

!!Say No to Stale Data!!

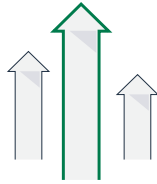
Implementation of automated GxP data pipelines using KNIME

Agenda

- Introduction to Sumitomo Pharma America
- Explore Traditional Data Flow
- Proposal of New Data Flow at SMPA
- Use Cases of KNIME at SMPA
- Components Built by SMPA
- GxP Validation Process

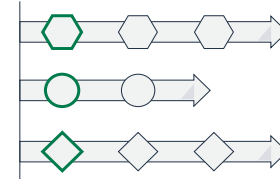
About Sumitomo Pharma America, Inc. (SMPA)

We are curious, bold, and share a passion for thinking *differently*.



Our Focus Areas

We accelerate scientific and therapeutic breakthroughs addressing areas of critical need.



High Impact Portfolio

We aim to address a wide range of conditions where treatment options remain limited or nonexistent.

Building on a Deep Sumitomo Pharma Heritage

Our parent company, Sumitomo Pharma Co., Ltd., was established in 1897 and has an extensive history of supporting health and wellbeing.

Our Areas of Focus

Women's Health



Urology



Oncology



Neurology



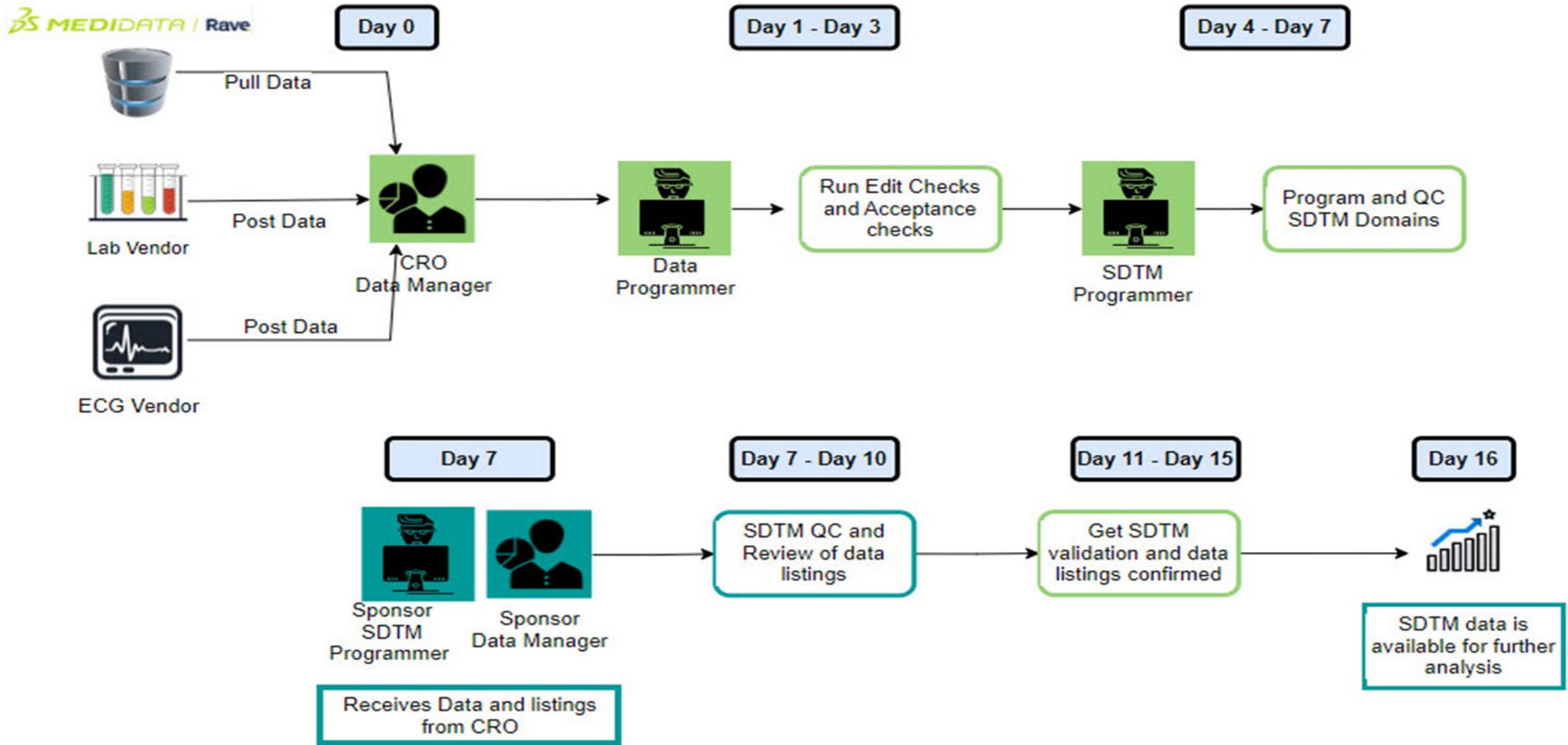
Psychiatry



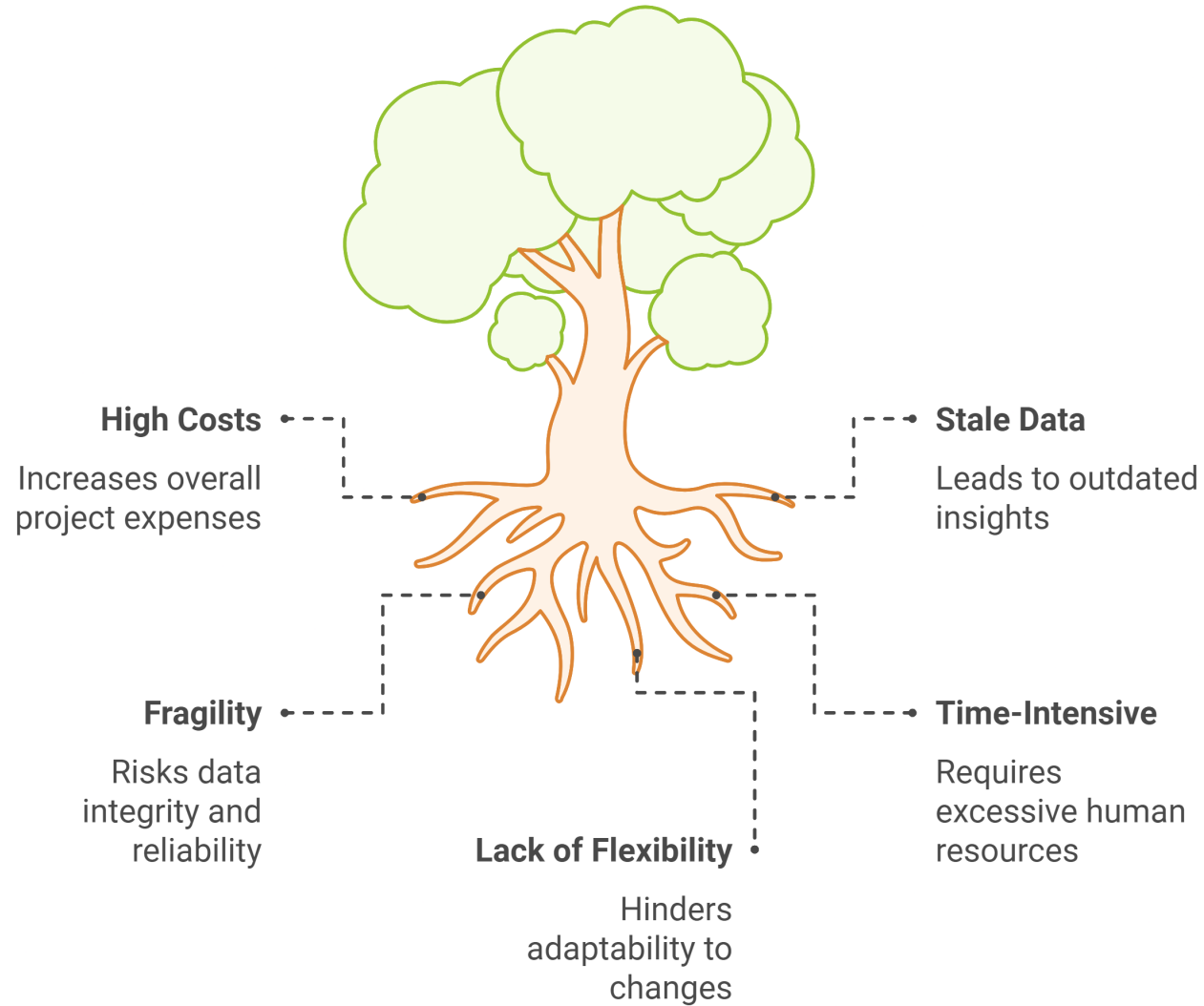
Rare Disease



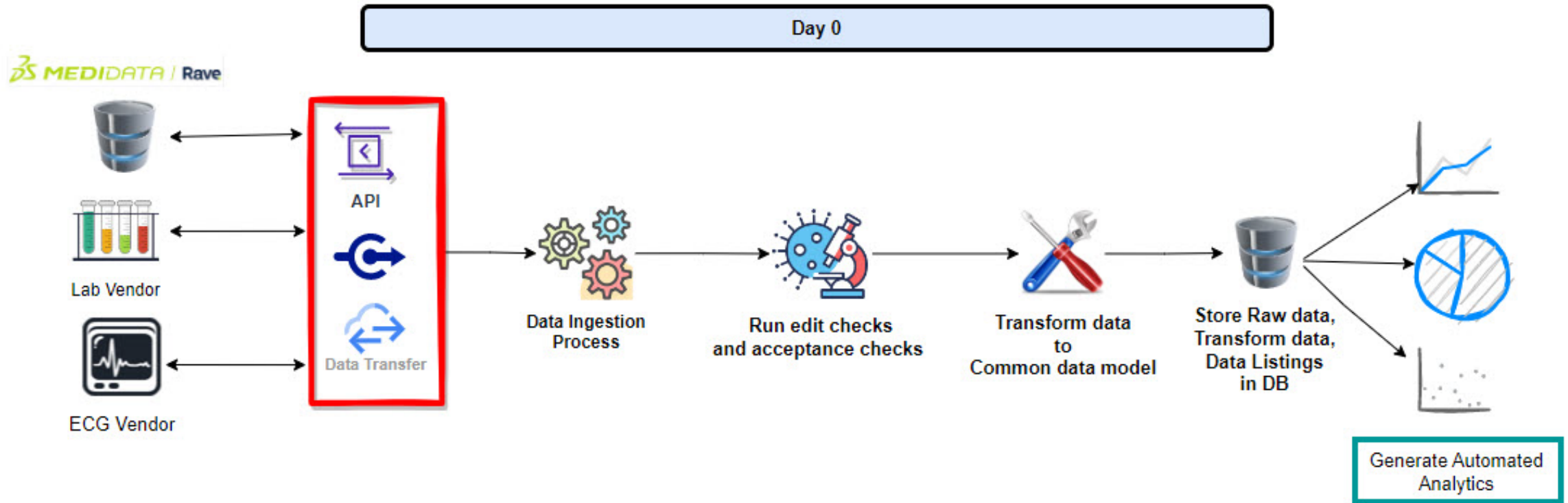
Traditional Data Flow



Challenges in Traditional Data Flow

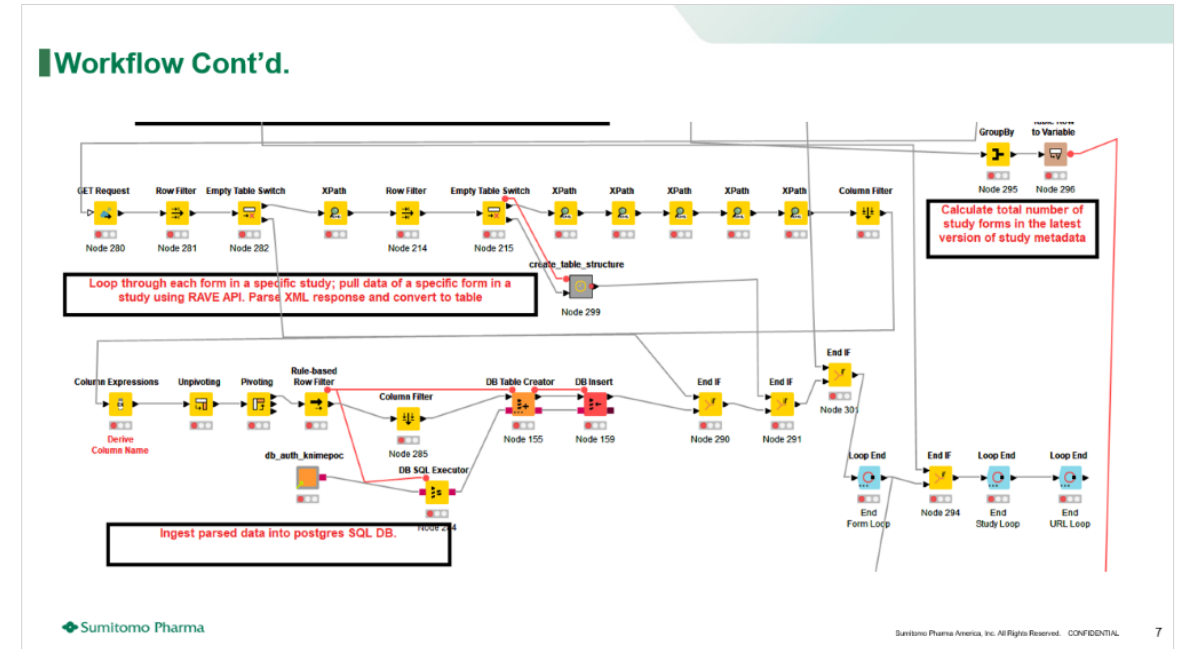
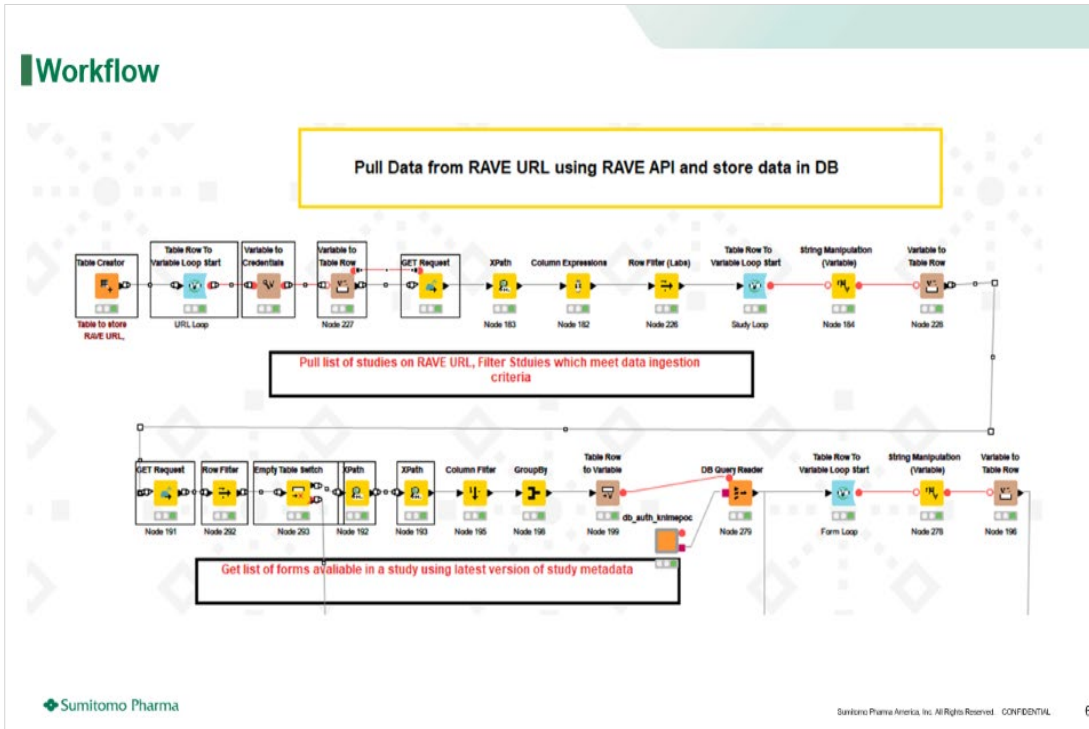


New Data Flow at Sumitomo Pharma America for Raw Data



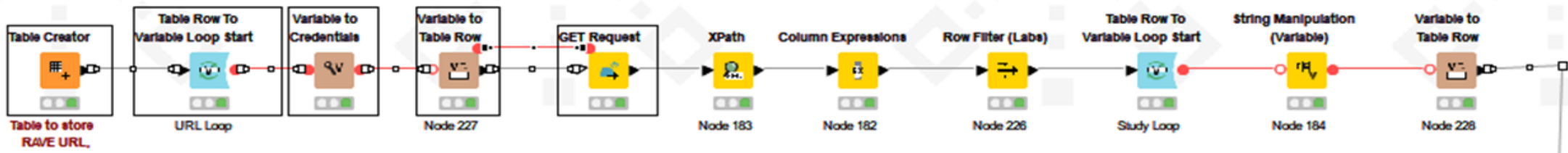
Use Case 1 – Ingest Data from RAVE

- Built and scheduled a KNIME workflow to pull data from RAVE URLs for the studies published on them on a daily frequency.
- RAVE web services provide data in XML format, workflow will transform data into tabular structure and store in SMPA data lake.

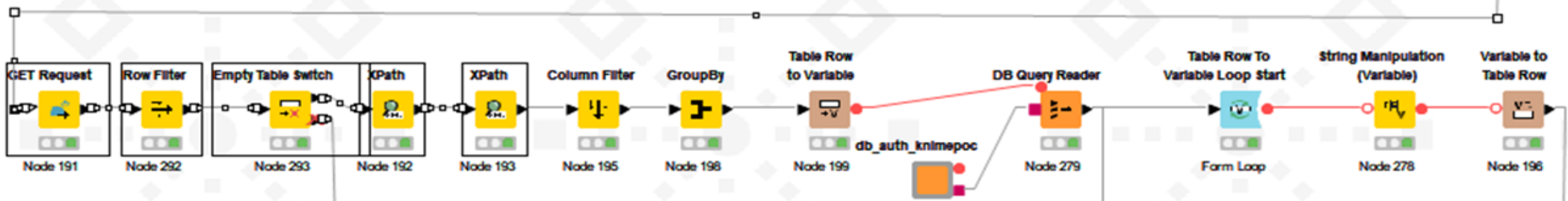


Workflow

Pull Data from RAVE URL using RAVE API and store data in DB

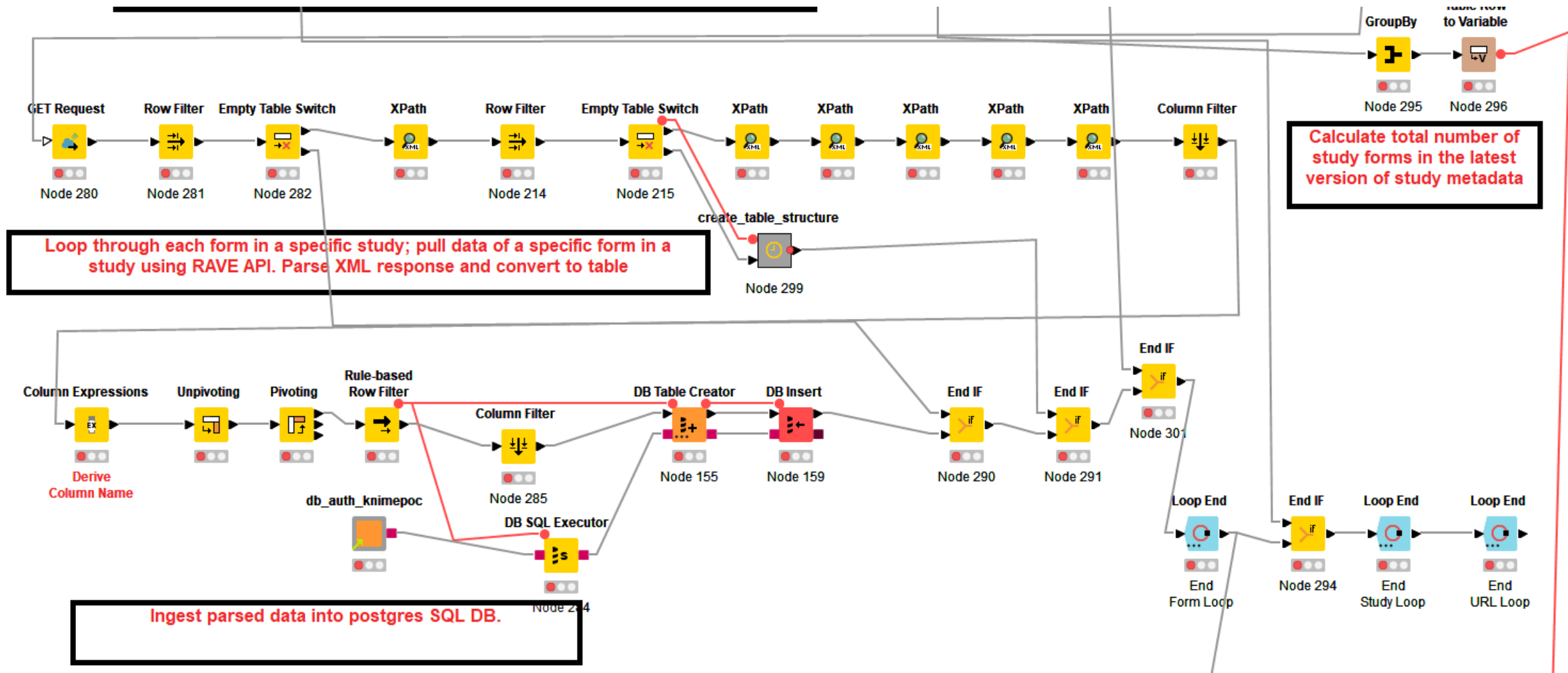


Pull list of studies on RAVE URL, Filter Studies which meet data ingestion criteria



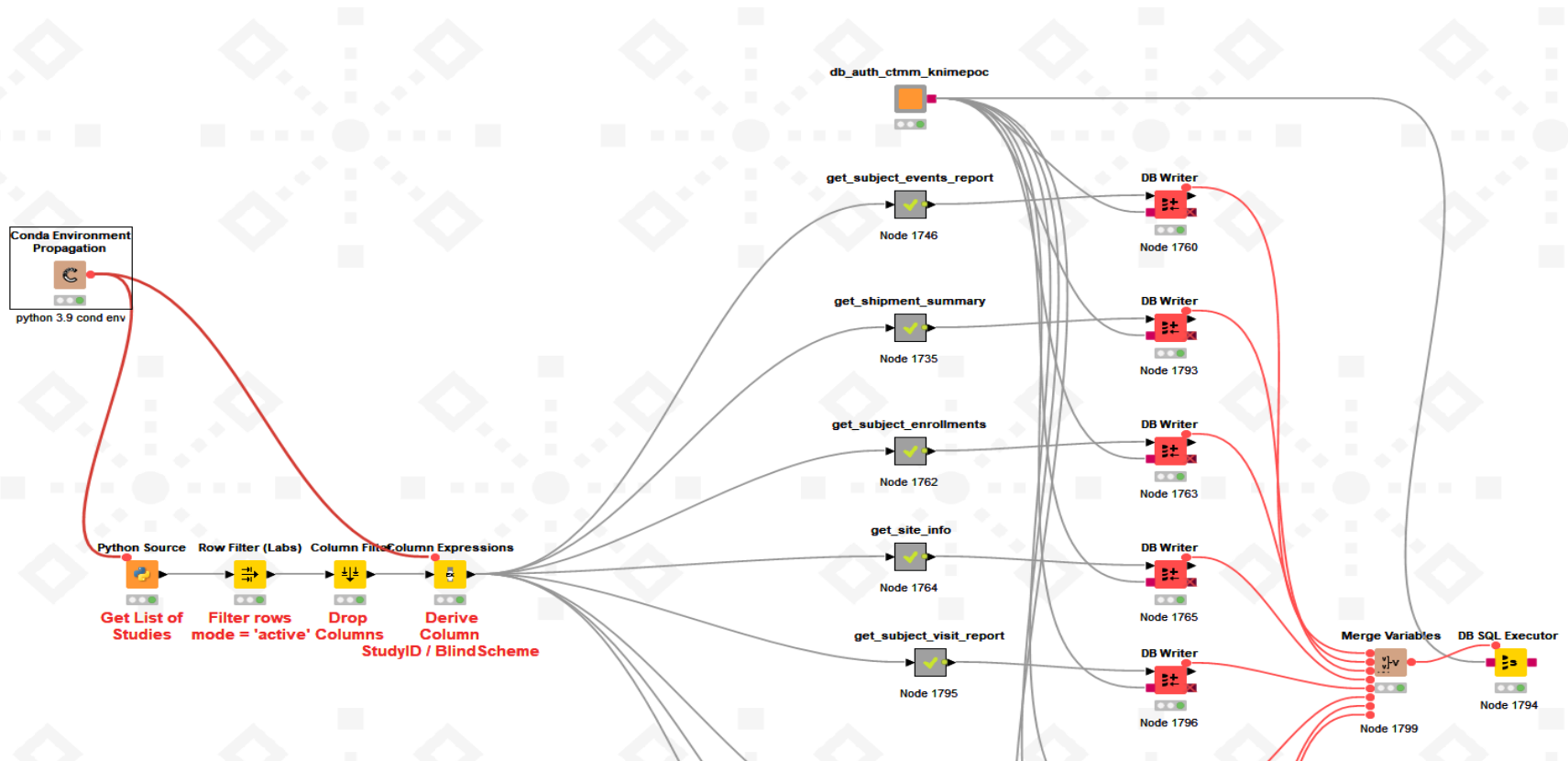
Get list of forms available in a study using latest version of study metadata

Workflow Cont'd.

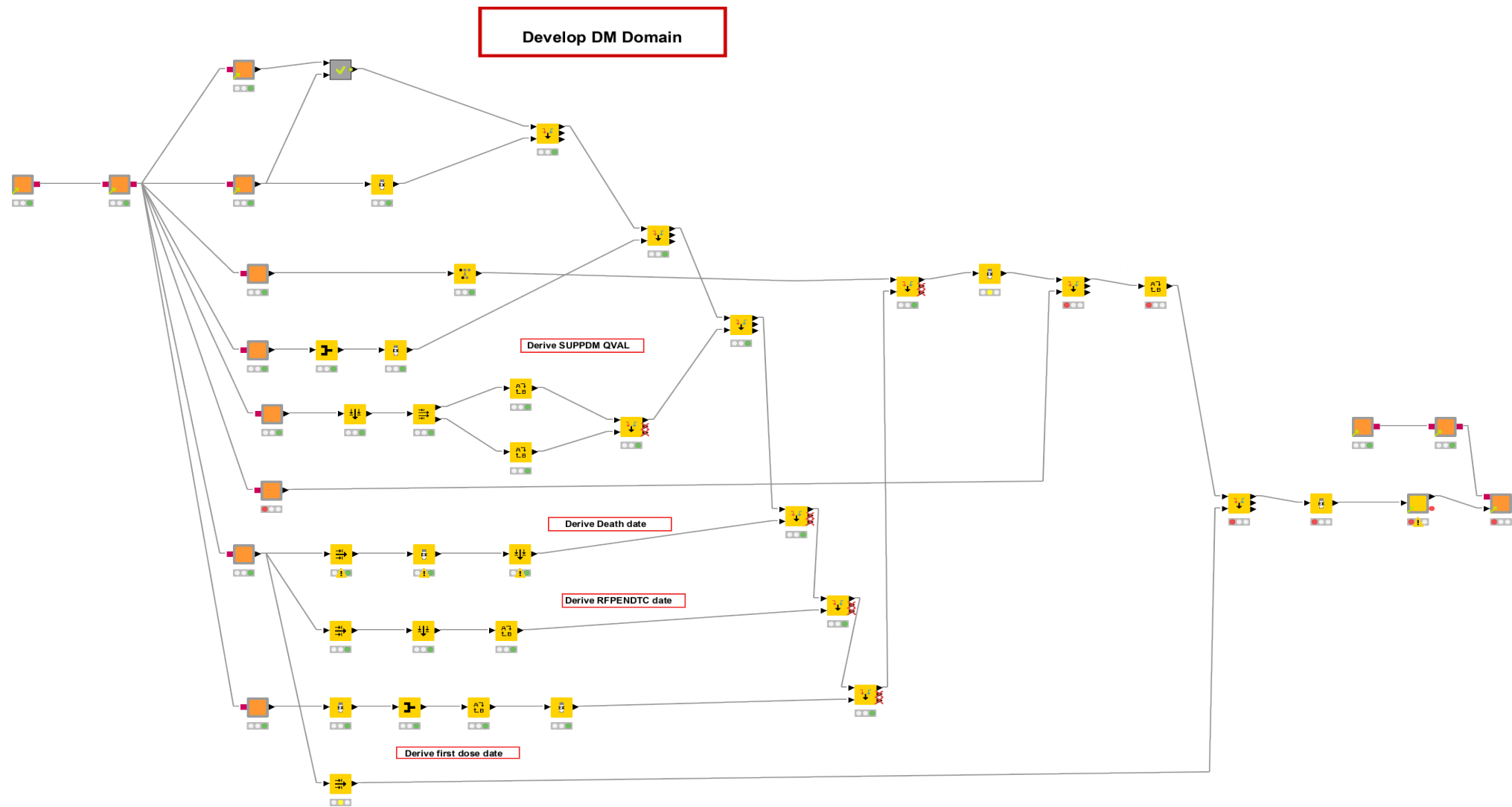


Use Case 2 – Ingest Data Oracle Clinical One

- Built KNIME workflow to ingest data from “**Oracle Clinical One**” for the SMPA studies.
- Data from about 15+ different source tables are being ingested daily.
- Business rules and Data Quality checks are designed in the WF to support different business needs.

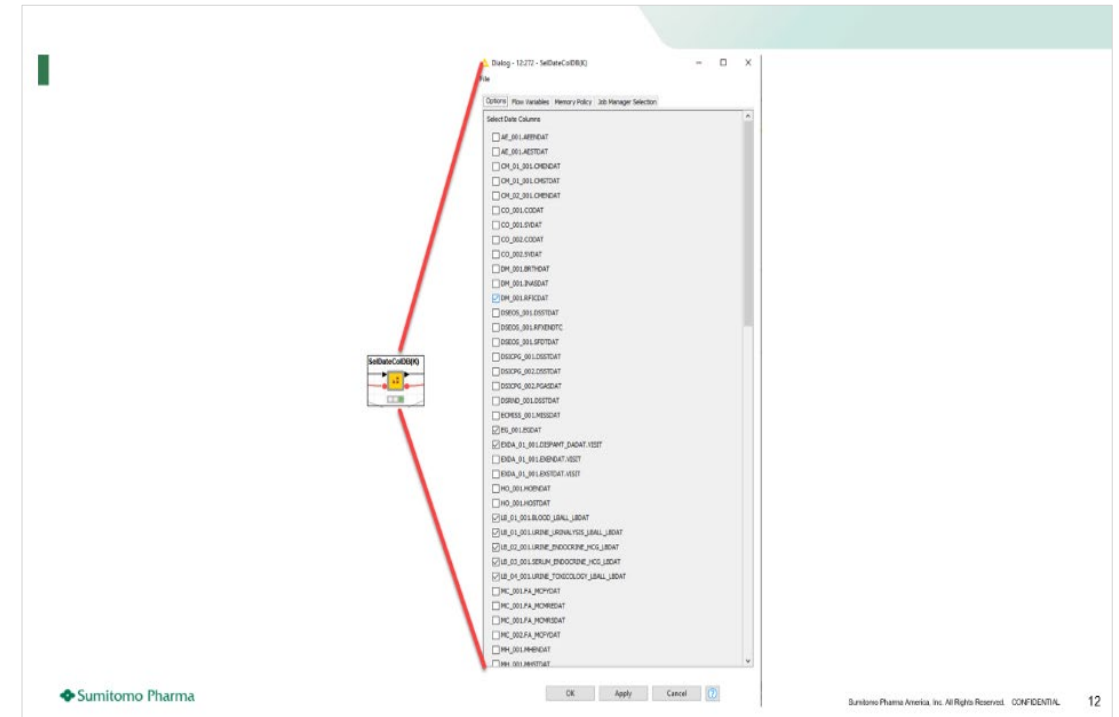
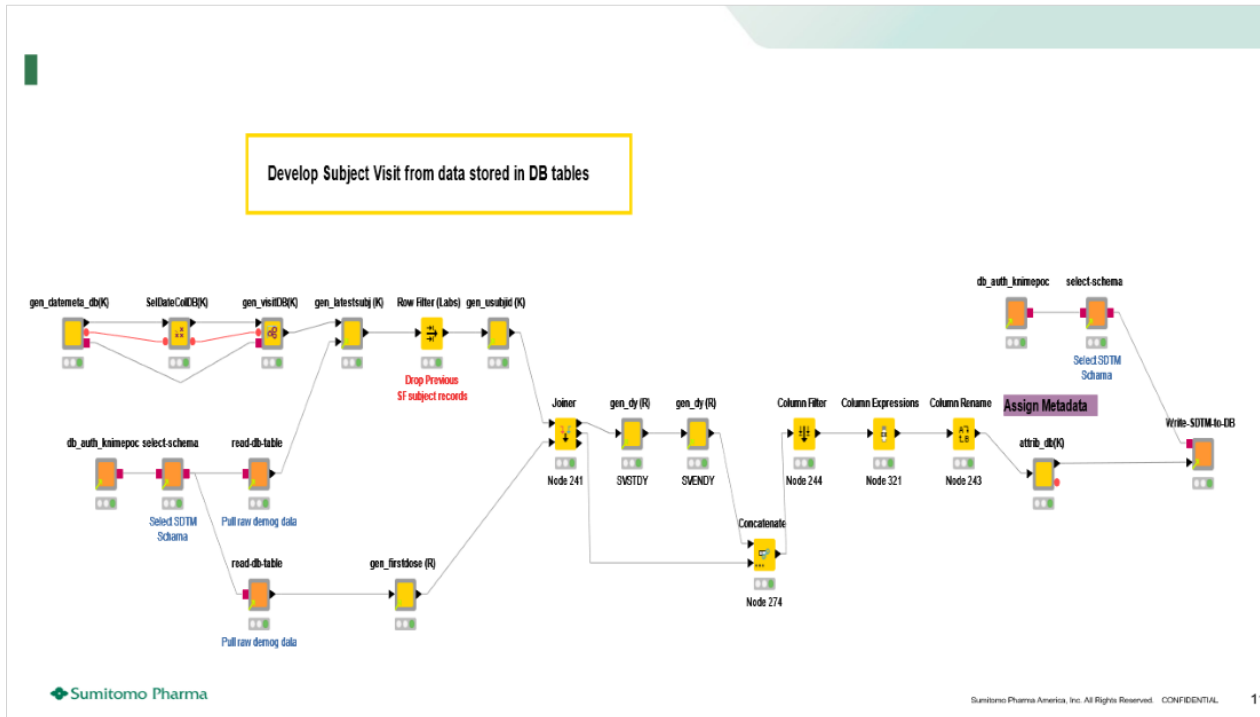


Use Case 3 – Develop SDTM Domain(s)

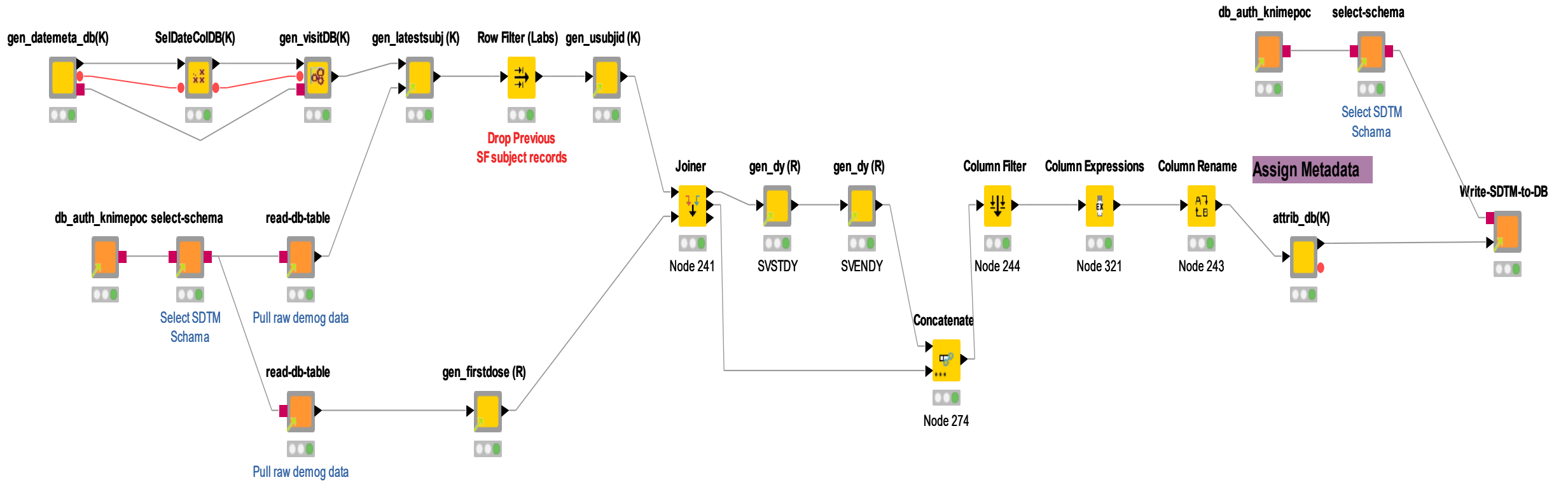


Use Case 3 – Develop Subject Visits (SV) Domain

- Developed SV domain workflow with user configuration that can be used across different studies



Develop Subject Visit from data stored in DB tables



Dialog - 12:272 - SelDateColDB(K)

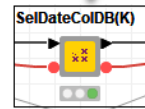
File

Options | Flow Variables | Memory Policy | Job Manager Selection

Select Date Columns

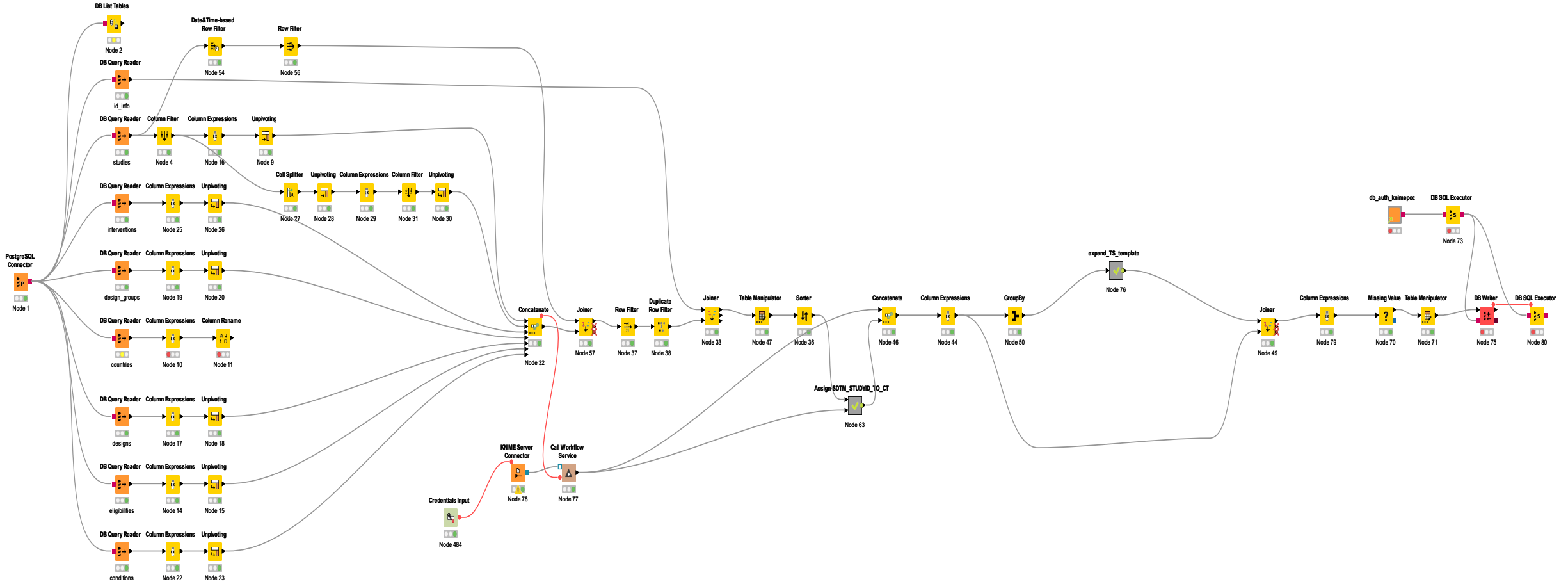
- AE_001.AEENDAT
- AE_001.AESTDAT
- CM_01_001.CMENDAT
- CM_01_001.CMSTDAT
- CM_02_001.CMENDAT
- CO_001.CODAT
- CO_001.SVDAT
- CO_002.CODAT
- CO_002.SVDAT
- DM_001.BRTHDAT
- DM_001.INASDAT
- DM_001.RFICDAT
- DSEOS_001.DSSTDAT
- DSEOS_001.RFXENDTC
- DSEOS_001.SFDTDAT
- DSICPG_001.DSSTDAT
- DSICPG_002.DSSTDAT
- DSICPG_002.PGASDAT
- DSRND_001.DSSTDAT
- ECMISS_001.MISSDAT
- EG_001.EGDAT
- EXDA_01_001.DISPAMT_DADAT.VISIT
- EXDA_01_001.EXENDAT.VISIT
- EXDA_01_001.EXSTDAT.VISIT
- HO_001.HOENDAT
- HO_001.HOSTDAT
- LB_01_001.BLOOD_LBALL_LBDAT
- LB_01_001.URINE_URINALYSIS_LBALL_LBDAT
- LB_02_001.URINE_ENDOCRINE_HCG_LBDAT
- LB_03_001.SERUM_ENDOCRINE_HCG_LBDAT
- LB_04_001.URINE_TOXICOLOGY_LBALL_LBDAT
- MC_001.FA_MCFYDAT
- MC_001.FA_MCMREDAT
- MC_001.FA_MCMRSDAT
- MC_002.FA_MCFYDAT
- MH_001.MHENDAT
- MH_001.MHSTDAT

OK Apply Cancel ?



Use Case 4 - Develop Automated Trial Design (TS) Domain

- Developed draft TS Domain from data sourced from AACT (ClinicalTrials.GOV) and SDTM datasets



Sample List of Components Built at Sumitomo Pharma America



Generate **USUBJID**



Generate **Study Day**

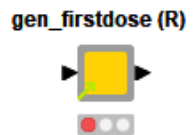
Components may have **configurations** to be managed by user . Configurations will add **flexibility to component**



Populate **LBTEST** and **LBTESTCD** from **LOINC** leveraging **CDISC Library API**



Generate **Latest Subject Number** when multiple screenings are done in the study



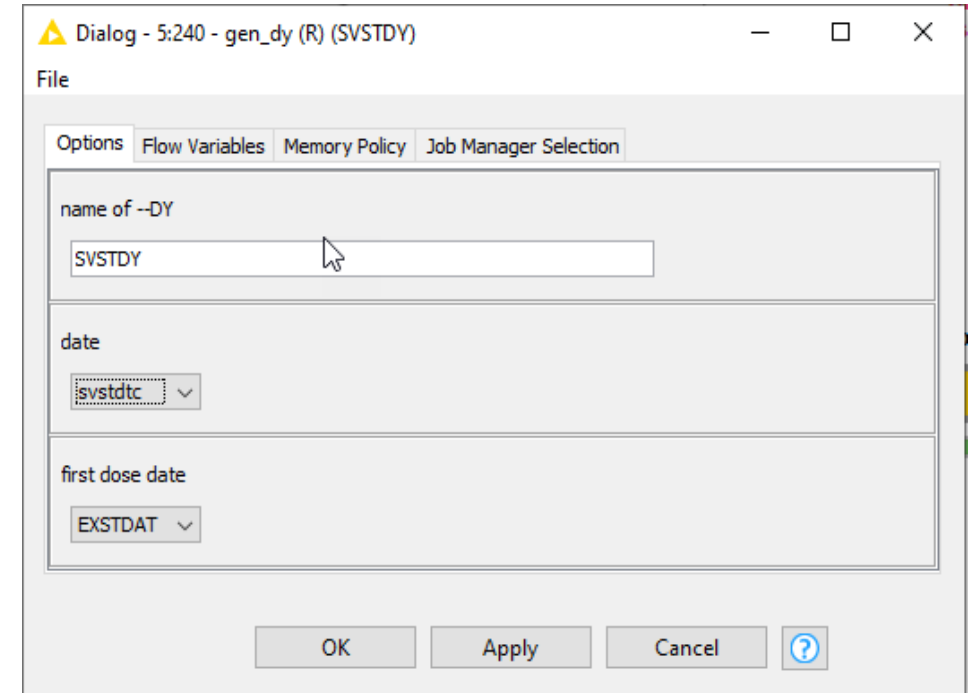
Generate First Dose of the Subject **RFSTDTC** and **RFXSTDTC**



Generate **EPOCH** based on dates and SE domain configuration



Assign **SDTM metadata** to the final data table



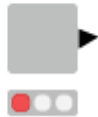
Sample List of Components Built at Sumitomo Pharma America

ucum based unit conversion



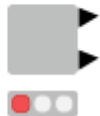
Convert results in one unit to another unit

search controlled terminology in CDISC Library

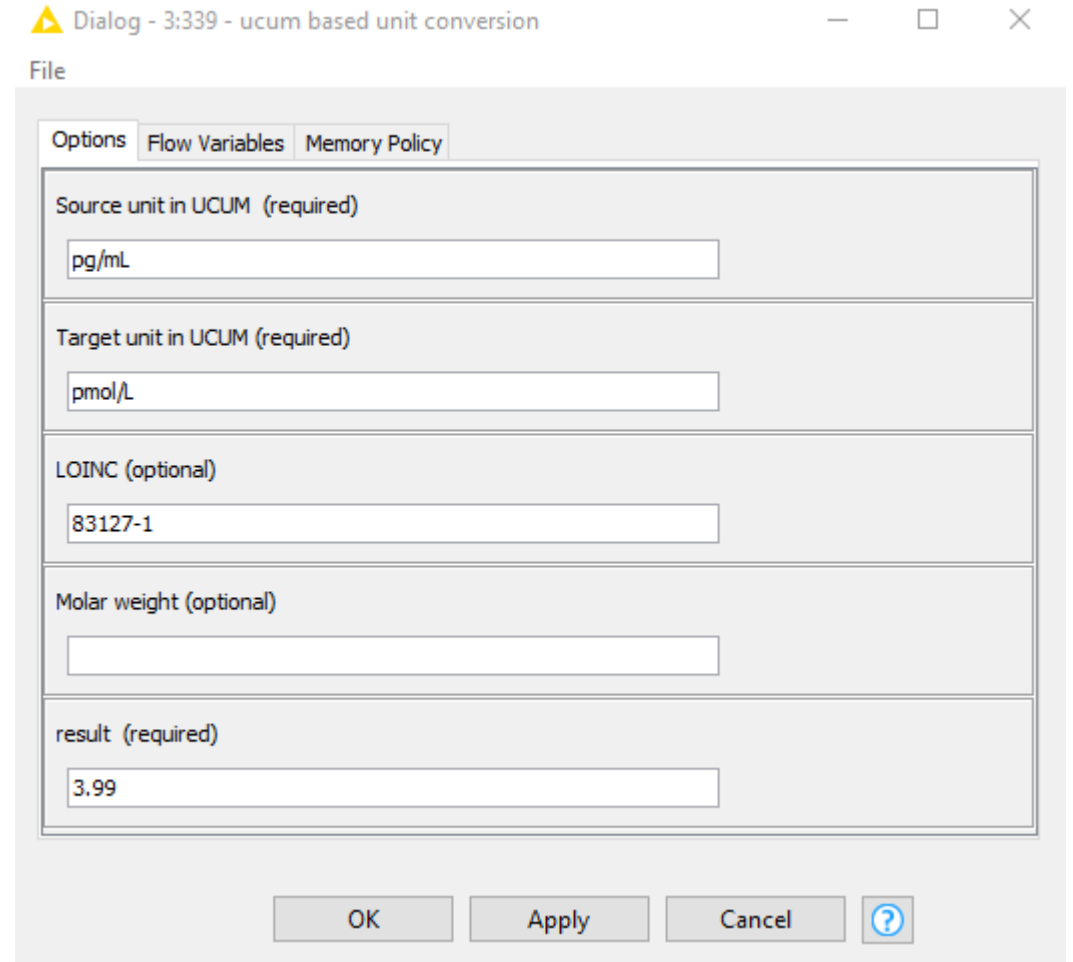


Search controlled terminology

Search LOINC in loinc.org and cdisc.org



Get LOINC metadata from loinc.org and CDISC library



Data Apps

- Data Apps provide a user interface to perform scalable and shareable data operations, such as visualization, data import, export, and preview.
- Provide actionable insights to end users whenever they need them.
- Through KNIME Server you can also share your Data App with 5, 10, or 1000 end-users.
- Data App to compare metadata of a study with GLV and generate discrepancy report.

KNIME WebPortal Monitoring Administration Sign out

Home > Dashboard > RAVE > rave metadata compare (8) > rave metadata compare (8) 2023-02-23 23:23:16

Select GLV Metadata: GLV_Standard_2022_01

Select Study Metadata: SEP361-302_Version_3_1_194PR2022

Update list of study metadata: Refresh

Compare Rave Metadata: Compare

GLV and Study ALS Comparison
SEP361-302

Issues count
SEP361-302 (125 issues)

SEP361-302 (125 issues)
Show: 25 entries
Search:

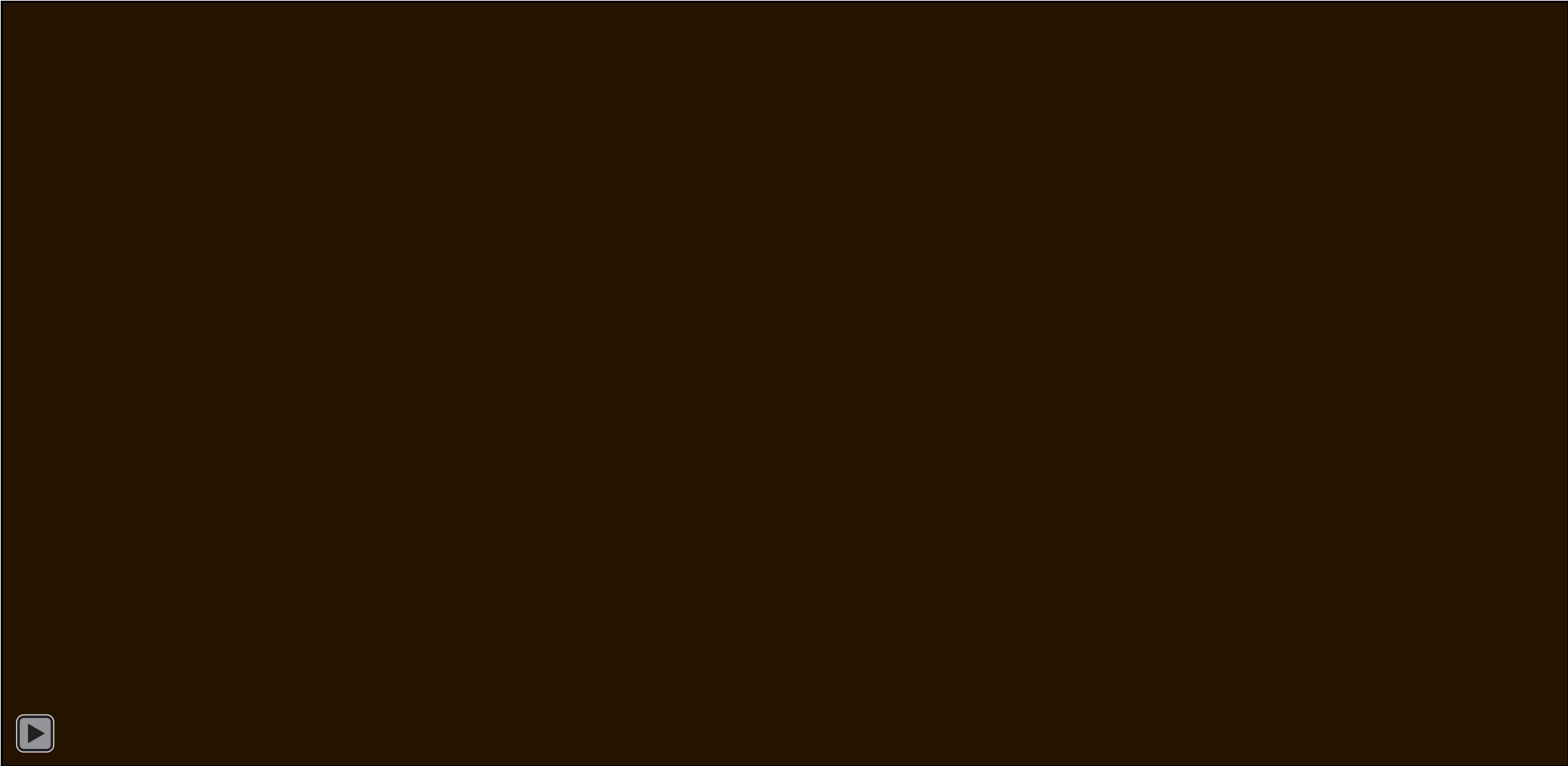
| checkID | Issues (N) | message |
|---------|------------|------------------------------------------------------------|
| 001 | 49 | FormOID is not standard |
| 002 | 0 | Standard FormOID is used but Form Name is not standard |
| 003 | 0 | Standard FormOID is used but Log Direction is not standard |
| 004 | 17 | Standard Form Name is used but FormOID is not standard |

Download
Download (xlsx)

Validation Report
Select Check ID from the graph or the summary to view the detail
Show: 10 entries
Search:
Delete this result Close

Sumitomo Pharma

Sumitomo Pharma America, Inc. All Rights Reserved. CONFIDENTIAL 17



GxP – Features on KNIME Server

Authentication

- Users can be authenticated using existing systems and can be added into defined user groups.

Versioning

- KNIME WF's can be versioned and restored to any previous versions on KNIME server.
- Different versions of KNIME WF's can be compared and generate report.

Access Control

- Folders and workflows can be provided with Read, Write, Execute permission to Users and User groups.

Audit

- All events on KNIME server including WF's ownerships, Folder and File access permissions, WF executions, WF updates are stored.
- All the audit data can be accessed and pulled out of server using predefined REST API.

Acknowledgements



- **Clinical Data Programming SMPA** - Cindy Ma, Jyotsna Bhate, Doris Menke, Sam Tomioka, Sravani Koganti and Jeff Xiong for their support in building workflows and components.
- **SMPA IT team** – For their extensive help in setting up infrastructure to support business needs.
- **KNIME Team** - Norbert Sas, Martyna Pawletta for their continued guidance and support.

References

KNIME Community Hub – Find workflows, nodes, components from user community

<https://hub.knime.com/>

KNIME Blog

<https://www.knime.com/blog?page=3>


Nodepit – Over 5K nodes, 19K workflows from user community and partners

<https://nodepit.com/>

<https://medium.com/low-code-for-advanced-data-science/why-knime-98c835afc186>

<https://medium.com/low-code-for-advanced-data-science>



 is a trademark of Sumitomo Pharma Co., Ltd., used under license.
SUMITOMO PHARMA is a trademark of Sumitomo Pharma Co., Ltd., used under license.
SUMITOMO is a registered trademark of Sumitomo Chemical Co., Ltd., used under license.
Sumitomo Pharma America, Inc. is a U.S. subsidiary of Sumitomo Pharma Co., Ltd.
© 2024 Sumitomo Pharma America, Inc. All rights reserved.