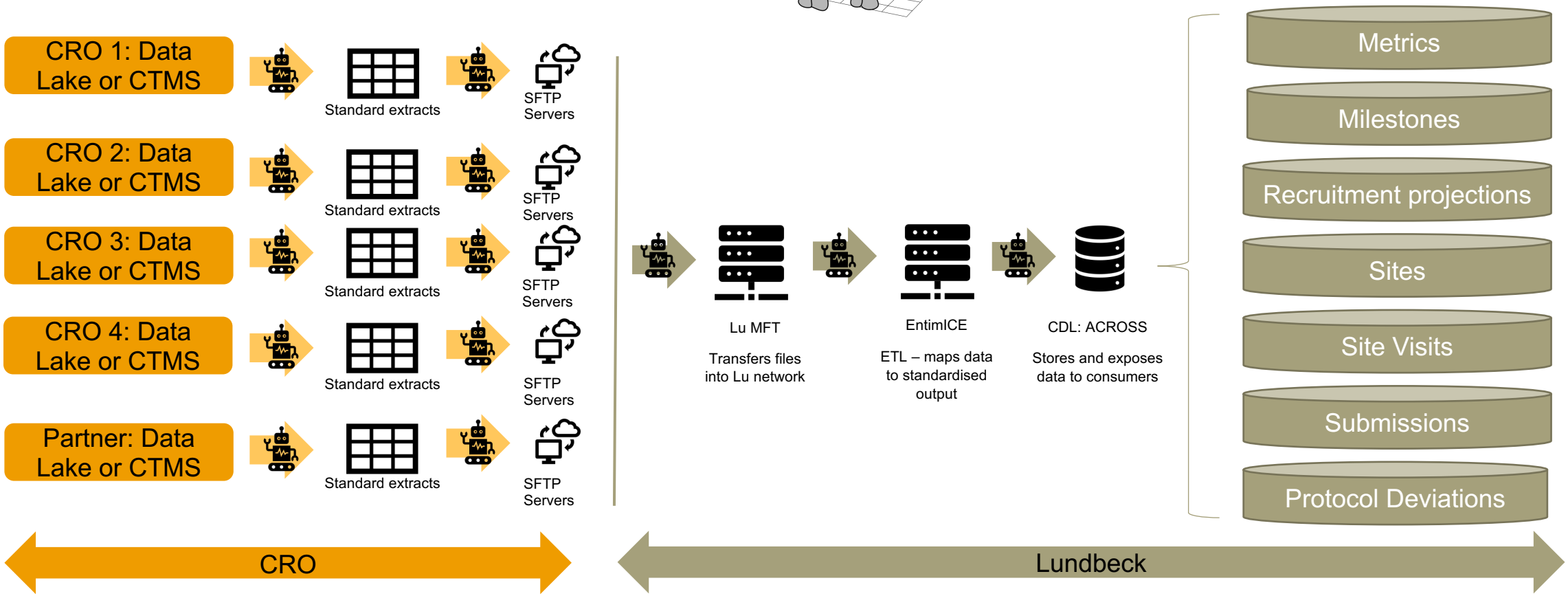


Hybrid CTMS

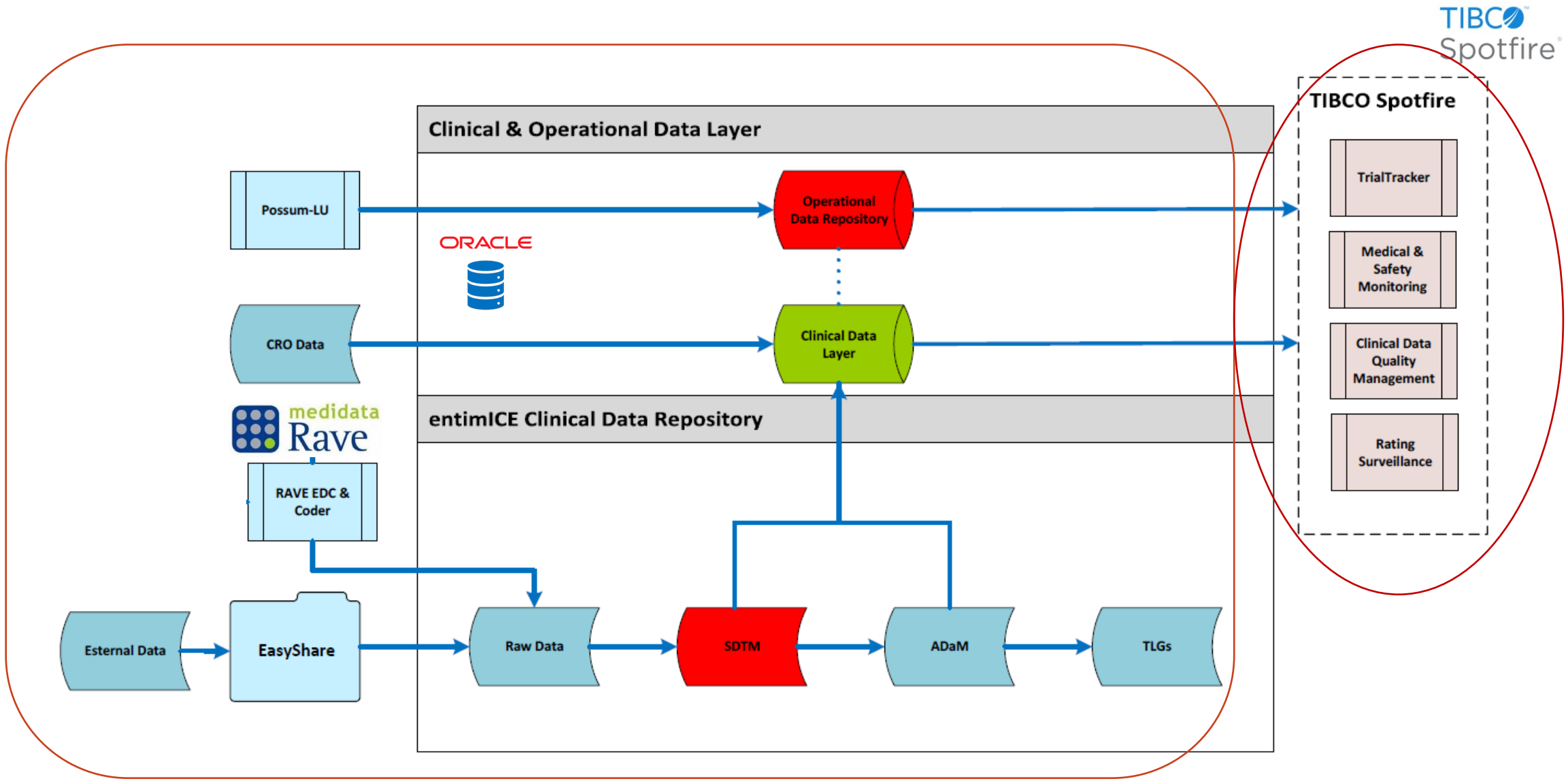

TrialTracker
Management reporting
Oversight of recruitment and CTMS data quality
Quality Tracker
Quality overview (under development)
StudyLIST
PV alerts
Expedited reporting
EU CTR
POSSUM
Lundbeck CTMS
Study-level Milestones & Metadata
Max 1d lag
ACROSS
CRO CTMS Data Collection
Ph. II-IV only
Max 2d lag
medidata Rave
eCRF
Actual recruitment
Max 1d lag
Clinical Data Layer
ORACLE
Lundbeck Clinical Data Lake
Spotfire
Data Visualisation Platform
Trialscope Disclose
Automated Clinical Trial Registry
PLUTO
eTMF
Site folders checked against CTMS

ACROSS (A CRO Sheep System)





Process overview



Medical Monitoring and Safety Review

Lundbeck Medical Monitoring and Safety Review

You may click the links below or the tabs at the top to browse this workbook in any order.

Population-Level Analysis

Overview

- Summary
- Safety Overview

Demographics and Disposition

- Screening Failures
- Demographics
- Disposition
- Withdrawals

Adverse Events

- Summary of AEs
- AEs by SOC and PT
- AEs by Duration
- AE/SAE Heatmap
- AEs leading to Withdrawal
- AEs of Special Interest

Medical History and CM

- Medical History
- Primary Indication
- CM by Name or Class
- CM for Primary Indication
- Disallowed Medication

Laboratory

- Lab Shift Plots - ULN
- Lab Shift Plots - LLN
- Hy's Law
- Lab Selector
- Labs out of Range
- Labs over time
- Lab CFB
- Lab PCS

ECGs

- ECG Interpretation
- ECGs over time
- ECG CFB
- ECG PCS

Vital Signs

- Vitals over time
- Vitals CFB
- Vitals PCS

Examinations

- Physical and Neurological Exam

Safety Scales

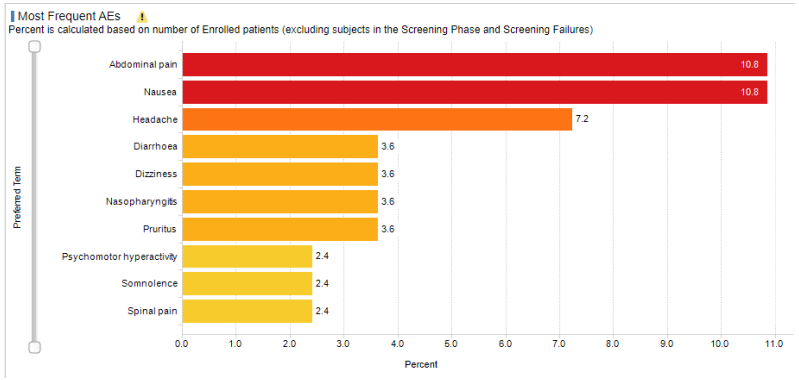
- AIMS
- BARS
- MSAS
- C-SSRS
- DESS

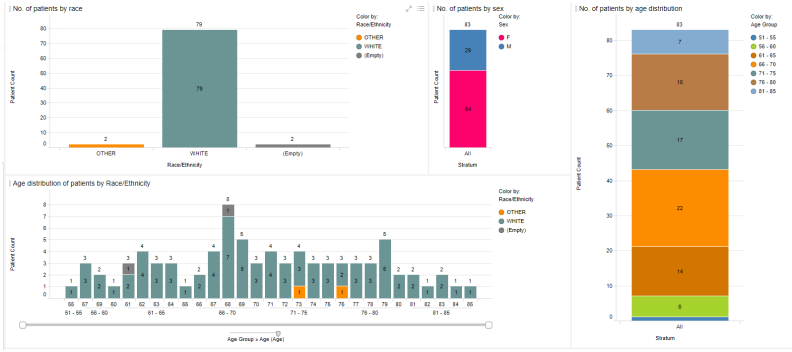
Pediatric

- PAERS
- PAERS over time
- TANNER
- GBIM
- GBIM over time
- Growth Chart

Patient-Level Analysis

- Patient Profile (Graph)
- Patient Profile (Listing)
- AEs, CM and Labs
- AEs, CM and Vitals
- AEs, CM and ECGs





Category for ...	Lab Test or E...	Lower Limit ...	Character Re...	Upper Limit ...	Reference Ra...	ULN	Age	Sex	Race	Country	Standard Units
LIPIDS	CHOL	3.24	6.24	5.18	HIGH NORMAL	1.20	32	M	ASIAN	CHN	mmol/L
LIPIDS	HDL	1.04	0.91	1.55	LOW NORMAL	0.59	32	M	ASIAN	CHN	mmol/L
LIPIDS	LDL	1.30	3.37	4.14	NORMAL	0.81	32	M	ASIAN	CHN	mmol/L
LIPIDS	TRIG	0.51	4.45	2.26	HIGH NORMAL	1.97	32	M	ASIAN	CHN	mmol/L
LIVER	ALP	43.00	94	115.00	NORMAL	0.82	32	M	ASIAN	CHN	U/L
LIVER	ALT	10.00	164	40.00	HIGH NORMAL	4.10	32	M	ASIAN	CHN	U/L
LIVER	AST	10.00	58	43.00	HIGH NORMAL	1.35	32	M	ASIAN	CHN	U/L
LIVER	BILDIR	0.00	5.1	6.80	NORMAL	0.75	32	M	ASIAN	CHN	umol/L
LIVER	BILI		12.3	18.80	NORMAL	0.65	32	M	ASIAN	CHN	umol/L
LIVER	GGT	10.00	70	49.00	HIGH NORMAL	1.43	32	M	ASIAN	CHN	U/L

Clinical Data Quality Metrics

Safety

- KRIs :: Adverse Events
- KRIs :: CM and MH
- KRIs :: PCS Values per AE
- KRIs :: PCS Values

Study Conduct

- KRIs :: Queries
- KRIs :: Query Response Time
- KRIs :: eCRF Data Entry Delay
- KRIs :: Missing Scales
- KRIs :: Missing Lab, ECG, VS, Exams
- KRIs :: Vital Signs Duplicates
- KRIs :: BP and Pulse Duplicates
- KRIs :: Vitals Rounded
- KRIs :: Unscheduled Visits

Recruitment

- KRIs :: Inactive Sites
- KRIs :: Randomization
- KRIs :: Screen Failures and Withdrawals

Efficacy

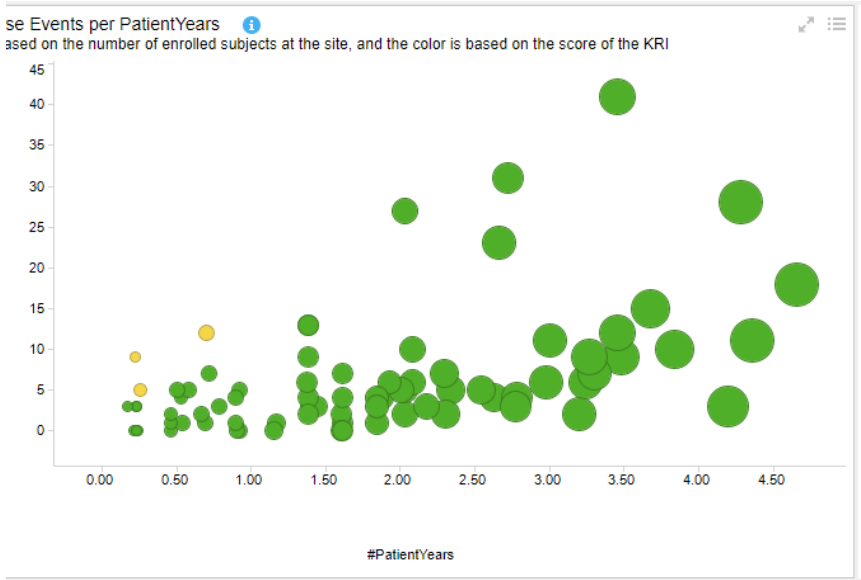
- KRIs :: Raters
- KRIs :: Primary Endpoint

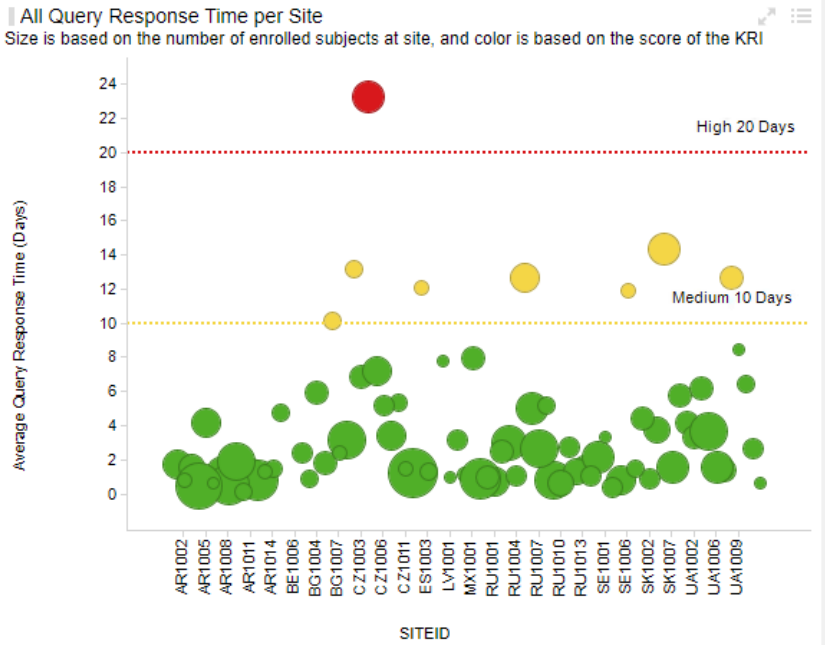
Summary Pages

- RBCM Report
- KRI Summary
- KRI Summary QA
- Metrics Summary
- Subject Summary

Other Pages

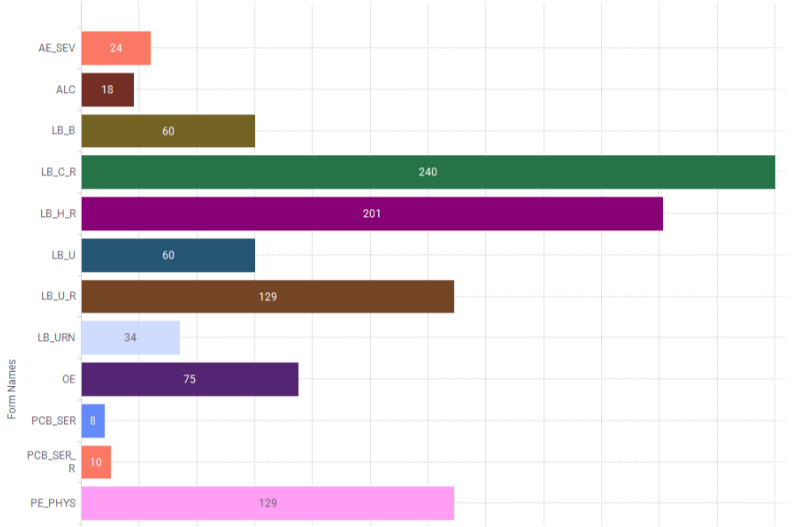
- Scales per Rater
- Raters at Multiple Sites
- Week Days
- ECG Interpretation

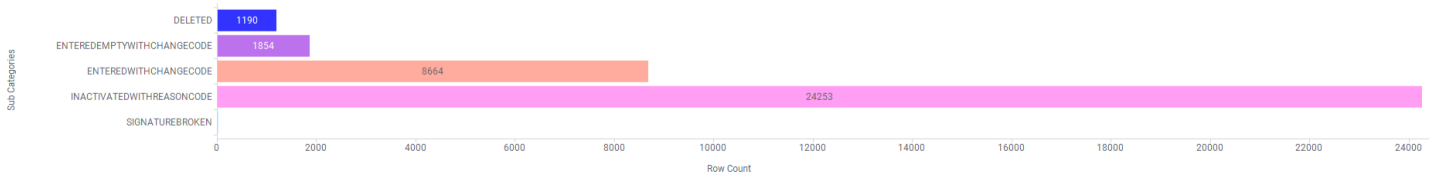


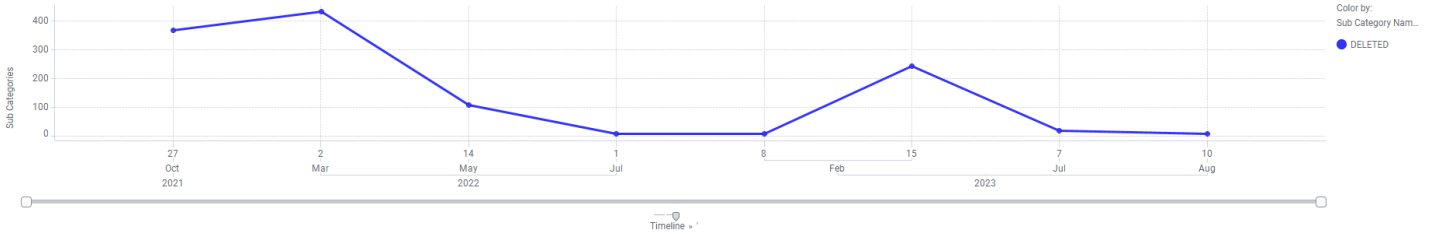


RAVE- Audit Trail Review

Which forms has been deleted or inactivated

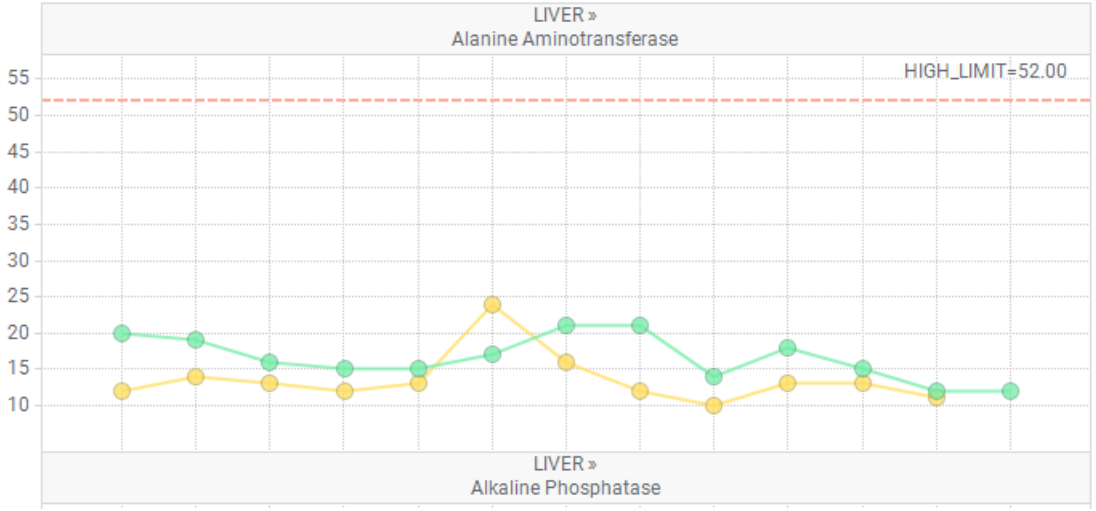


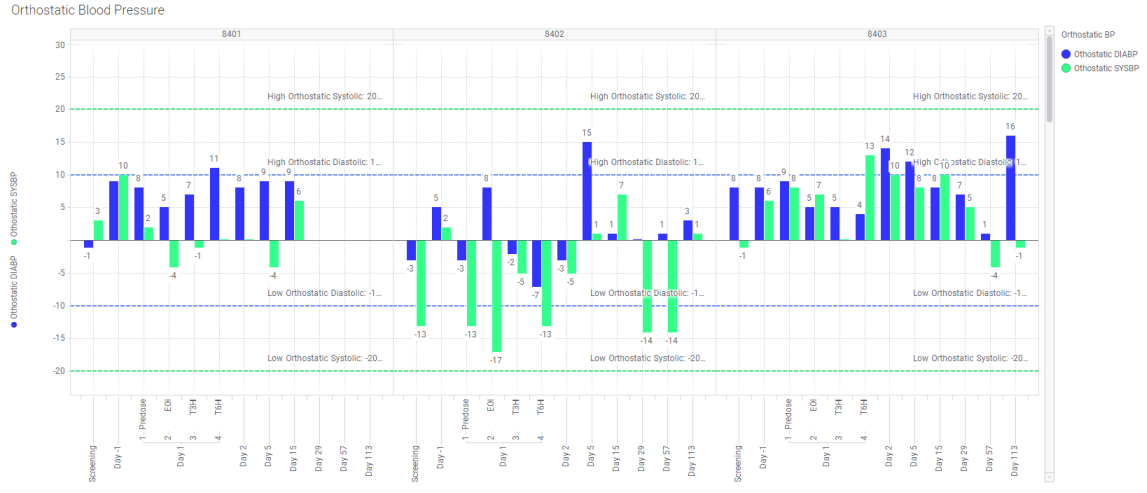




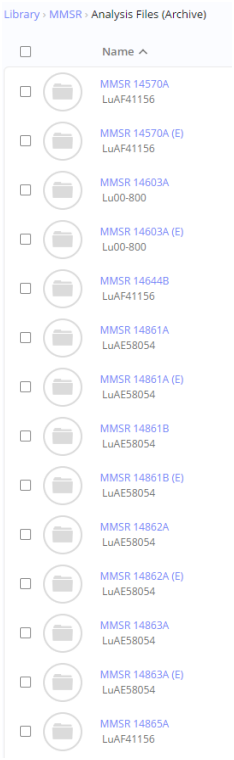
Master Templates

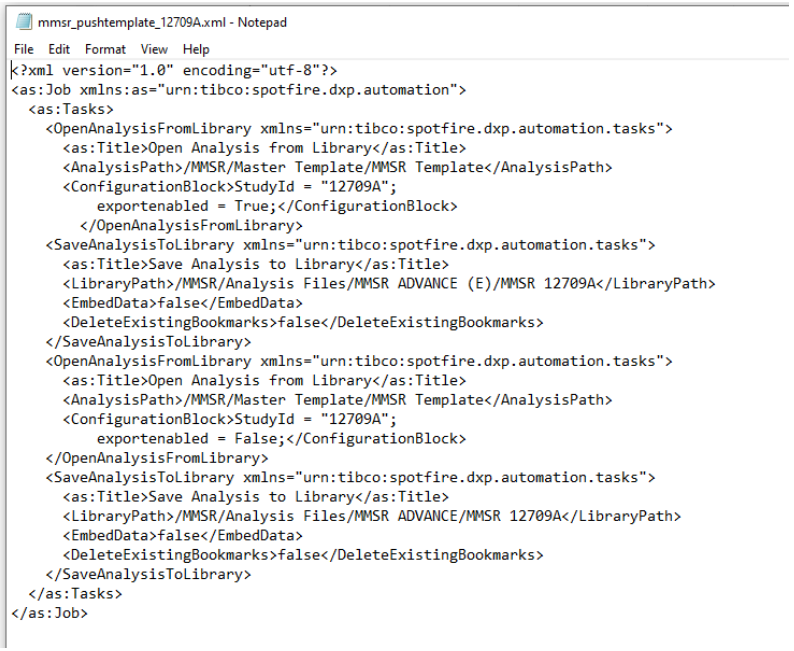
- Spotfire Applications
  - Medical Monitoring
  - Phase 1 Dosing Conferences
  - Risk Based Quality Management
  - Rave Audit Trail

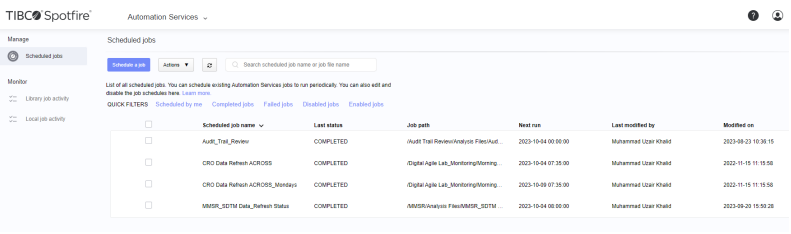




Automation Jobs to create study based applications







Monitoring Data Refresh in Spotfire

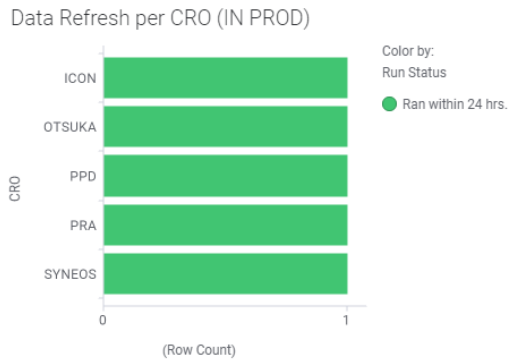
Priority	Name	Type	Details	Schedule	Resource pool	Last status
0	StudyLIST	Analysis file	/StudyLIST/Analysis Files/StudyLI...	Update with new data every 480 ...	Default routing	Successfully loaded o
1	TrialTracker	Analysis file	/Digital Agile Lab/Master Templat...	Update with new data every 525 ...	Default routing	Successfully loaded o
2	TrialTracker_mobile	Analysis file	/Digital Agile Lab/Master Templat...	Update with new data every 510 ...	Default routing	Successfully loaded o
3	QualityTracker	Analysis file	/QualityTracker/Production/Qualit...	Update with new data every 510 ...	Default routing	Successfully loaded o
4	Safety Review Tool	Analysis file	/Safety Review Tool/Analysis File/...	Update with new data every 510 ...	Default routing	Successfully loaded o
5	Biobank	Analysis file	/Biobank/Analysis Files/Biobank	Update with new data every 870 ...	Default routing	Successfully loaded o
7	CRO Data Refresh	Analysis file	/Digital Agile Lab_Monitoring/Mor...	Update with new data every 390 ...	Default routing	Successfully loaded o
8	RAVE_AUDIT_19678A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
9	RAVE_AUDIT_18898A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
10	RAVE_AUDIT_19366A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
11	RAVE_AUDIT_18331A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
14	RAVE_AUDIT_19356A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
15	RAVE_AUDIT_19385A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
17	RAVE_AUDIT_19379A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
18	RAVE_AUDIT_19702A	Analysis file	/Audit Trail Review/Analysis Files/...	Update with new data every 600 ...	Default routing	Successfully loaded o
21	Trial Resource Planning	Analysis file	/Trial Resource Planning/Master T...	Update with new data every 525 ...	Default routing	Successfully loaded o
22	CDQM 18898A	Analysis file	/CDQM/Analysis Files (Archive)/C...	Update with new data every 600 ...	Default routing	Successfully loaded o

Monitoring Data Refresh per CRO

This email contains:

- Data refresh in PROD
- ACROSS load error
- RA Email Status

Data Refresh per CRO (IN PROD):



T_LOADTIME:

T_LOADTIME

CRO	LOAD_TS	Ru
ICON	03-10-2023 04:57:58	Ra
OTSUKA	03-10-2023 06:19:35	Ra
PRA	03-10-2023 03:59:38	Ra
PPD	03-10-2023 05:45:16	Ra
SYNEOS	03-10-2023 03:47:16	Ra

T_LOAD_ERROR:

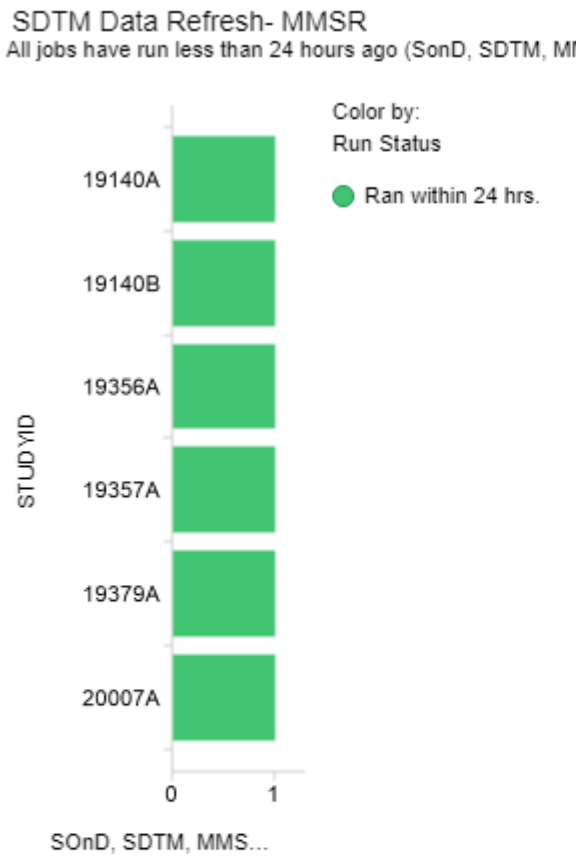
T_LOAD_ERROR

ID	JOB_ID	ALERT_TYPE	CRO	INVALID_ROW	ORIGIN	LOAD_COMM...	DATE_CREAT...	CREATED_BY	DAT
195358,00	0,00	No new Data L...	ICON				02-09-2023 04...	ACROSS_SAC	
195356,00	0,00	No new Data L...	ICON				11-08-2023 04...	ACROSS_SAC	

Data table:
T_LOAD_ERROR

Monitoring SDTM Data Refresh per study

SDTM Data Refresh- MMSR:



4 columns from Loadstatus Information link (allstudies):

4 columns from Loadstatus Information link (allstudies)

STUDYID	MMSR_FrontPageDate	MMSR_Data_RunStatus	SUCCESSFL
19140B	03-10-2023 02:38:56	Ran within 24 hrs.	1
19357A	03-10-2023 04:01:36	Ran within 24 hrs.	1
19140A	03-10-2023 00:55:55	Ran within 24 hrs.	1
19356A	03-10-2023 02:13:47	Ran within 24 hrs.	1
20007A	03-10-2023 02:02:29	Ran within 24 hrs.	1
19379A	03-10-2023 03:01:51	Ran within 24 hrs.	1

Conclusion

- Streamlined data flow reduces manual efforts
- Improved data quality
- Automated data management enhances clinical trial efficiency
- Improved data quality and timely analysis
- Enhances data integrity and regulatory compliance
- Facilitates efficient and transparent trial monitoring and auditing

Questions