Open for Innovation
KNIME

Welcome to
Working with the RDKit in KNIME Analytics Platform

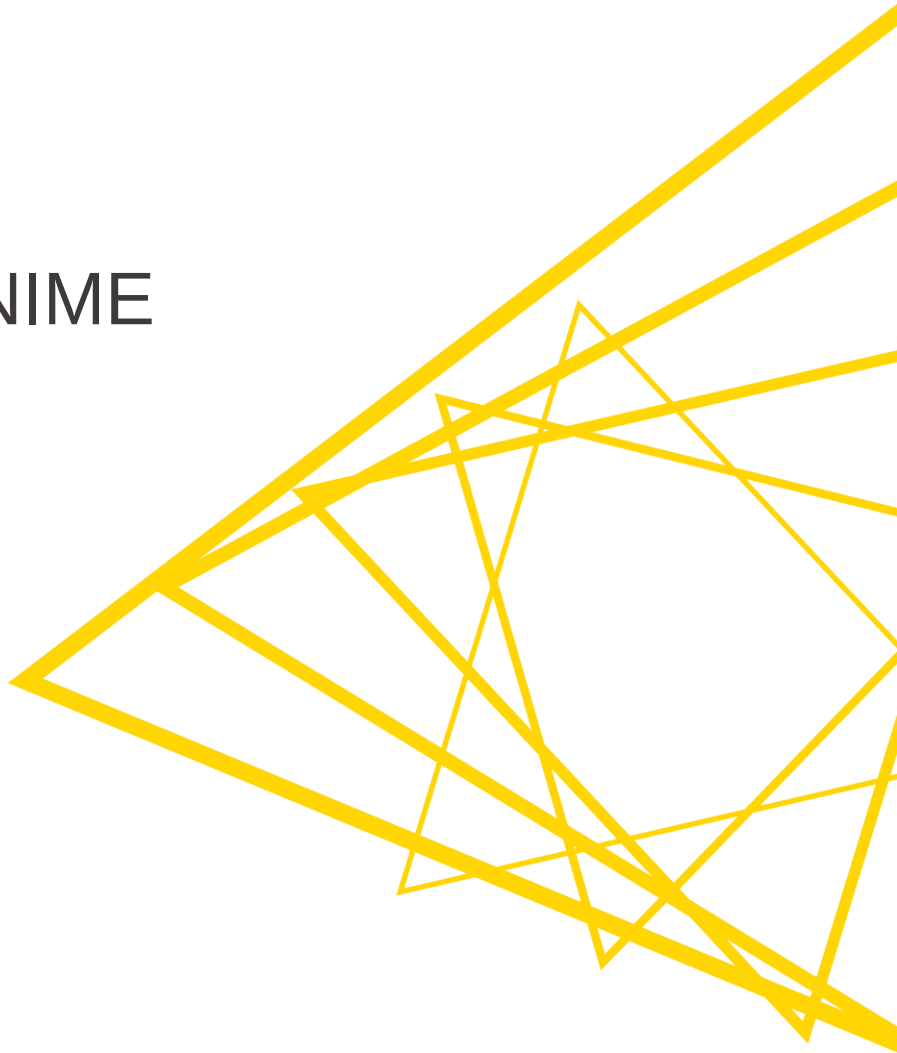
Going live at:

Chicago 11:00 am

San Francisco 9:00 am

New York 12:00 pm

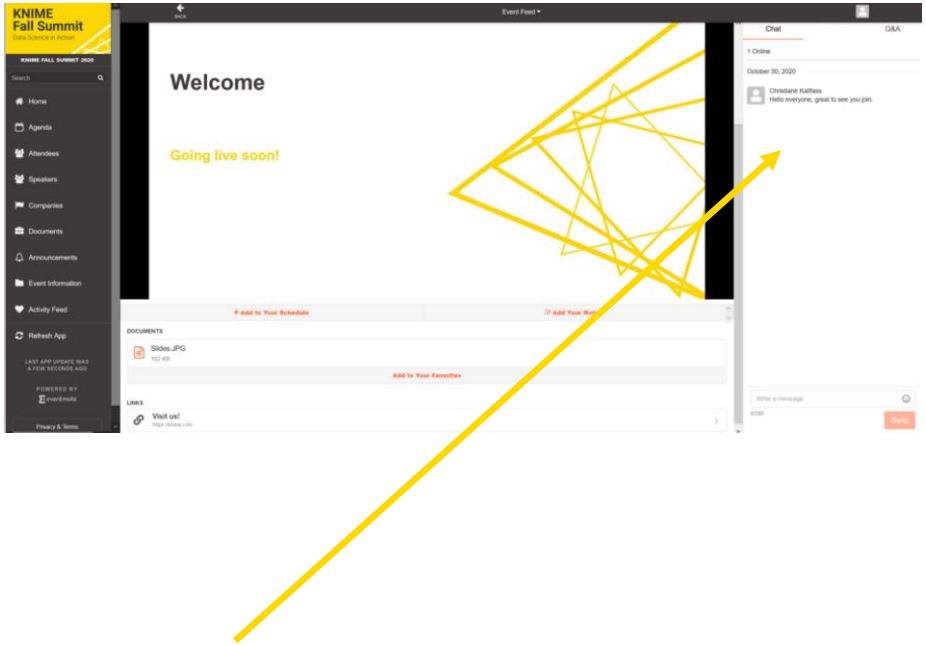
Berlin 6:00 pm



Housekeeping

- Post in the chat where you are dialing in from and discuss with other attendees
- Questions? Post them in the Q&A

Questions will be answered after the presentation.



© 2020 KNIME AG. All rights reserved.

The agenda

- Brief introduction to KNIME
- **Part I**: Alice

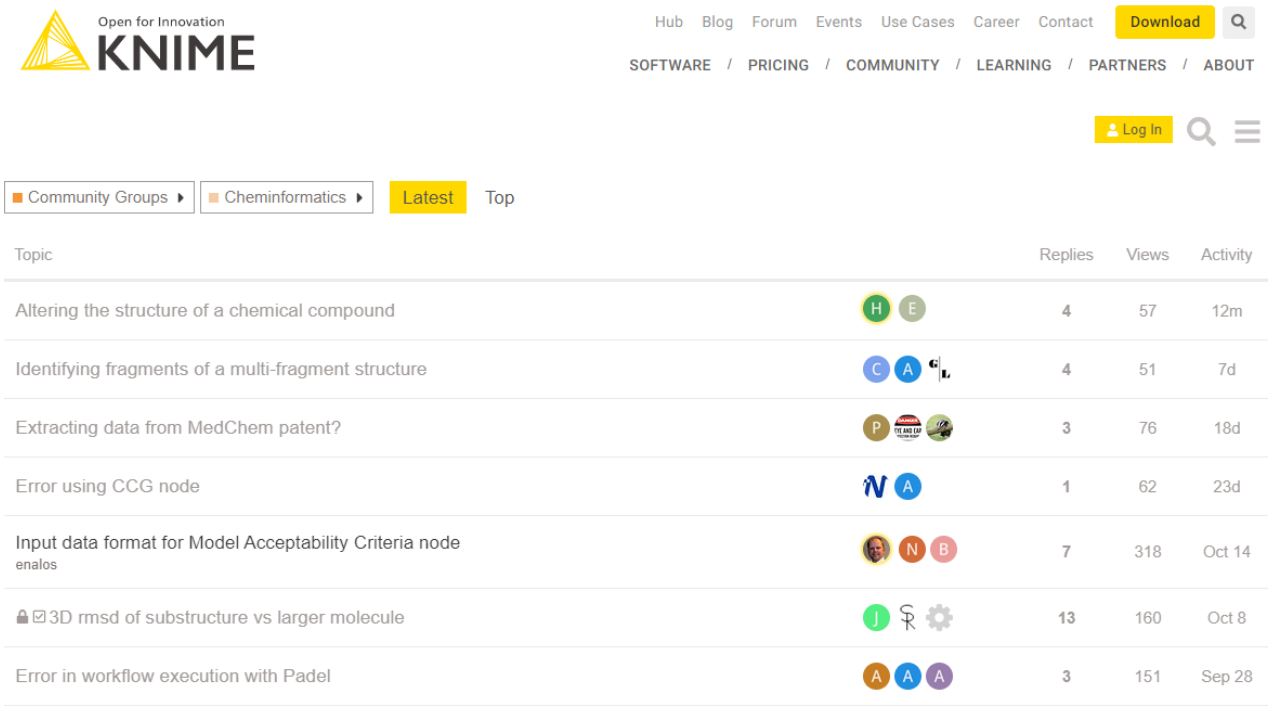
  Reactions in RDKit
- **Part II**: Greg

  3D visualization of molecules

© 2020 KNIME AG. All rights reserved.

Post your questions in the forum

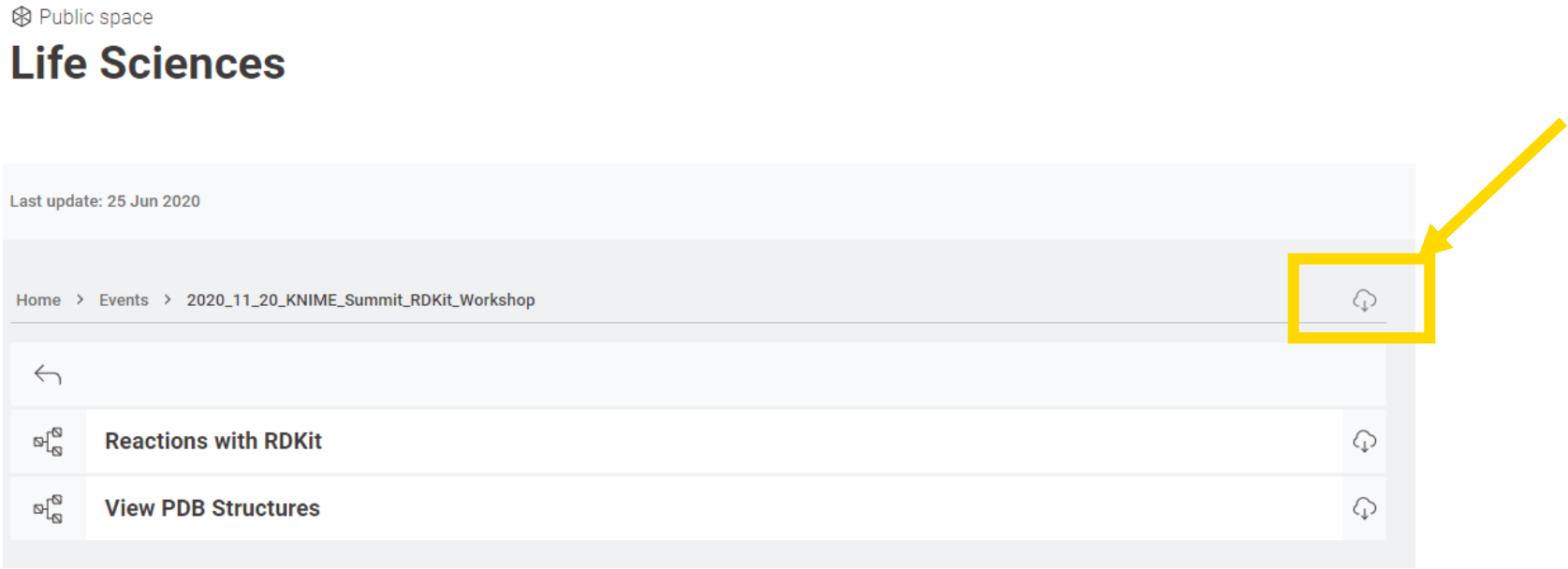
https://forum.knime.com/c/special-interest-groups/cheminformatics/36



© 2020 KNIME AG. All rights reserved.

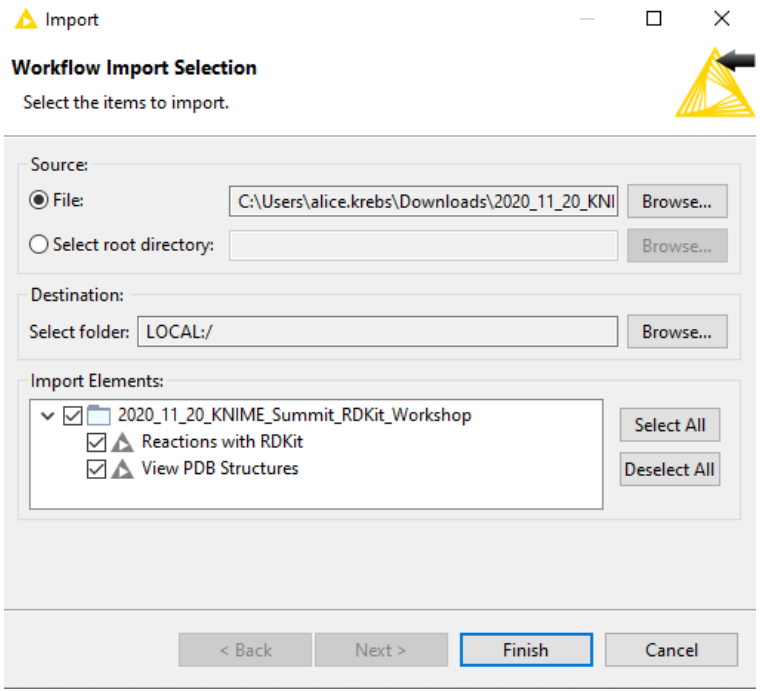
The workflows

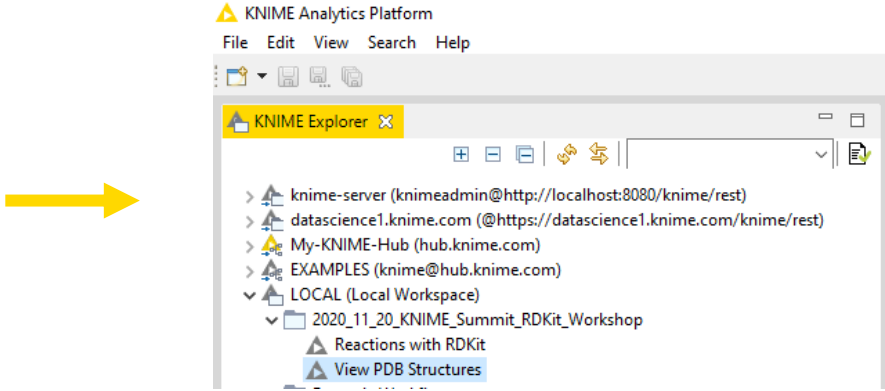
https://hub.knime.com/knime/spaces/Life%20Sciences/latest/Events/2020_11_20_KNIME_Summit_RDKit_Workshop/



© 2020 KNIME AG. All rights reserved.

Download



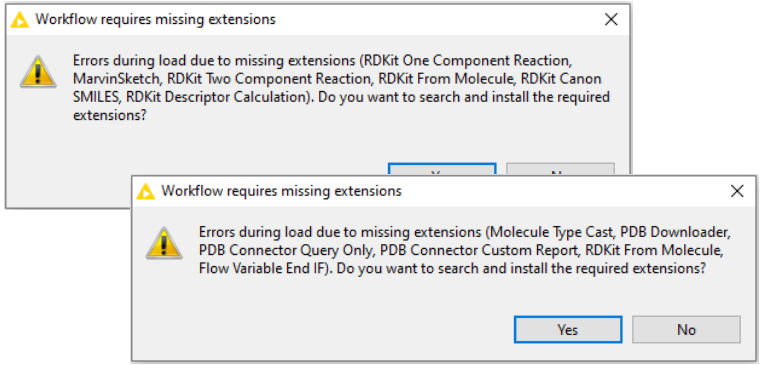


© 2020 KNIME AG. All rights reserved.

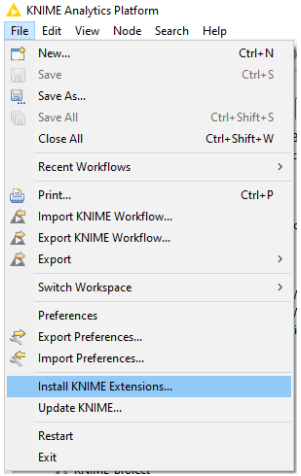
Open for Innovation KNIME

Install extensions

- Open the workflows…



- … or via File > Install KNIME Extensions



  - ChemAxon/Infocom Marvin Extensions Feature
  - RDKit KNIME integration
  - KNIME Base Chemistry Types & Nodes
  - Vernalis KNIME Nodes

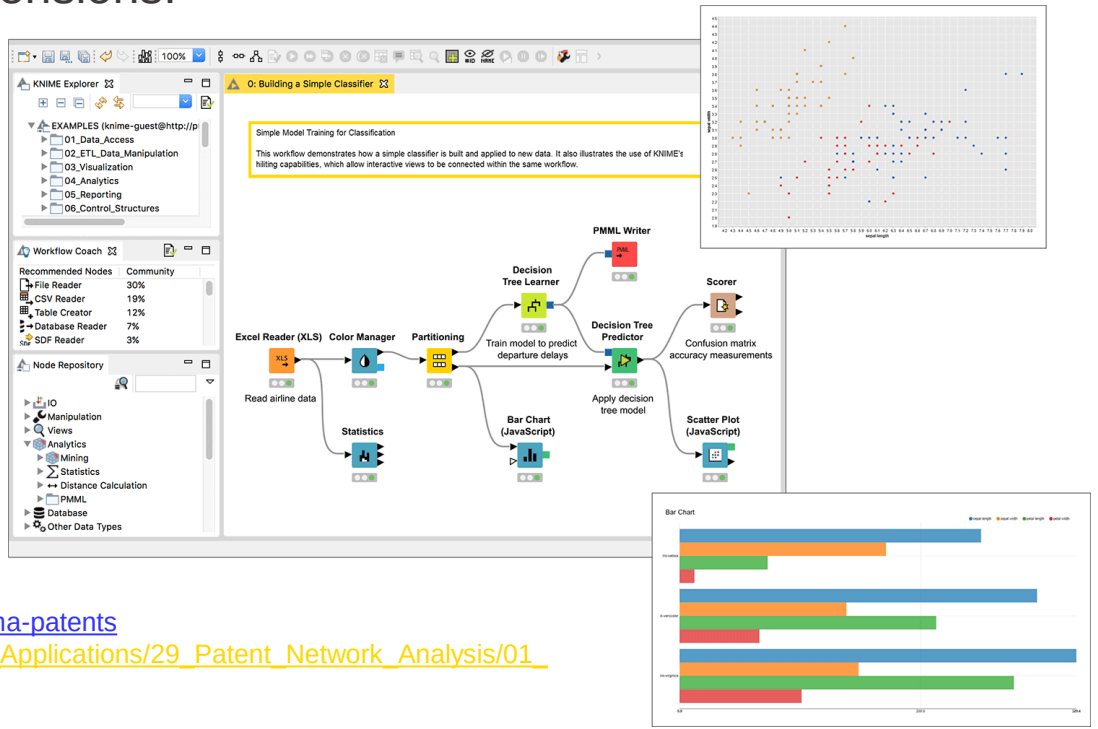
© 2020 KNIME AG. All rights reserved.

Open for Innovation
KNIME

Introduction to KNIME

What is KNIME Analytics Platform?

- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
  - Text Mining
  - Network Mining
  - Cheminformatics
  - Many integrations,
    such as Java, R, Python,
    Weka, Keras, Plotly, H2O, etc.



- https://www.knime.com/competitive-intelligence-for-pharma-patents
- https://hub.knime.com/knime/spaces/Examples/latest/50_Applications/29_Patent_Network_Analysis/01_Tarceva_neighbor_network_from_SureChEMBL

© 2020 KNIME AG. All rights reserved.

KNIME Software - One Ecosystem

Create

Gather & Wrangle

Model & Visualize

Productionize

Deploy & Manage

Consume & Optimize

KNIME Analytics Platform

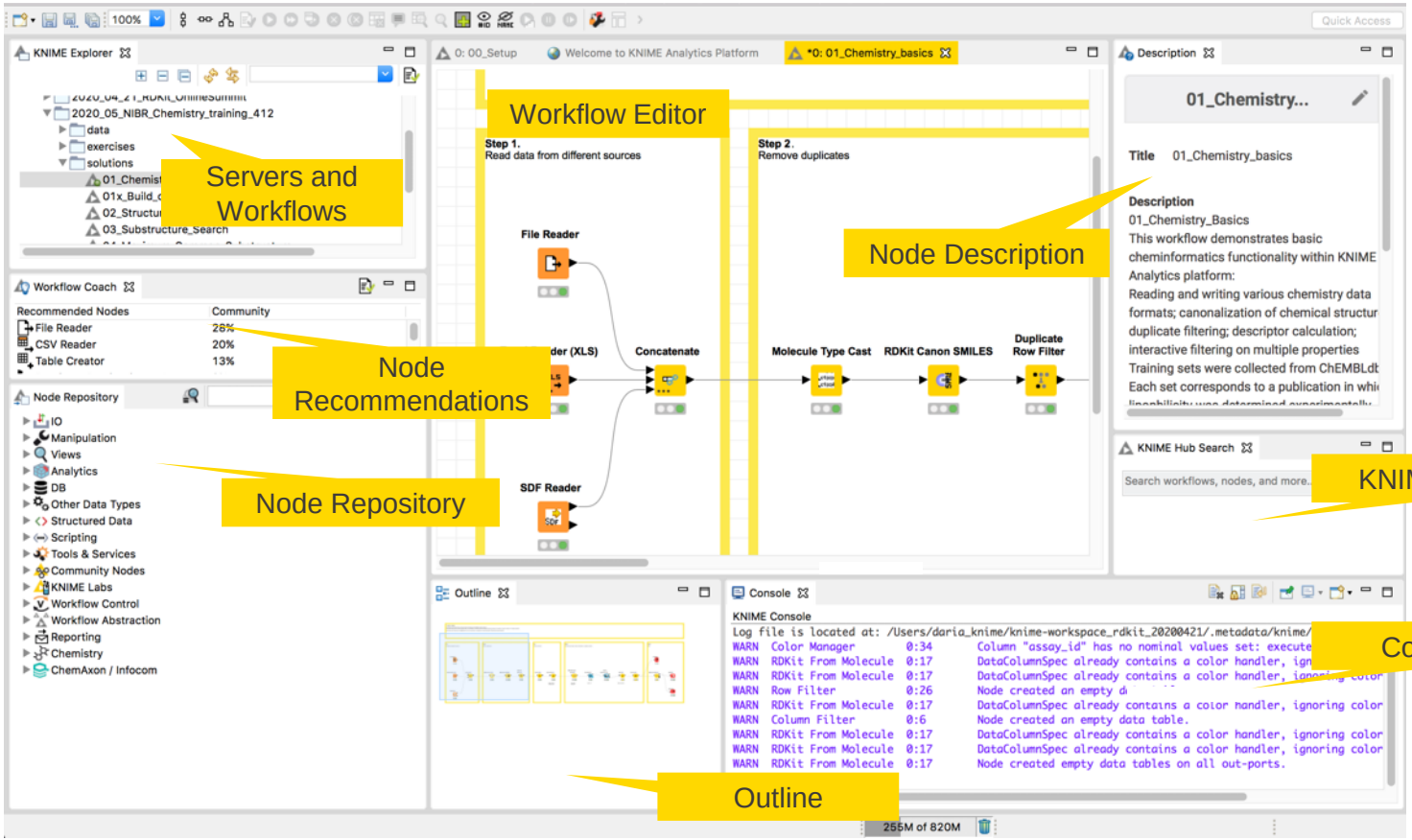
- KNIME Extensions
- KNIME Integrations
- Community Extensions
- Partner Extensions

KNIME Server

- KNIME WebPortal
- Data Science as a Service

© 2020 KNIME AG. All rights reserved.

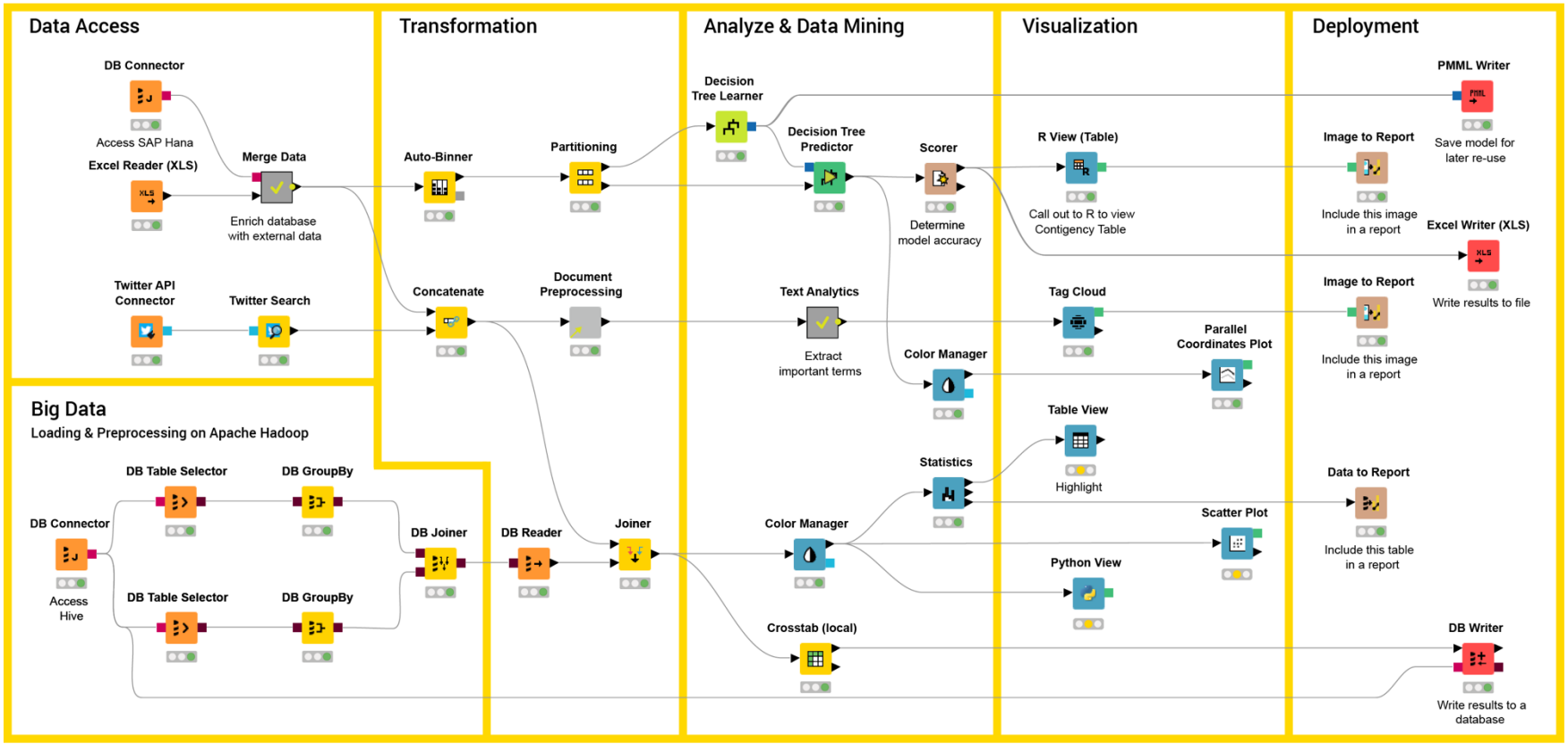
The KNIME Workbench



© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

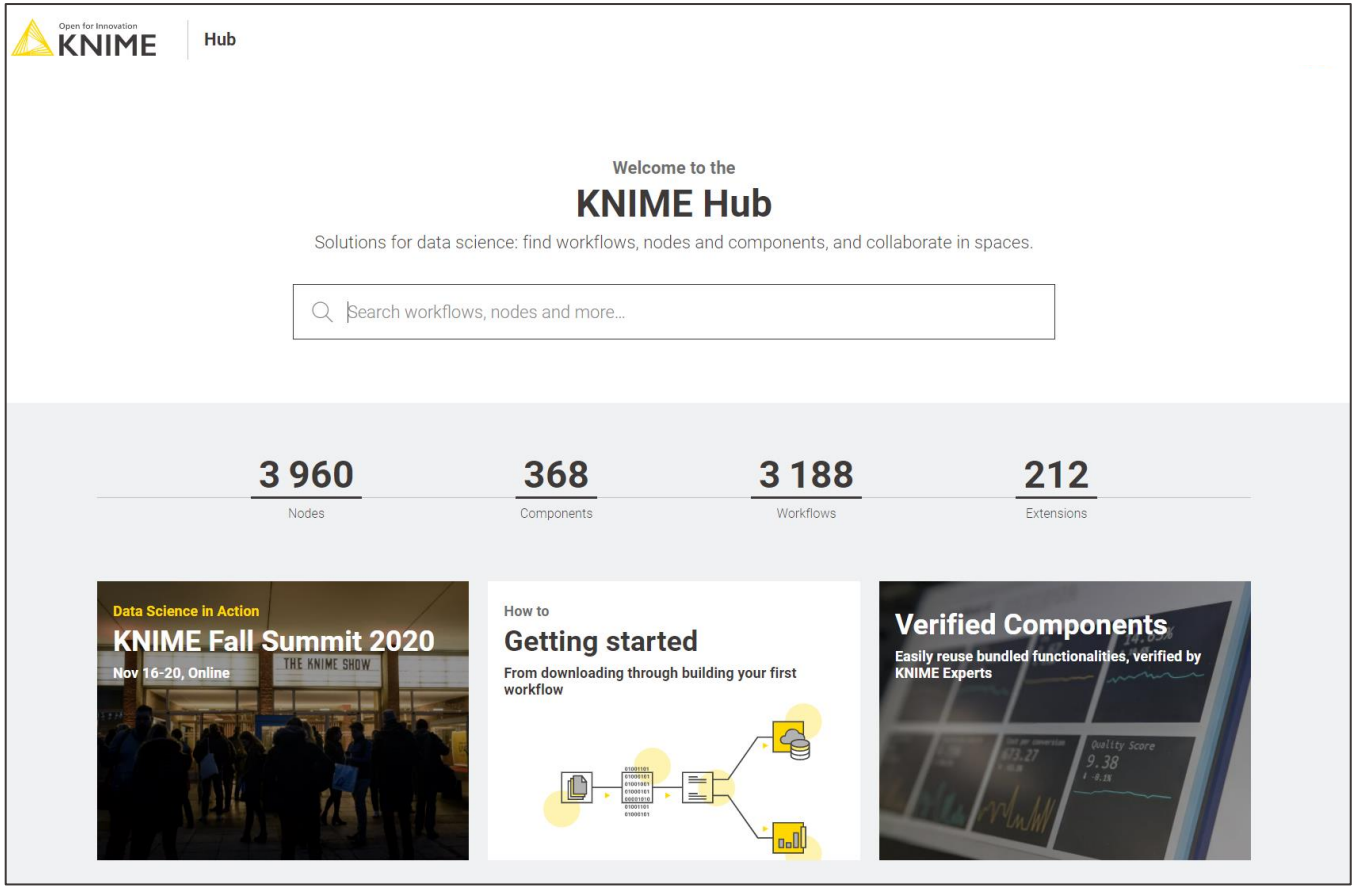
3000+ Nodes for all Steps of End-To-End Data Science



© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

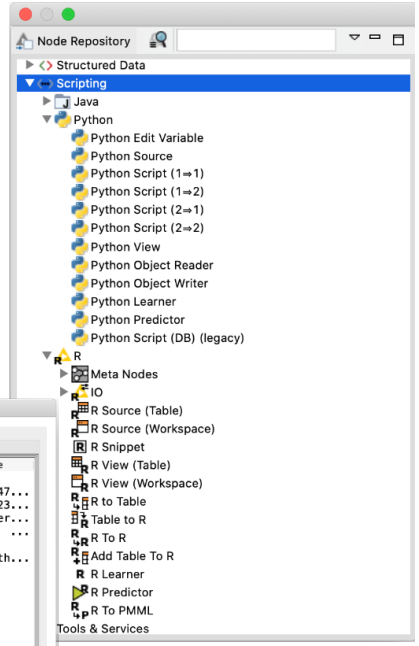
KNIME Hub: Searching, Sharing, and Collaborating

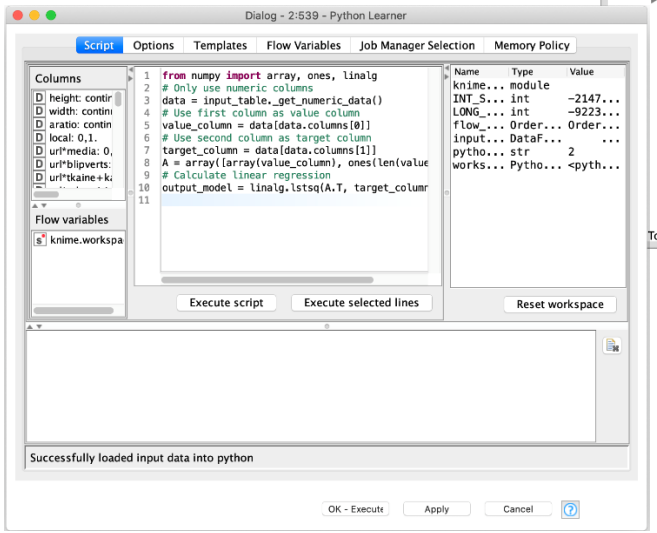


© 2020 KNIME AG. All rights reserved.

Scripting Integrations: R and Python

- Run R or Python code in KNIME Analytics Platform
- Works with existing Python and R installations
- Syntax highlighting support
- Different nodes for many tasks

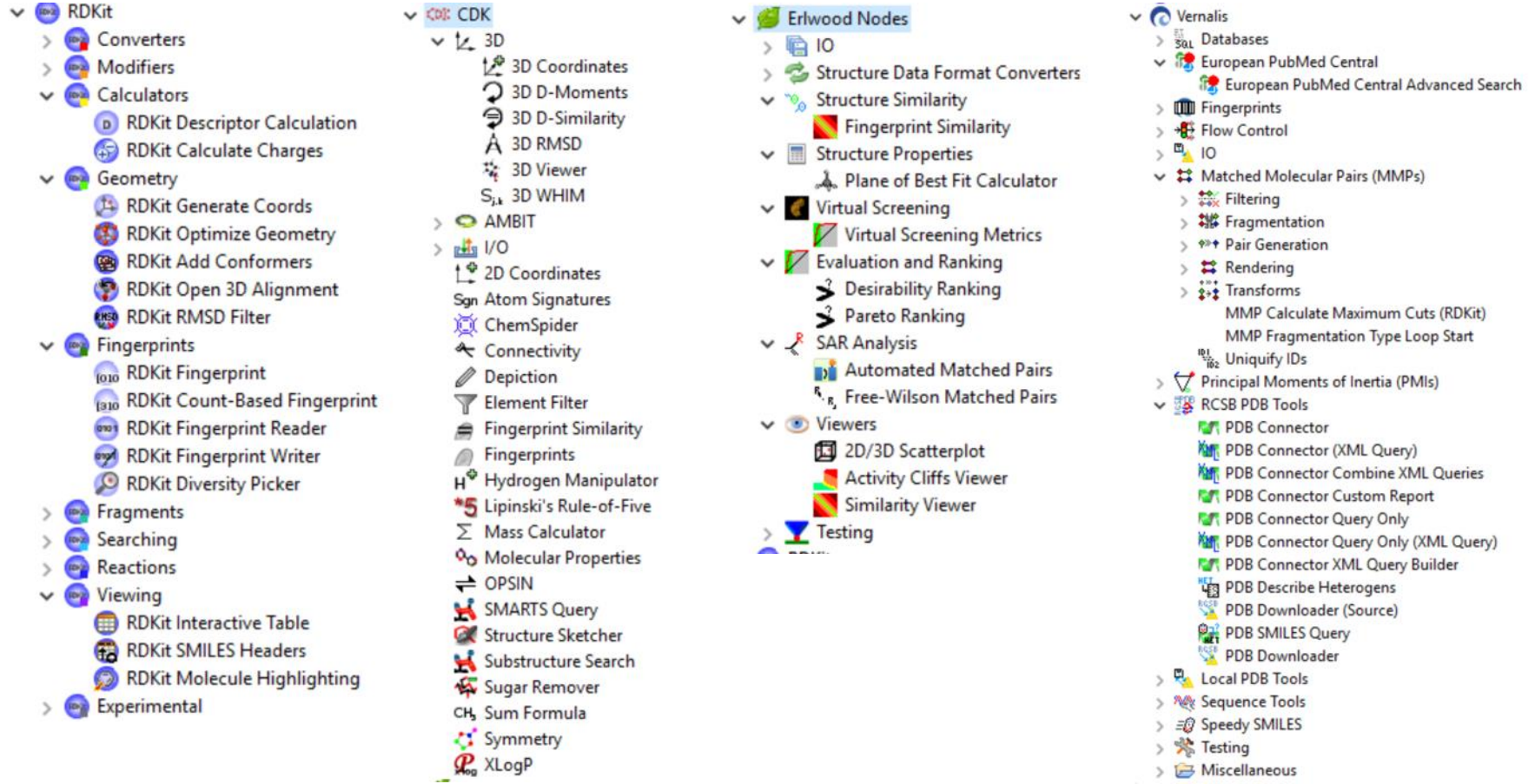




© 2020 KNIME AG. All rights reserved.

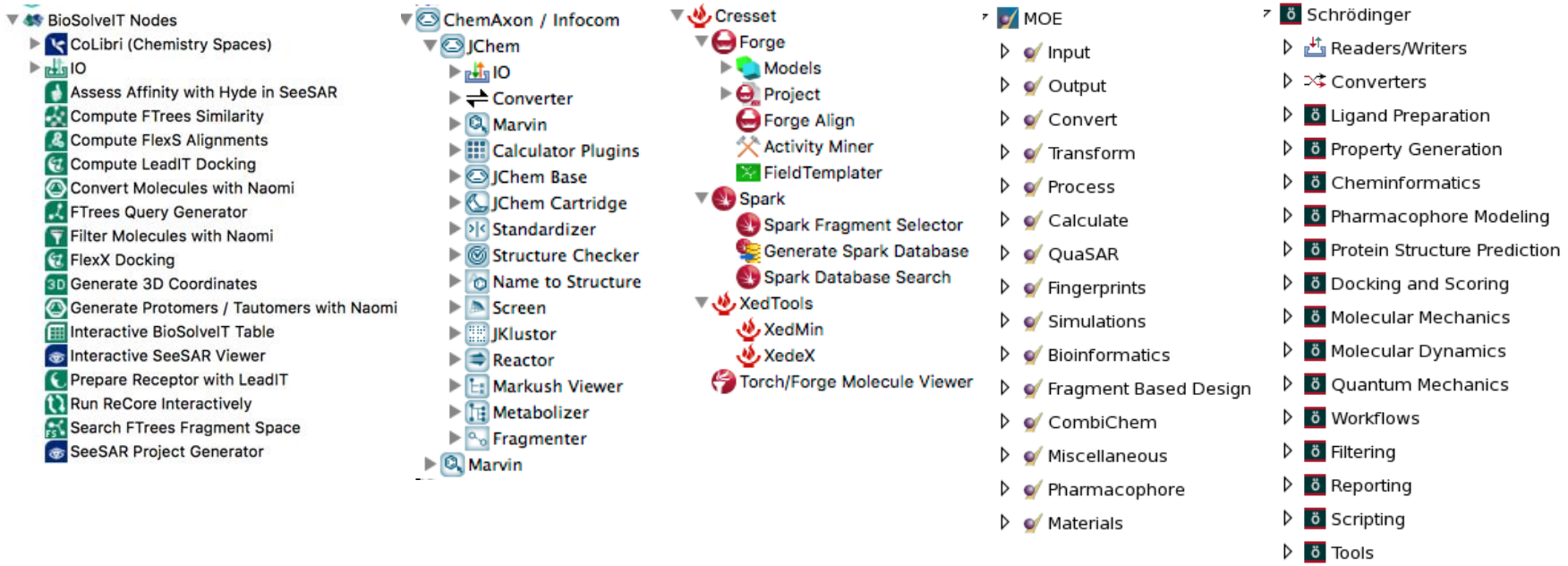
Open for Innovation KNIME

Selected Open Source Extensions for Cheminformatics



© 2020 KNIME AG. All rights reserved.

Selected Commercial Extensions for Cheminformatics

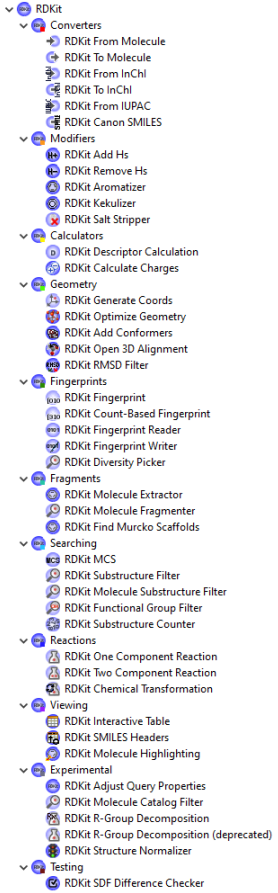


© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

What is the RDKit?

- Open source cheminformatics library in C++
- Wrappers for KNIME maintained by the open source community
- Useful for:
  - Descriptor calculation
  - Cleaning structures
  - InChI conversion
  - Canonical SMILES
  - Fingerprints
  - Scaffolds/substructures
  - Reaction simulation
  - and more…



https://www.rdkit.org

© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

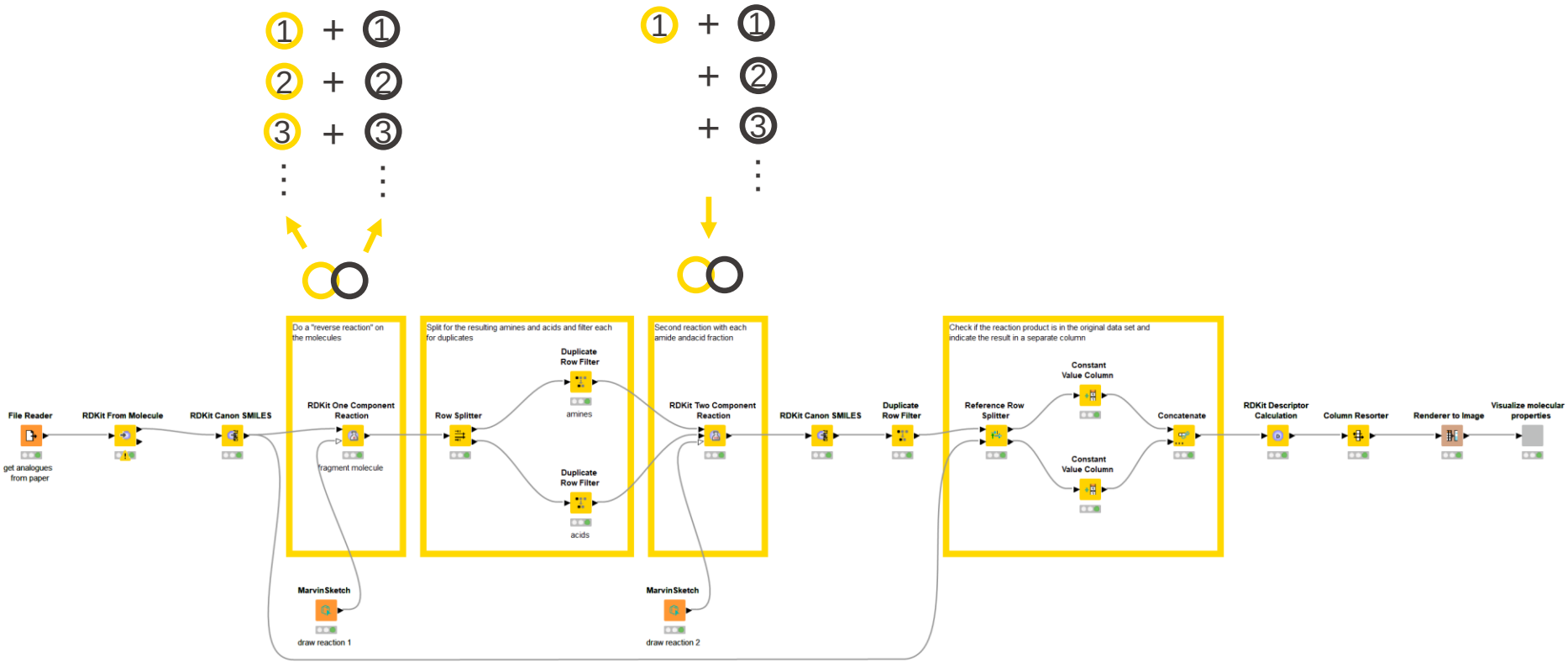
Open for Innovation
KNIME

Reactions with RDKit

Alice Krebs

KNIME Fall Summit Webinar
20 November 2020

The workflow



https://hub.knime.com/knime/spaces/Life%20Sciences/latest/Events/2020_11_20_KNIME_Summit_RDKit_Workshop/Reactions%20with%20RDKit

© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

The data

Journal of Medicinal Chemistry

pubs.acs.org/jmc

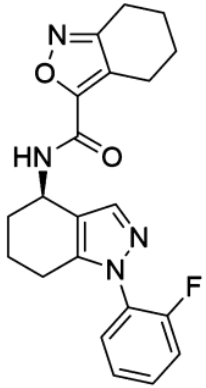
Article

This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

ACS AUTHORCHOICE

Optimization of Tetrahydroindazoles as Inhibitors of Human Dihydroorotate Dehydrogenase and Evaluation of Their Activity and In Vitro Metabolic Stability

Gergana Popova,* Marcus J. G. W. Ladds, Lars Johansson, Aljona Saleh, Johanna Larsson, Lars Sandberg, Sara Häggblad Sahlberg, Weixing Qian, Hjalmar Gullberg, Neeraj Garg, Anna-Lena Gustavsson, Martin Haraldsson, David Lane, Ulrika Yngve, and Sonia Lain



(*R*)-HZ05

https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01658

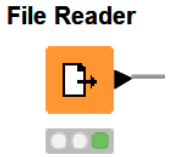
© 2020 KNIME AG. All rights reserved.

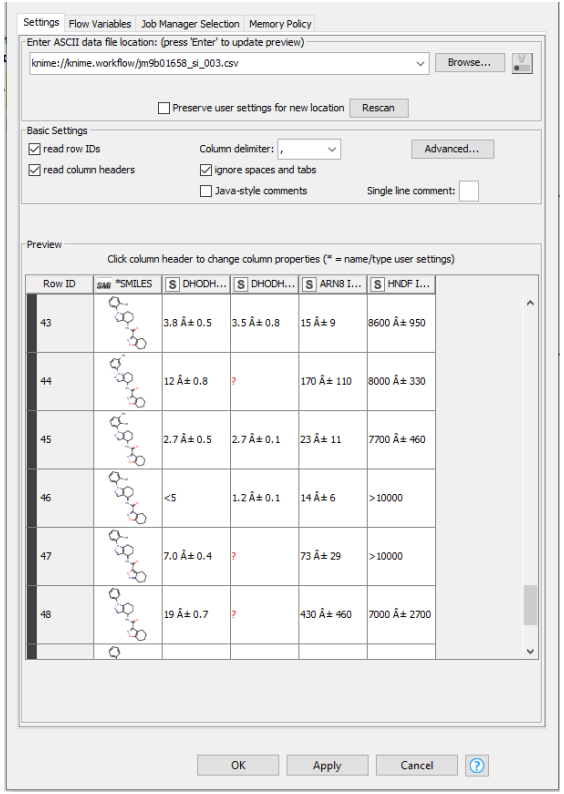
Open for Innovation KNIME

The nodes – File Reader

- tip: workflow-relative paths to the data

Local: "C:\Users\alice.krebs\knime-workspace\2020_11_20_KNIME_Summit_RDKit_Workshop\Reactions with RDKit\jm9b01658_si_003.csv"



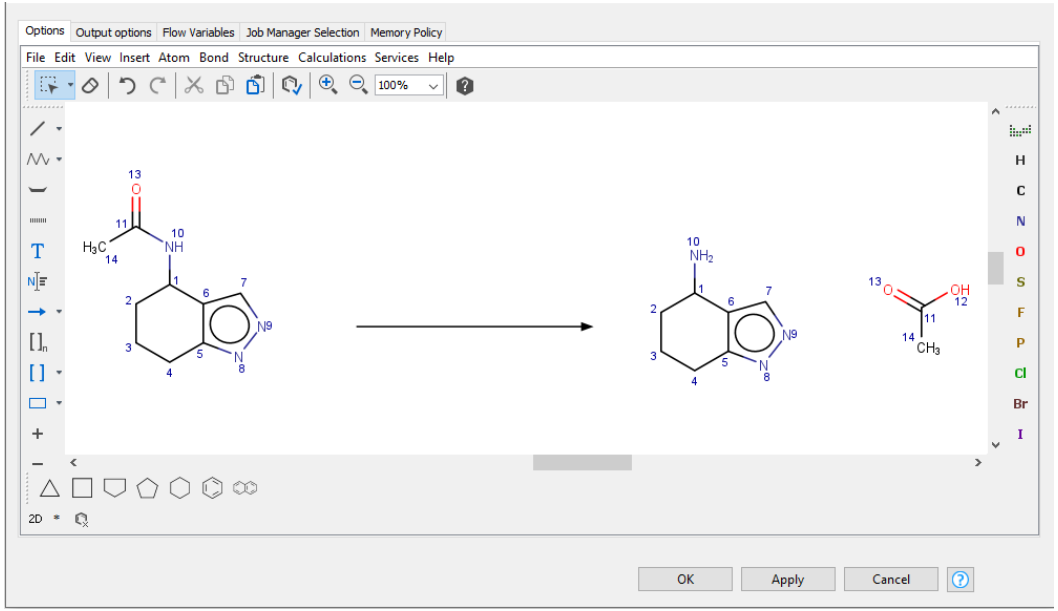


https://www.youtube.com/watch?v=U9sP4g4yGwY&t=124s

© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

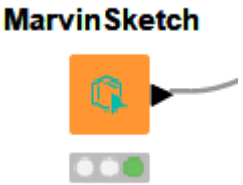
The nodes – Marvin Sketch

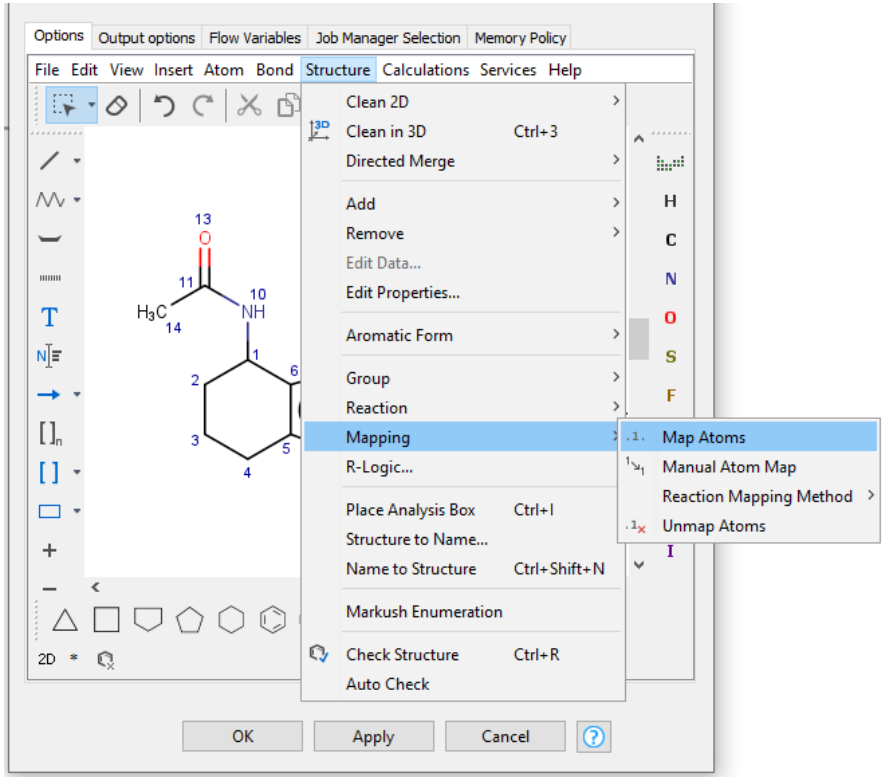


© 2020 KNIME AG. All rights reserved.

The nodes – Marvin Sketch

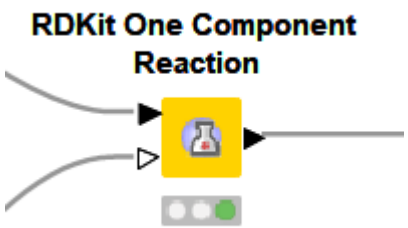
- Mapping the atoms





© 2020 KNIME AG. All rights reserved.

The nodes – RDKit One Component Reaction



Options | Flow Variables | Job Manager Selection | Memory Policy

Reaction

Reactant RDKit Mol column: mol

RDKit Rxn column: Molecule

Randomization

Randomize reactants

Maximum number of random reactions: 100

Random seed (or -1 to be ignored): -1

Other Options

Uniquify products

OK | Apply | Cancel

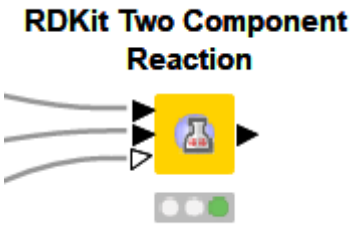
Table "Output" - Rows: 54 | Spec - Columns: 4 | Properties | Flow Variables

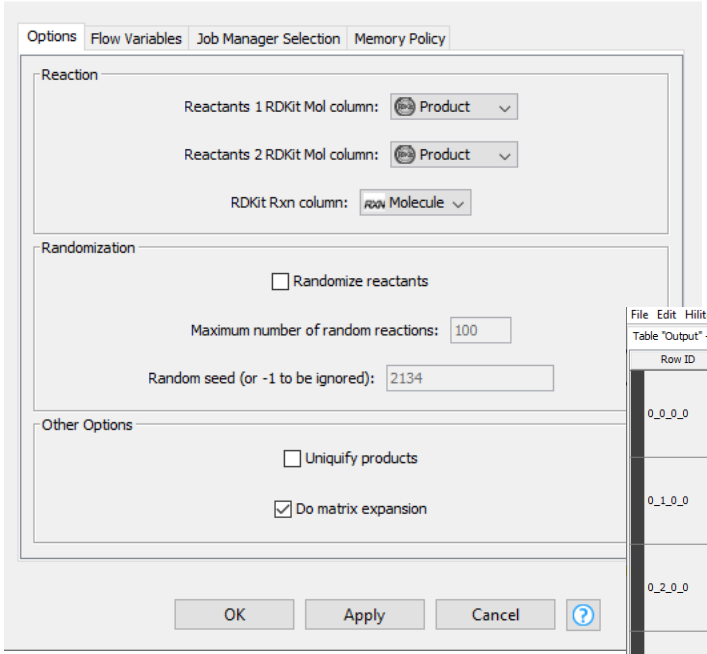
Row ID	Product	Product Index	Reactant 1 sequence index	Reactant 1
24_0_0		0	24	
24_0_1		1	24	
25_0_0		0	25	
25_0_1		1	25	

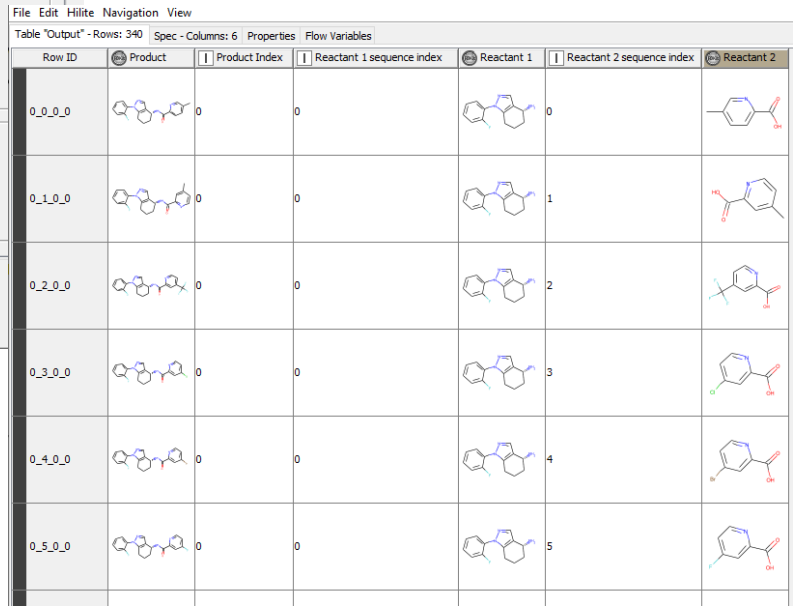
© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

The nodes – RDKit Two Component Reaction







© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

Open for Innovation
KNIME

3D Visualization of Molecules

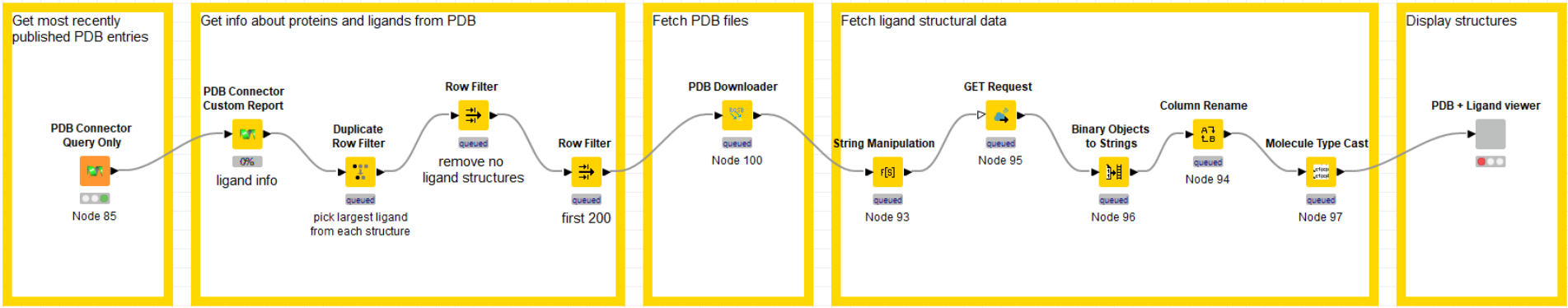
Greg Landrum

KNIME Fall Summit Webinar
20 November 2020

The use case

- Build a browser for the most recent protein–ligand crystal structures from the PDB

- The workflow:
  - Download most recent x-ray structures from the PDB that contain a co-crystallized “ligand”
  - Download the 3D structure for the ligand
  - Provide an interactive view allowing the ligands and proteins to be viewed in 3D

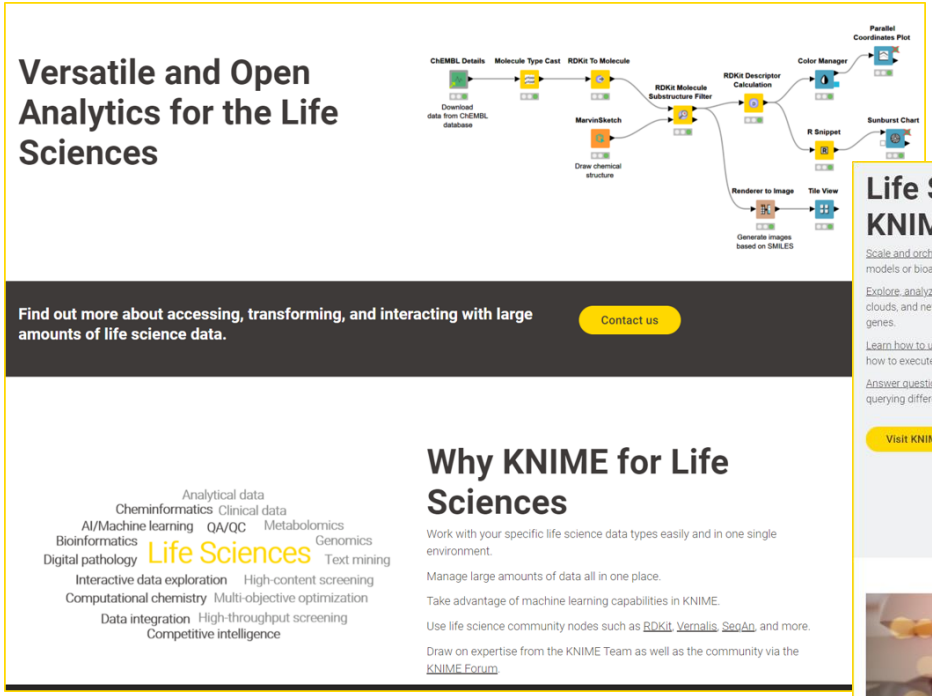


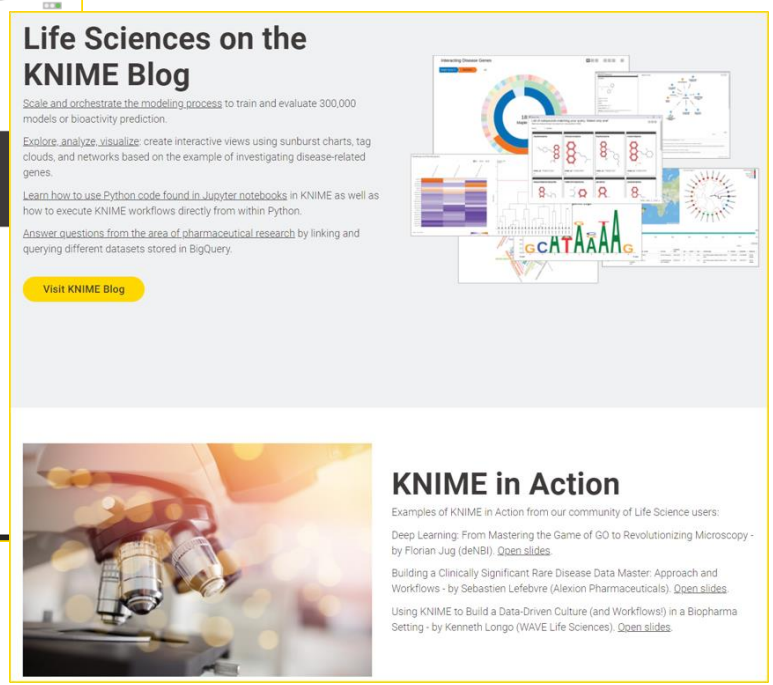
© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

Why KNIME for Life Sciences?

https://www.knime.com/why-knime-for-life-science





© 2020 KNIME AG. All rights reserved.

Additional Resources

KNIME pages (https://www.knime.com)

- RESOURCES **LEARNING HUB** https://www.knime.com/learning-hub
- RESOURCES **HUB** https://hub.knime.com/
- BOOK **WILL THEY BLEND** https://www.knime.com/knimepress/will-they-blend

KNIME Tech pages

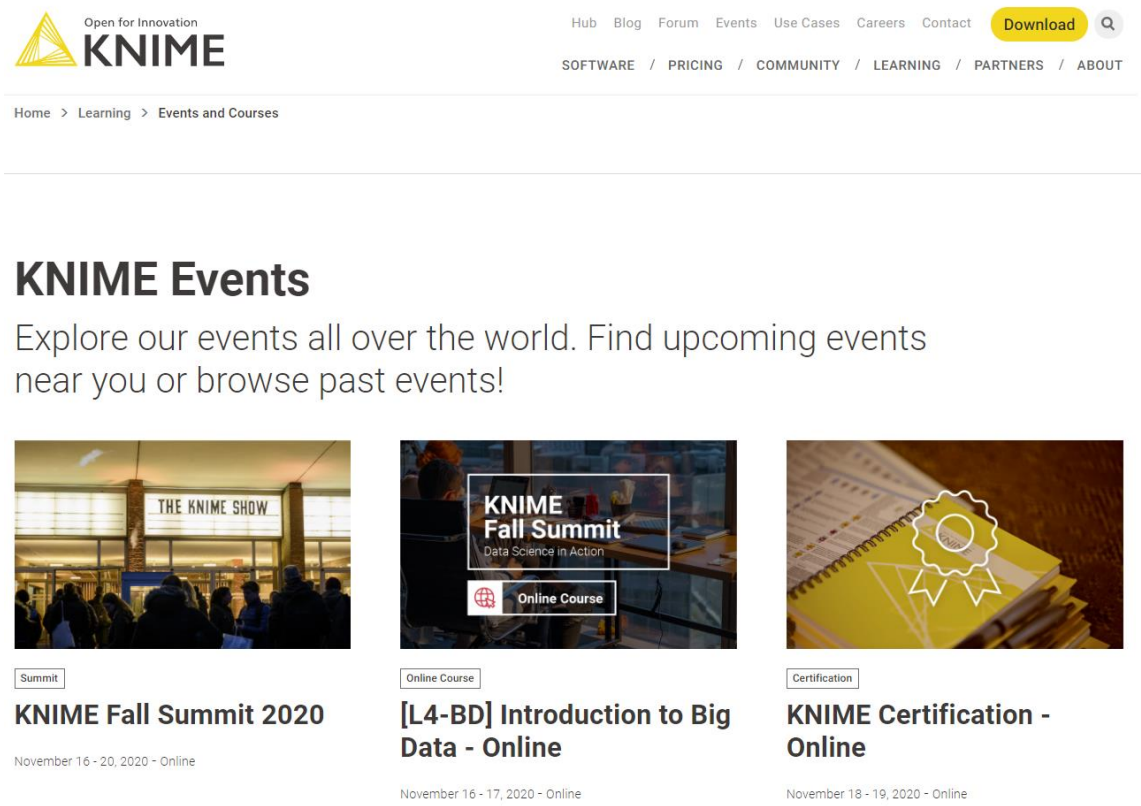
FORUM for questions and answers https://forum.knime.com

- **DOCUMENTATION** for docs, FAQ, changelogs, ... https://docs.knime.com/
- **COMMUNITY CONTRIBUTIONS** for dev instructions and third party nodes https://www.knime.com/community

KNIME TV on **YouTube** https://www.youtube.com/user/KNIMETV

© 2020 KNIME AG. All rights reserved.

KNIME courses



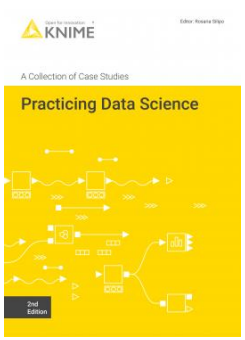
https://www.knime.com/learning/events

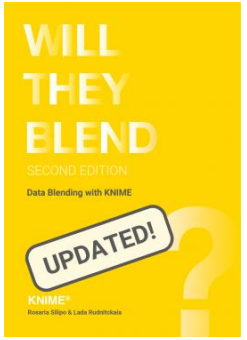
© 2020 KNIME AG. All rights reserved.

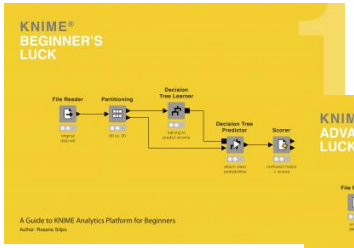
KNIME Books

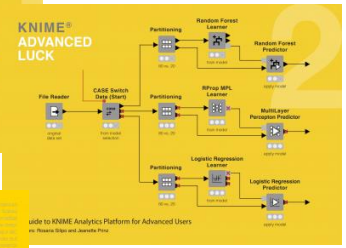
- Course books downloadable from **KNIME Press**
- https://www.knime.com/knimepress

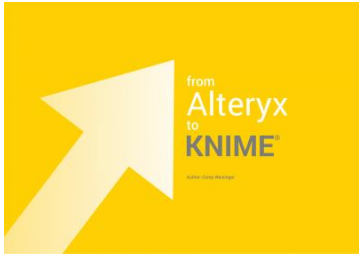
- Code: **FALL-SUMMIT-WORKSHOP**
  Valid for: All Books
  Expires: Jan 31, 2021

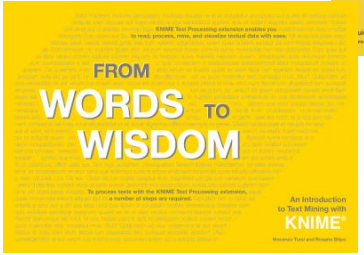












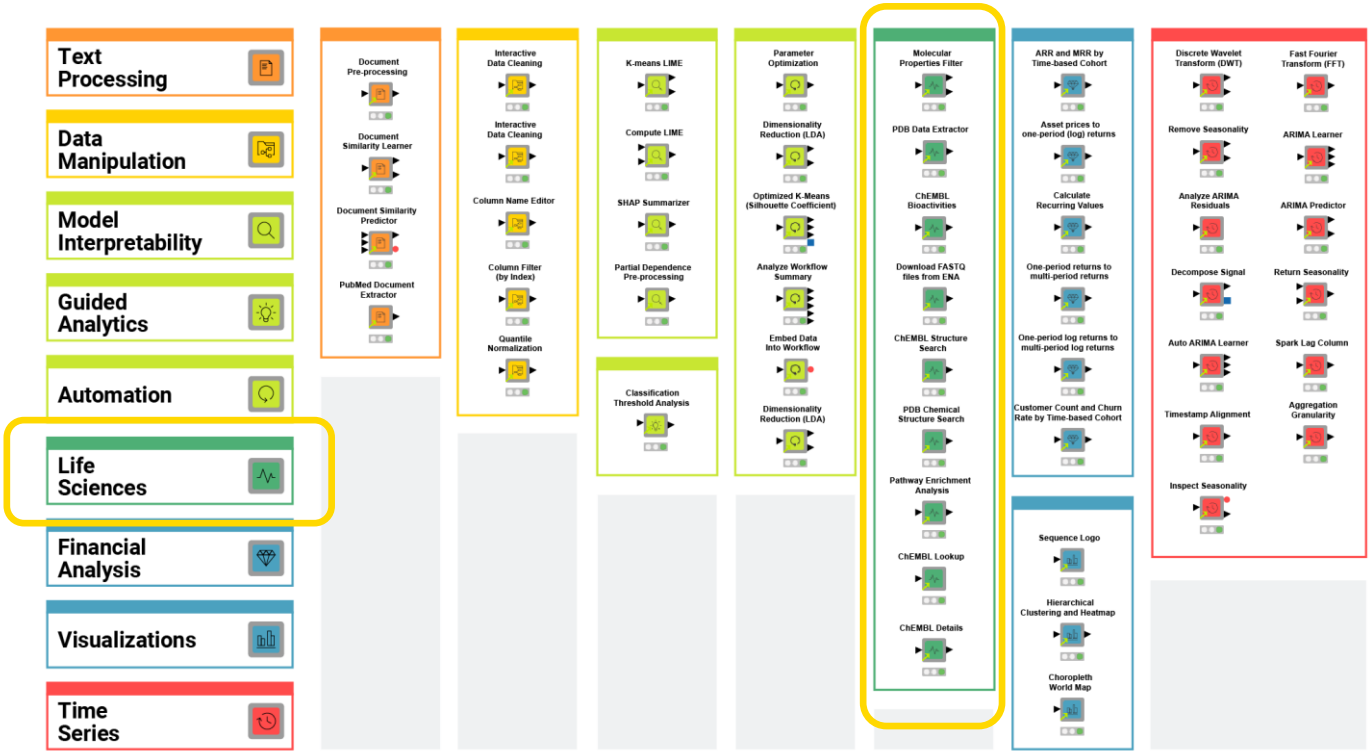
© 2020 KNIME AG. All rights reserved.

Open for Innovation
KNIME

Thank you for joining!

KNIME Verified Components

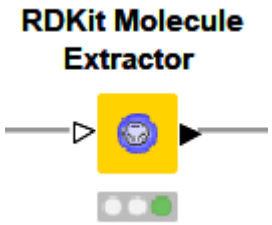
- Verified Components reuse bundled functionalities, verified by KNIME experts
- Released and updated on the KNIME Hub

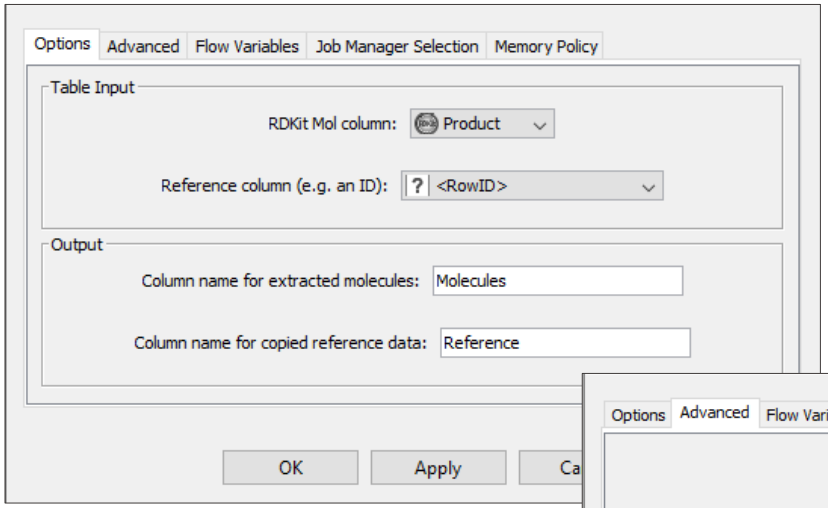


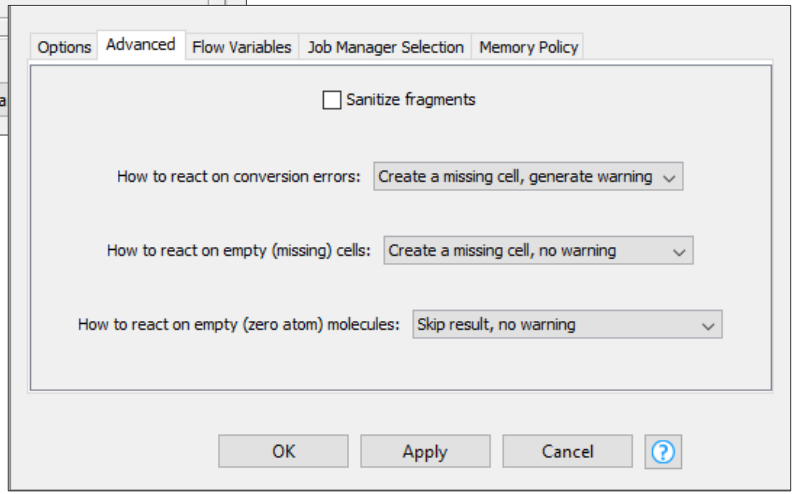
© 2020 KNIME AG. All rights reserved.

Open for Innovation KNIME

The nodes – RDKit Molecule Extractor



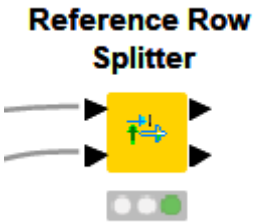


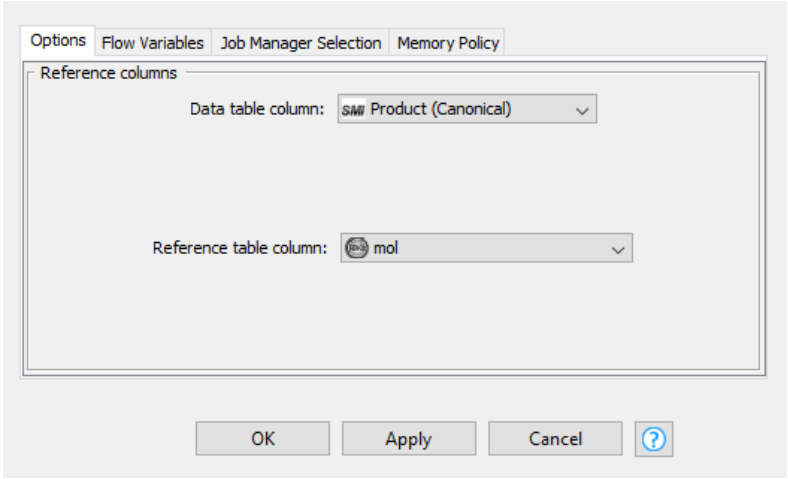


© 2020 KNIME AG. All rights reserved.

Open for Innovation
KNIME

The nodes – Reference Row Splitter





© 2020 KNIME AG. All rights reserved.

Open for Innovation
KNIME