Implementing a Cheminformatics Workflow Environment from Scratch in Medium-Sized Pharma Company - Lessons Learned

Dora Šribar

Berlin, November 4th 2021

Nuvisan Innovation Campus Berlin

- Fully integrated Science contract research organization specializing in drug discovery and development
- Areas of research:
  - Lead Discovery
  - Life Science Chemistry
  - Therapeutic Research
  - Preclinical Compound Profiling
  - Digital Life Science
- We needed the platform that will:
  - Enable the implementation of different workflows for the processing, analysis and predictive modeling of chemical data
  - Enable sharing of the developed workflows across the organization with ensured data privacy
  - Be budget-friendly

INNOVATION CAMPUS BERLIN
A NUVISAN Company

KNIME at Innovation Campus Berlin: Infrastructure

KNIME Analytics Platform

- Implementation of project-specific workflows within Digital Life Science

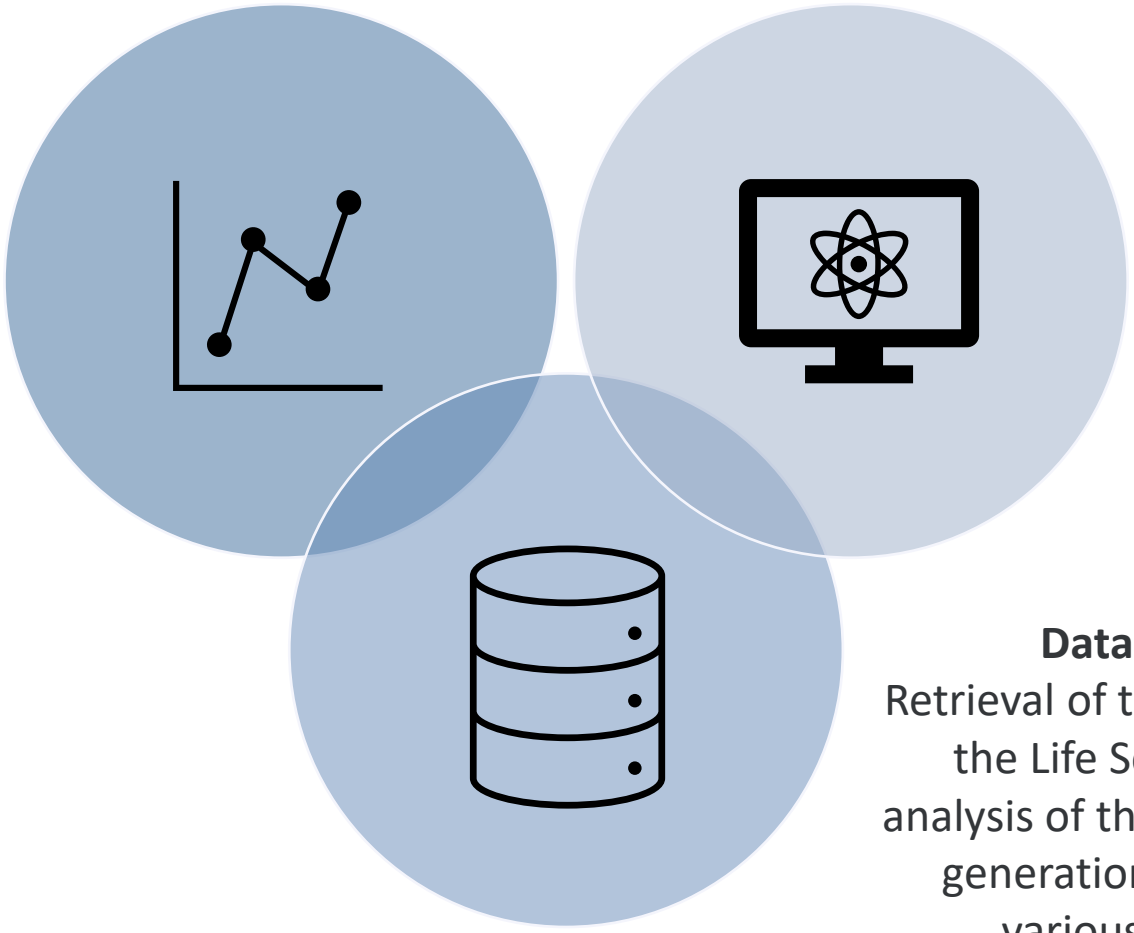
KNIME Server

- Implementation of various apps that are available to different departments
- Currently accessible through WebPortal, integration with other analytics platforms (e.g., TIBCO Spotfire®) is planned
- Technical details:
  - KNIME Server Large license split into two licenses for development and production server
  - Hosted on Virtual Machine with Ubuntu 20.04.3 LTS

INNOVATION CAMPUS BERLIN
A NUVISAN Company

KNIME at Innovation Campus Berlin: Use cases

Data wrangling and analysis
Processing of different file formats, integration of different data sources, generation of user-friendly output,…

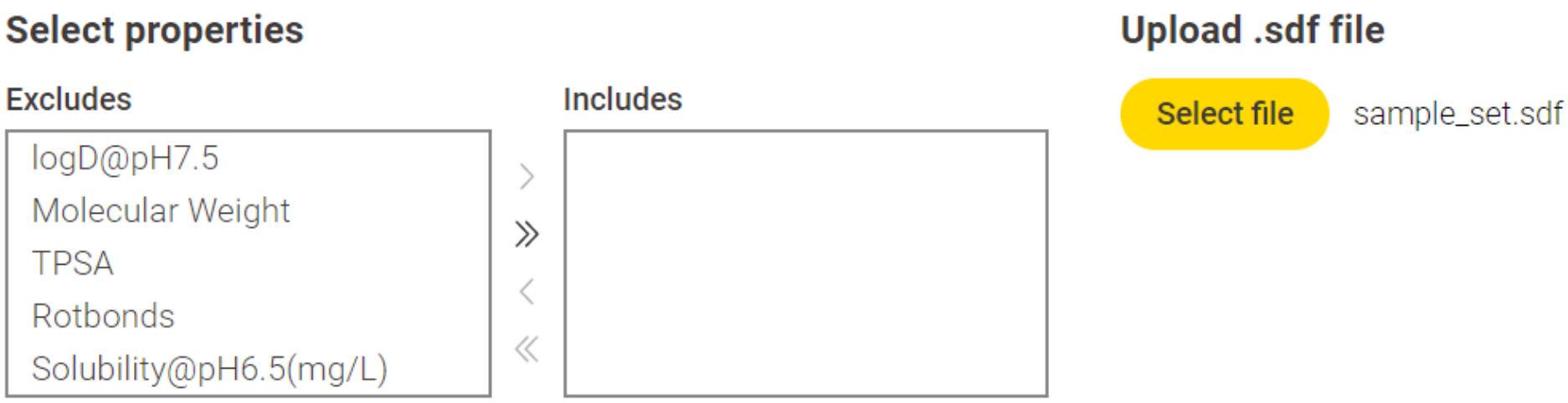


Cheminformatics-specific application
Prediction of molecular properties, clustering, tools for structure-activity relationship (SAR) analysis,..

Database mining
Retrieval of the information from the Life Science Database, analysis of the retrieved data and generation of the output in various file formats,..

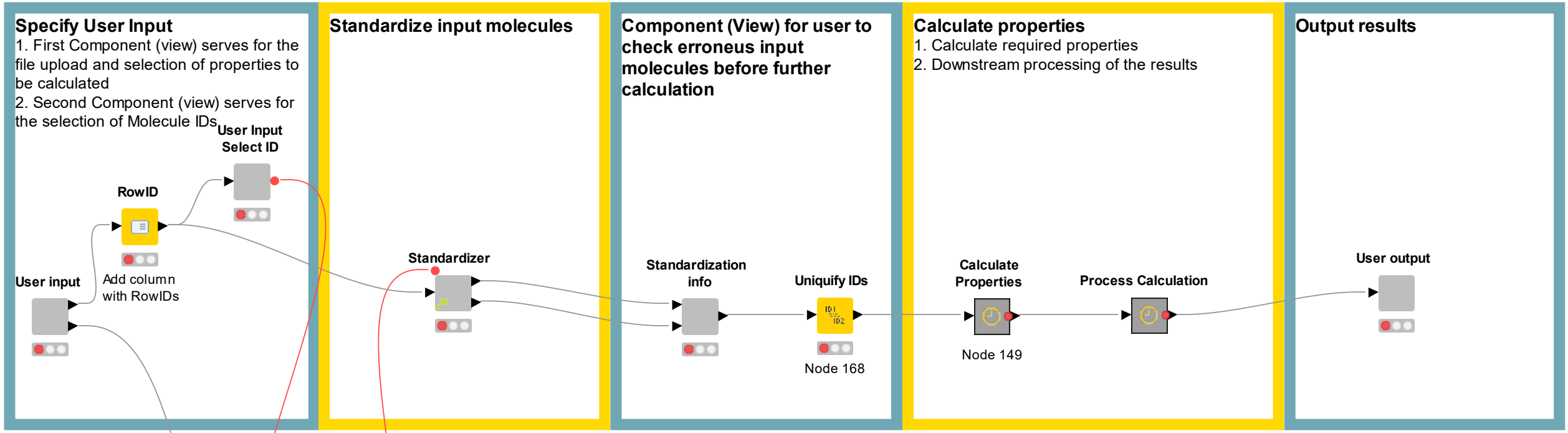
INNOVATION CAMPUS BERLIN
A NUVISAN Company

Example workflow: Calculation of molecular properties



User input: File upload and selection of properties in the WebPortal

Example workflow: Overview



Calculation of different properties

Calculation of different properties (Molecular Weight, Solubility, LogD, CYP inhibition,…) either directly via various chemistry integrations in KNIME (e.g. RDKit, CDK, Indigo, Vernalis,..) or by implementing machine learning models based on the *in-house* life science data

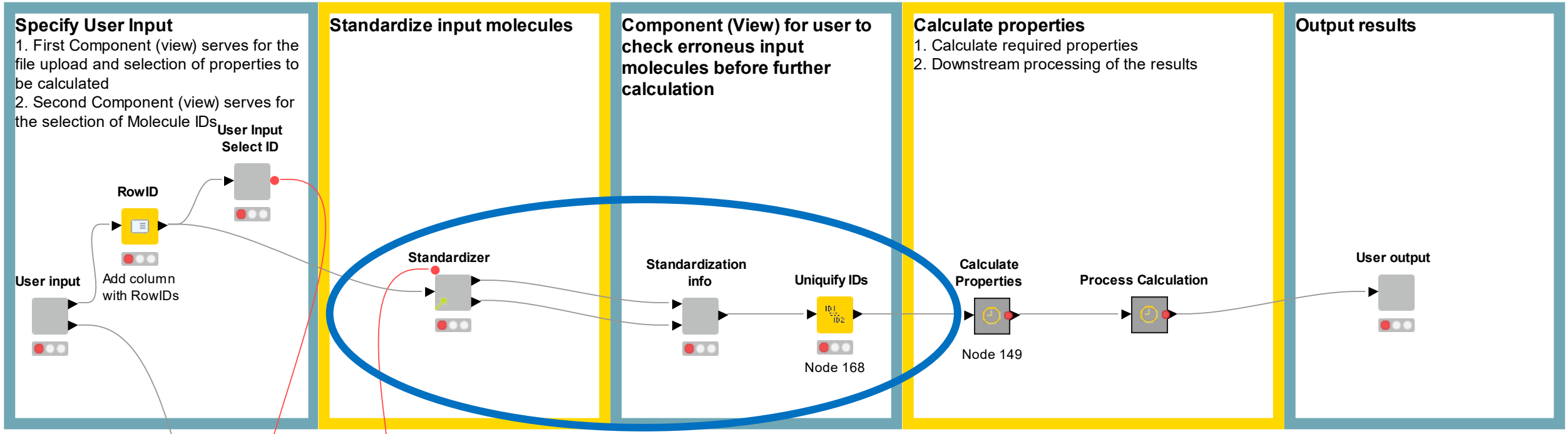
INNOVATION CAMPUS BERLIN
A NUVISAN Company

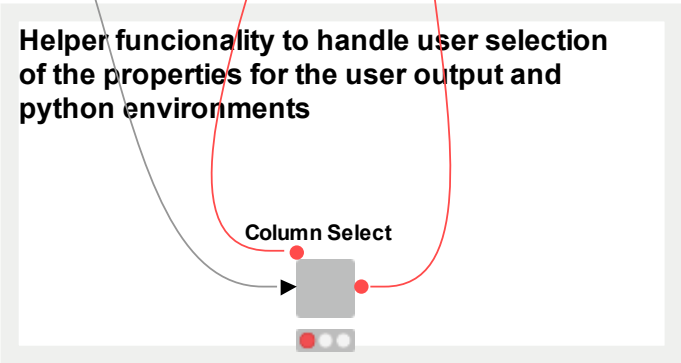
Example worklow: User input

User Input

INNOVATION CAMPUS BERLIN
A NUVISAN Company

Example workflow: Standardization of the molecules



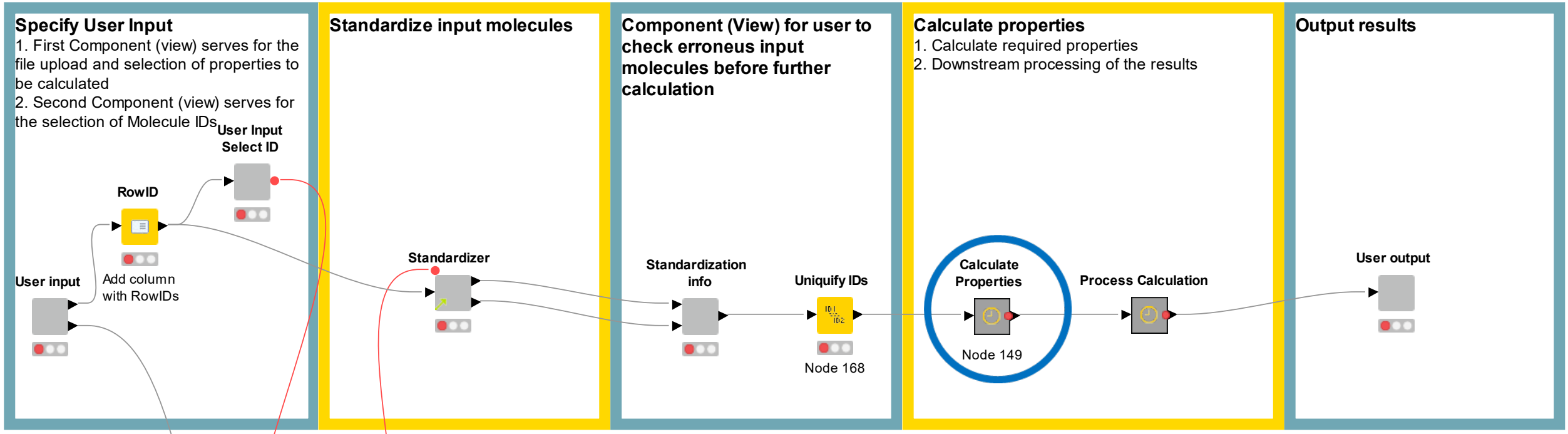


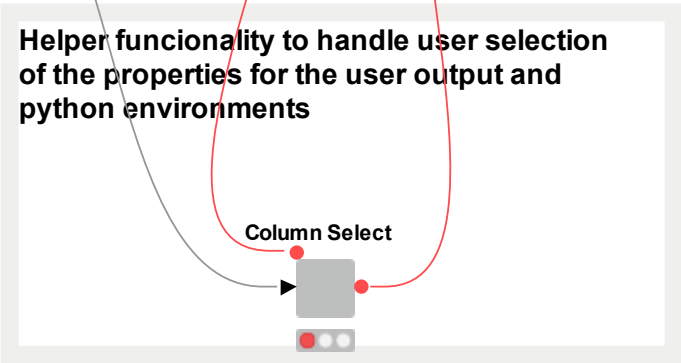
Standardizer

Structure standardization and check-up is mostly based on this very useful Python package:
https://github.com/chembl/ChEMBL_Structure_Pipeline

INNOVATION CAMPUS BERLIN
A NUVISAN Company

Example workflow: Calculation of properties



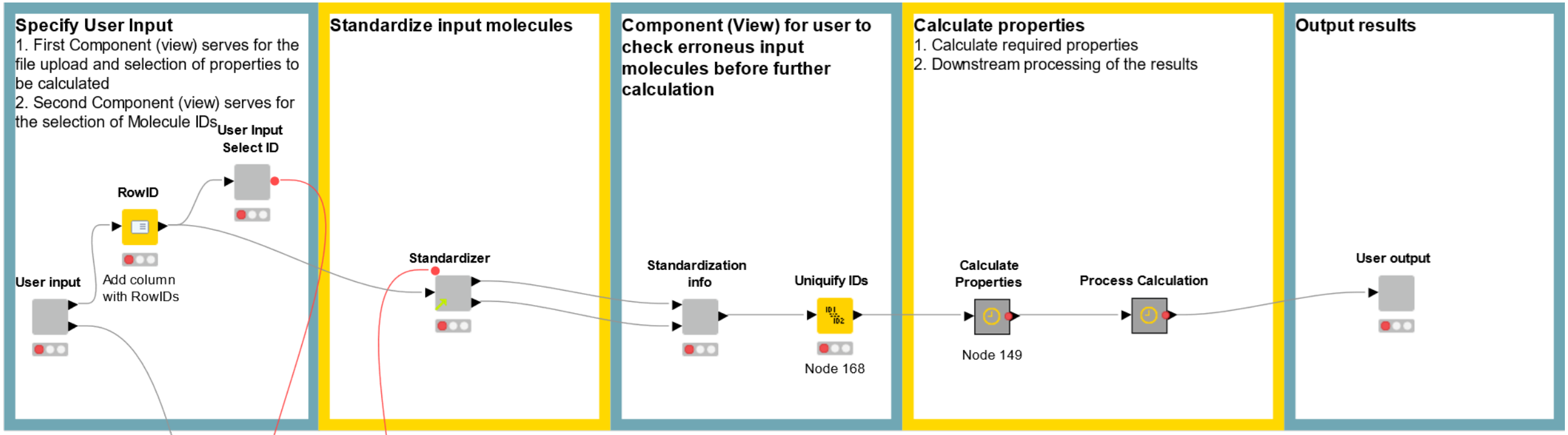


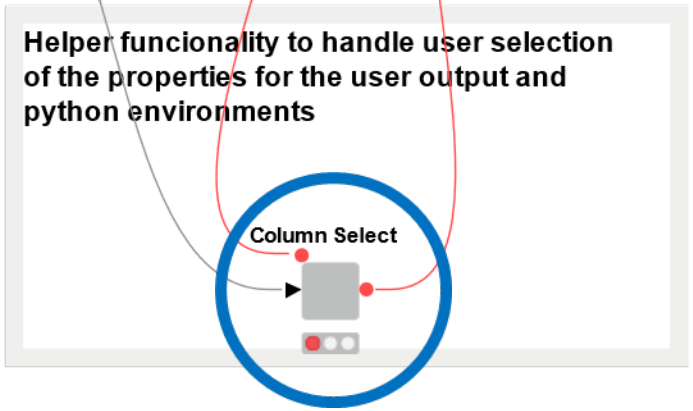
Integration with other programming languages

Machine learning models and different python-based scripts (or other programming languages) can be implemented easily using Python Integration Nodes

INNOVATION CAMPUS BERLIN
A NUVISAN Company

Example workflow: Specifying different environments



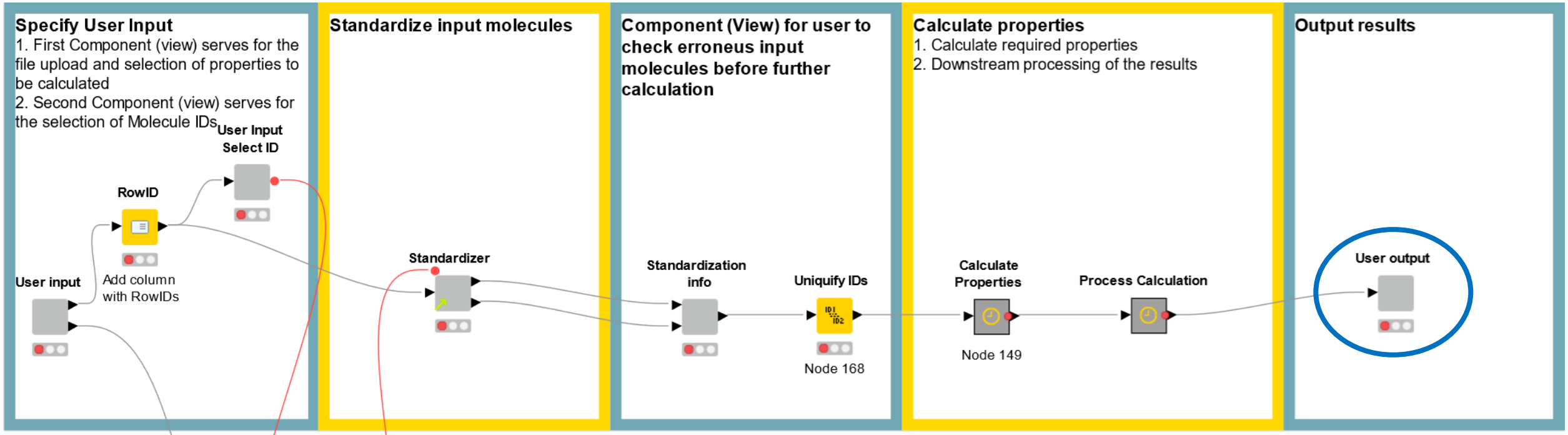


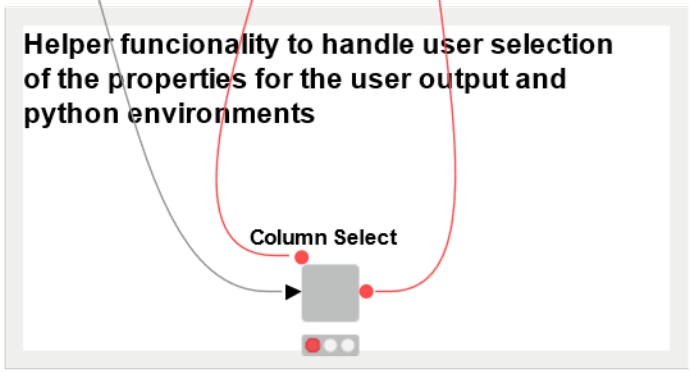
Using multiple python environments

Different environments (e.g conda or pyenv) per each node can be used by specifying the paths using flow variables

https://docs.knime.com/latest/python_installation_guide/index.html#configure_python_integration

Example workflow: User output





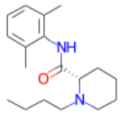
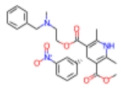
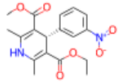
Example workflow: User output in WebPortal

Download

Download Results (sdf)

Results

Show entries Search:

	Structure	ID	Molecular Weight	Rotbonds
☐		levobupivacaine	288.22	5
☐		(S)-nicardipine	479.21	9
☐		(S)-nitrendipine	360.13	5

User output: Display of results in WebPortal

If you want to color the cells in the WebPortal, check this cool snippet!
https://kni.me/w/Q-E_6XmNXVqQ_WgV

Additionally,

If you are interested in pptx, pdf,… reports, check the BIRT integration in KNIME:
https://www.knime.com/getting-started-knime-birt

Lessons learned

- Always check the output of the nodes! Anything that can go wrong will go wrong.
- In case you need to implement specific, more complex functionality:
  - Check https://hub.knime.com/
  - If there is no available KNIME implementation, try to implement it yourself (Python, R or whatever programming language you prefer) and integrate in KNIME environment. There is a good chance that someone already implemented what you are looking for in other programming language.
- If you have any questions or encounter any problems:
  - First check https://forum.knime.com/
  - If you still don´t find a solution, contact KNIME support
  - This may save you lots of time and energy

INNOVATION CAMPUS BERLIN
A NUVISAN Company

Team

- Vysalini Akshintala
- Andreas Bender
- Chetan Chintha
- Jörg Fanghänel
- **Andreas Göhrig**
- **Nadine Homeyer**
- **Christopher Linn**
- David Machalz
- Franz von Nussbaum
- Gaus Reza
- Holger Steuber
- **Franziska Witte**

Also, we are hiring! Check the positions on https://career.nuvisan.com/.