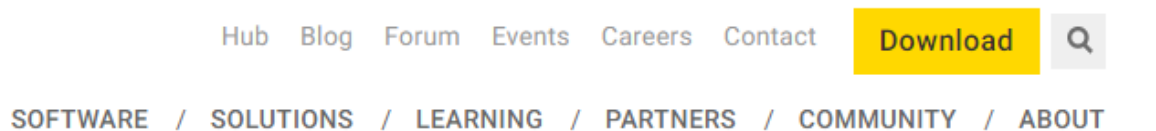


KNIME SOP

This SOP is intended as a general guide to help beginners get started with the software. It details: the installation process of KNIME and basic extensions; setting up the workspace in logical manner; how to import and use existing workflows (e.g. a bespoke PAINS filter that can be modified to suit the needs of a project); and finally general and more advanced knowledge on how to use KNIME.

Downloading the software

1. Visit the KNIME website (<https://www.knime.com/>)
2. Click the download icon in the top right corner:



- a. Fill out registration details (only email is compulsory) and then click submit

Open for Innovation
KNIME

Hub Blog Forum Events Careers Contact Download

SOFTWARE / SOLUTIONS / LEARNING / PARTNERS / COMMUNITY / ABOUT

Home

1 Register for Help & Updates 2 Download KNIME 3 Get Started

New to the KNIME family? Let us help you get started with a short series of introductory emails. These messages will get you up and running as quickly as possible and introduce you to resources that will maximize your success with the KNIME Analytics Platform.

First Name

Last Name

E-Mail ^{*}

Country

- b. Download appropriate version. For windows select the 64 bit installer:

Windows	
KNIME Analytics Platform for Windows (installer) <i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	64 Bit (441.03 MB) 32 Bit (437.42 MB)

- c. Accept and download (takes a few minutes):

Home > Downloads

You decided to download the installer for KNIME Analytics Platform for Windows (installer) 64 Bit (441.03 MB).

The installer adds an icon to the desktop and suggests suitable memory settings

If you want to run the KNIME installer or self-extracting archive for Windows you might experience some difficulty because of the Microsoft SmartScreen filter which was introduced with Internet Explorer 9 and Windows 8. [Find out how to solve the problem.](#)

I have read and accept the [privacy policy](#) and the [terms and conditions](#)

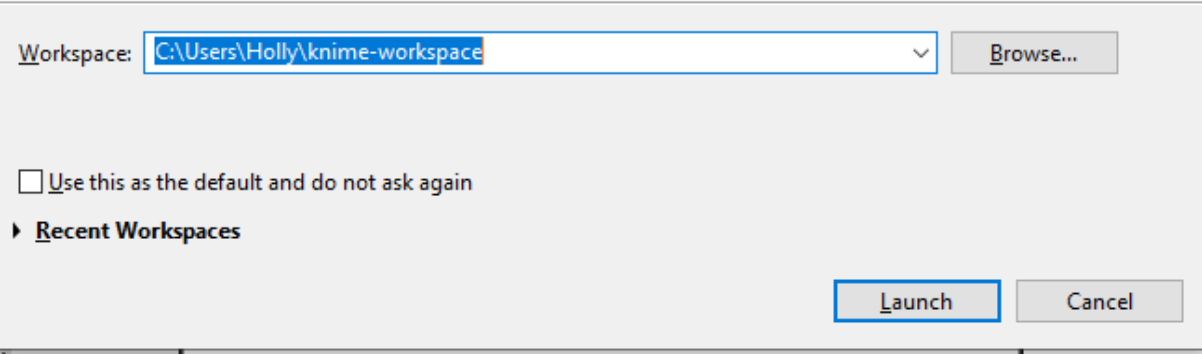
 **Download**

- d. Allow app to make changes to your device
 - e. Accept the agreement
 - f. Choose location to install KNIME (click next)
 - g. Start menu folder – either click next or choose not to have this
 - h. Additional tasks – make sure file extensions and URLs are all ticked
 - i. Proceed with suggested memory space
 - j. Install
 - k. Finish
3. Launch KNIME (either from desktop icon if you selected that or type 'KNIME' in the computer search bar)
 - a. The platform launcher should pop up, and click launch:

KNIME Analytics Platform Launcher ✕

Select a directory as workspace

KNIME Analytics Platform uses the workspace directory to store its preferences and development artifacts.



Workspace:

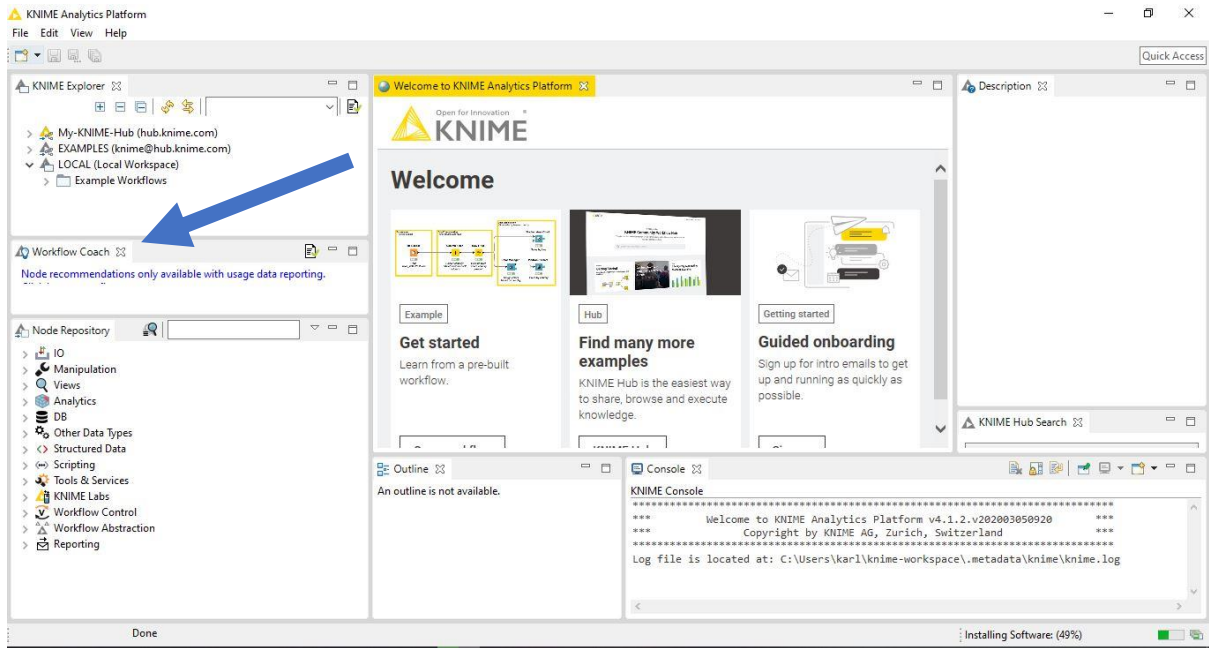
Use this as the default and do not ask again

▶ Recent Workspaces

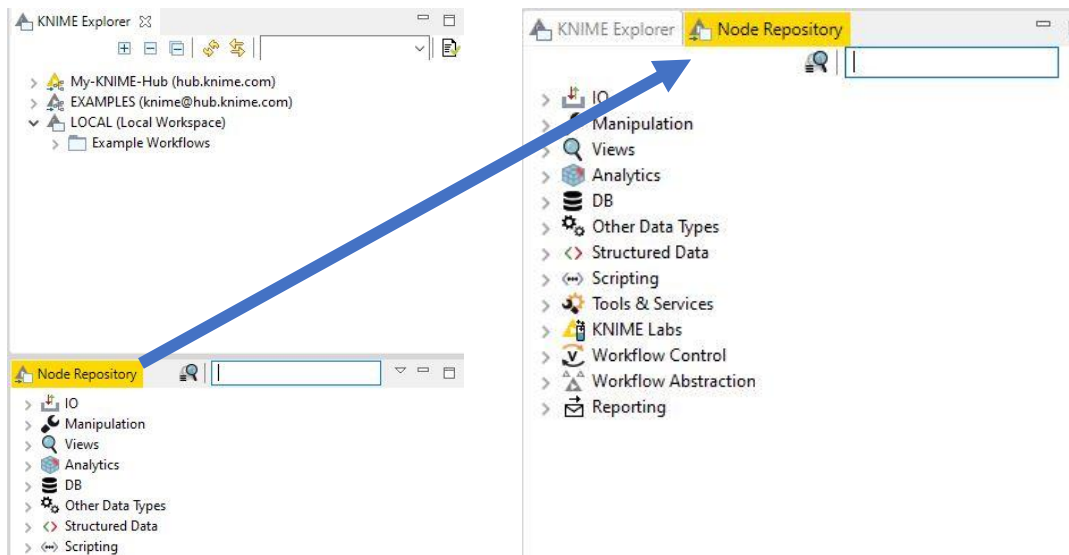
Tidying the workspace for efficient working

Once set up these modifications should be retained every time KNIME is launched. This is not the only way to configure the panel set up, but I find it most useful as gives the most space for the workflow.

1. Close workflow coach (use the cross):

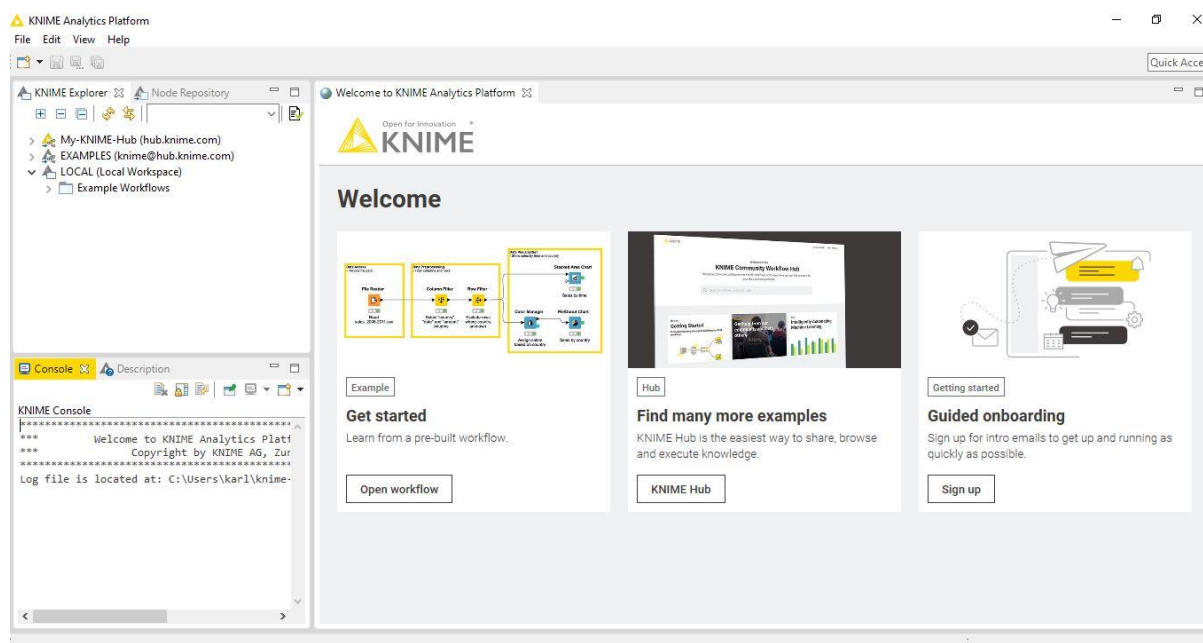


2. Drag Node Repository next to KNIME explorer:



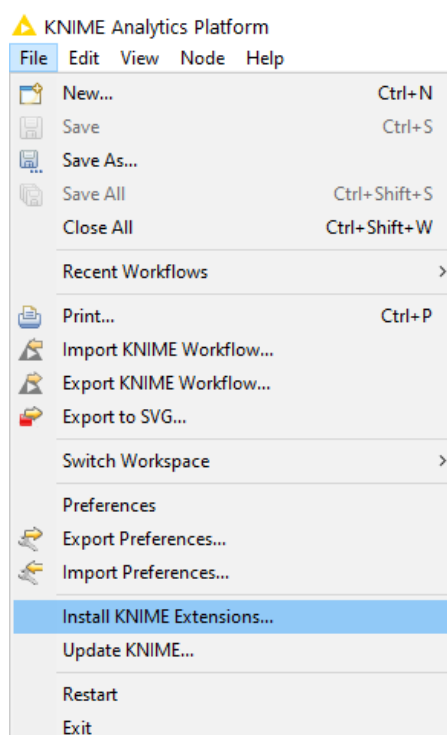
3. Close KNIME hub search
4. Close Outline
5. Drag 'Console' and 'Description' to bottom left of the screen

Window should look like this upon completion:

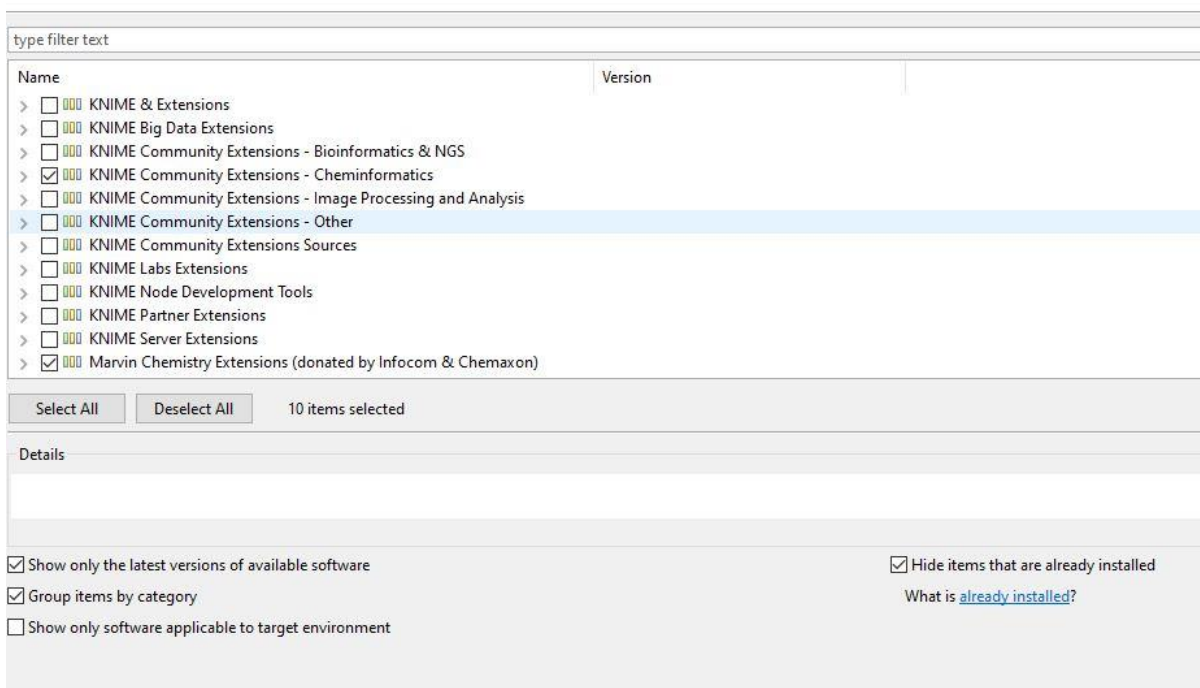


Installing chemistry extensions

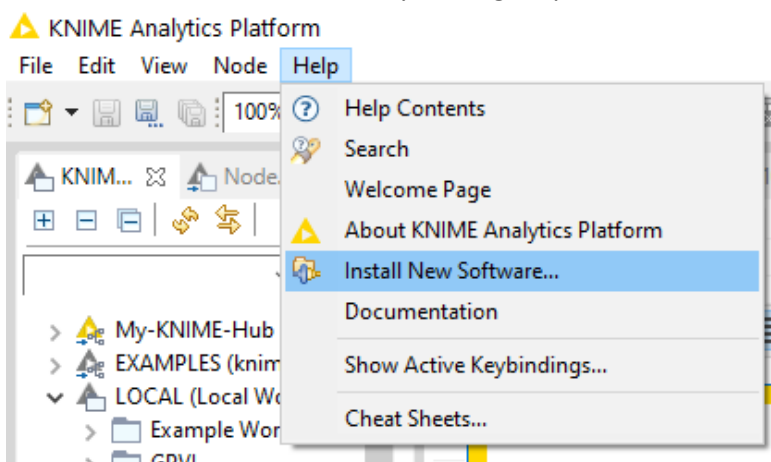
1. File > Install KNIME Extensions



2. Select extensions of interest using the tick boxes. The cheminformatics and Marvin chemistry extensions are recommended as a minimum:



3. Click next and then next, accept agreements, finish
4. Close KNIME and restart for changes to take effect
5. Other extensions can be added by clicking Help > Install New Software:



6. Type in web address for new extension in the 'work with' section then click 'add'

Install

Available Software
Select a site or enter the location of a site.

Work with: Add... Manage...

type filter text

Name	Version
<input type="checkbox"/> ⓘ There is no site selected.	

Select All Deselect All

Details

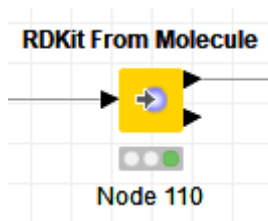
Show only the latest versions of available software
 Group items by category
 Show only software applicable to target environment
 Contact all update sites during install to find required software

Hide items that are already installed
[What is already installed?](#)

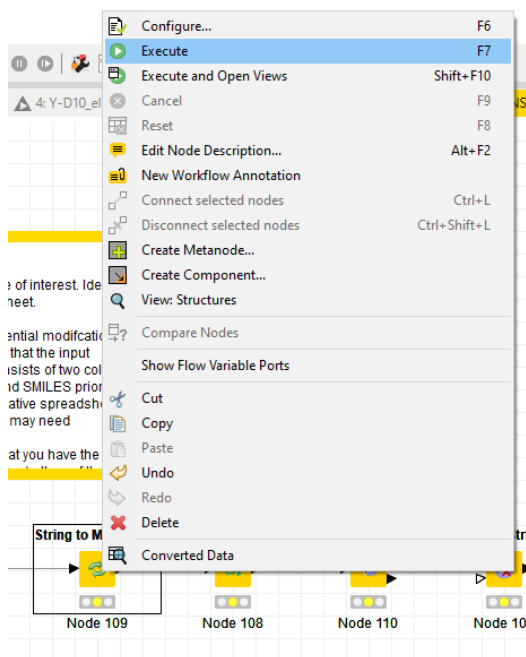
< Back Next > Finish Cancel

Basic KNIME knowledge

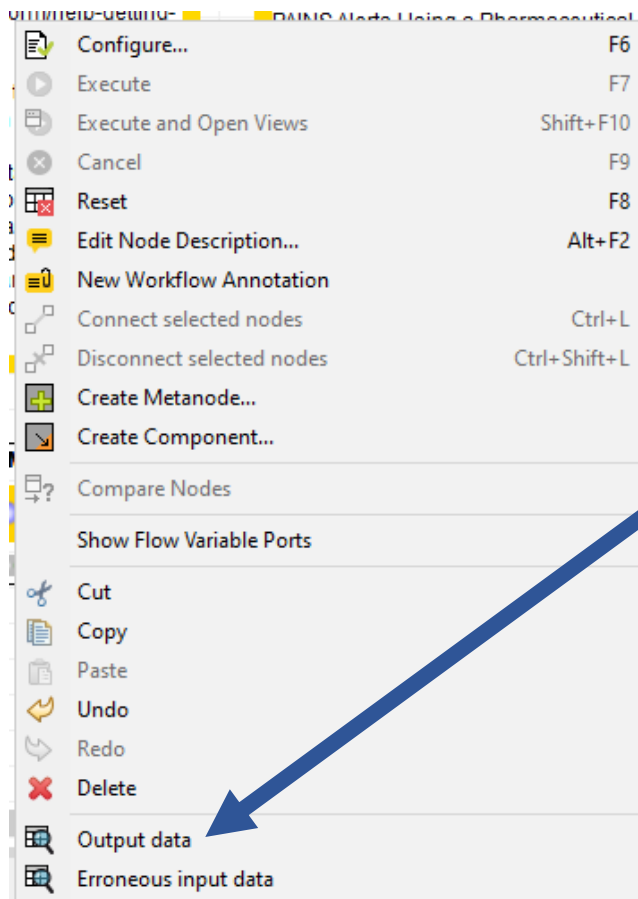
- Headed arrows show the data inputs/outputs from nodes:



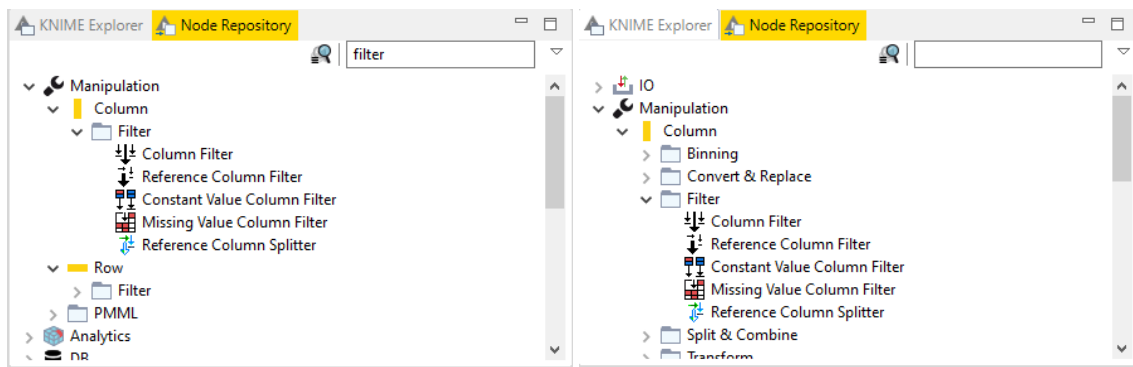
- Running nodes: right click > execute or execute and open views to see the output of this node. Alternatively select then hold fn and f7.



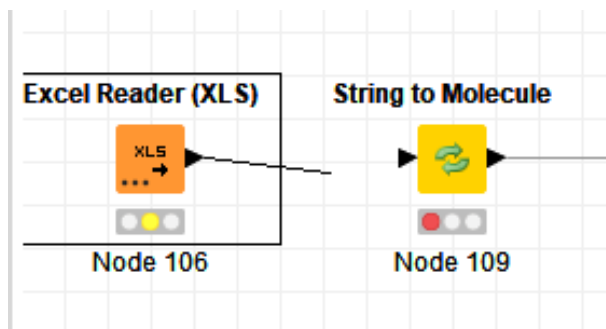
- Viewing node outputs: right click and then select the bottom option that has a spreadsheet and magnifying glass



- Searching for nodes: Navigate to the node repository and type a key word in the search bar (e.g. filter), or navigate manually (e.g. to find column filters: manipulation > column > filter)






- Connecting nodes: Drag inputs/outputs from one node to another

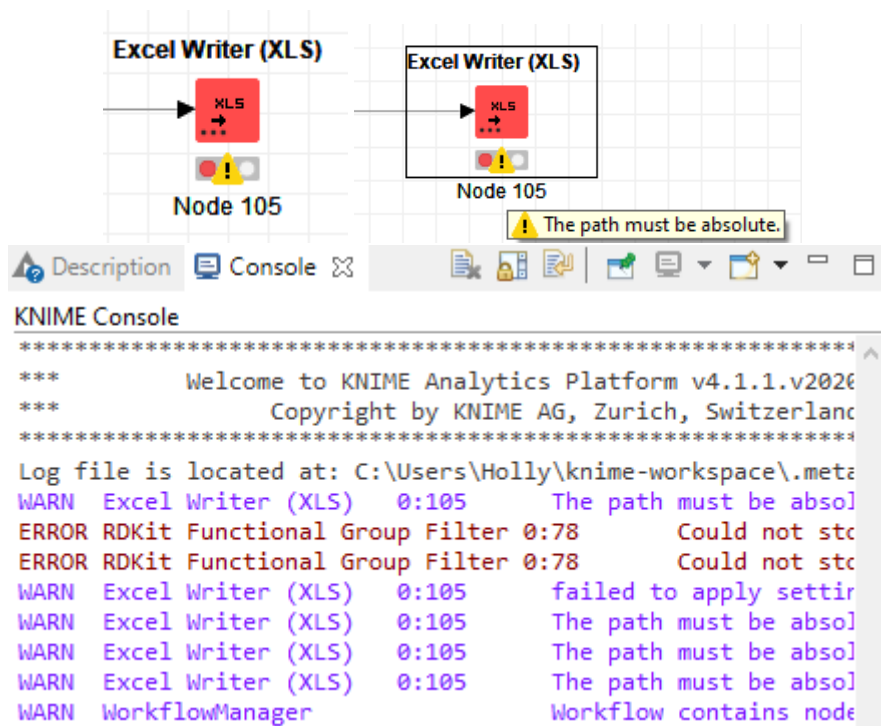


- Node traffic lights:

Understanding the traffic light system:

-  Not configured: Node is not yet configured and cannot be executed with its current settings
-  Configured: Node has been correctly configured and may be executed at any time
-  Executed: Node has been successfully executed and results can be viewed and used in downstream nodes

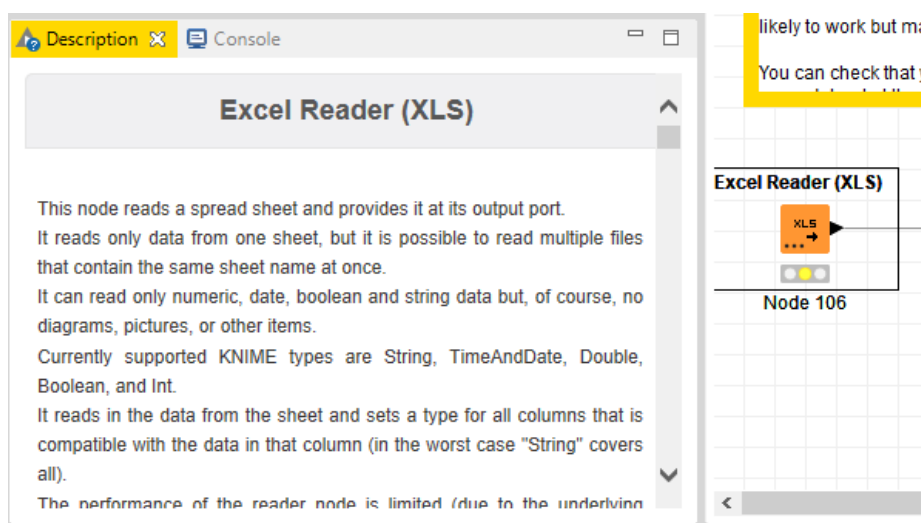
- Node errors are denoted with a warning triangle. If you hover over this it will detail the issue. The console may offer further description of the problem.



The screenshot shows two 'Excel Writer (XLS)' nodes in a workflow, both labeled 'Node 105'. Each node has a red traffic light icon and a yellow warning triangle. A tooltip for the right node reads 'The path must be absolute.' Below the nodes is the 'KNIME Console' window, which displays the following log entries:

```
*****  
***      Welcome to KNIME Analytics Platform v4.1.1.v2020  
***      Copyright by KNIME AG, Zurich, Switzerland  
*****  
Log file is located at: C:\Users\Holly\knime-workspace\.meta  
WARN Excel Writer (XLS)  0:105      The path must be absol  
ERROR RDKit Functional Group Filter 0:78      Could not stc  
ERROR RDKit Functional Group Filter 0:78      Could not stc  
WARN Excel Writer (XLS)  0:105      failed to apply settir  
WARN Excel Writer (XLS)  0:105      The path must be absol  
WARN Excel Writer (XLS)  0:105      The path must be absol  
WARN Excel Writer (XLS)  0:105      The path must be absol  
WARN WorkflowManager      Workflow contains node
```

- Node descriptions: Click on a node and select the 'Description' tab to understand the function of the node:



The screenshot shows the 'Description' tab for an 'Excel Reader (XLS)' node, labeled 'Node 106'. The description text reads:

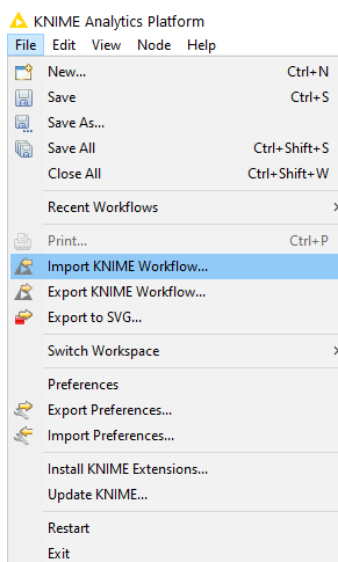
This node reads a spreadsheet and provides it at its output port. It reads only data from one sheet, but it is possible to read multiple files that contain the same sheet name at once. It can read only numeric, date, boolean and string data but, of course, no diagrams, pictures, or other items. Currently supported KNIME types are String, TimeAndDate, Double, Boolean, and Int. It reads in the data from the sheet and sets a type for all columns that is compatible with the data in that column (in the worst case "String" covers all). The performance of the reader node is limited (due to the underlyin

Importing existing workflows

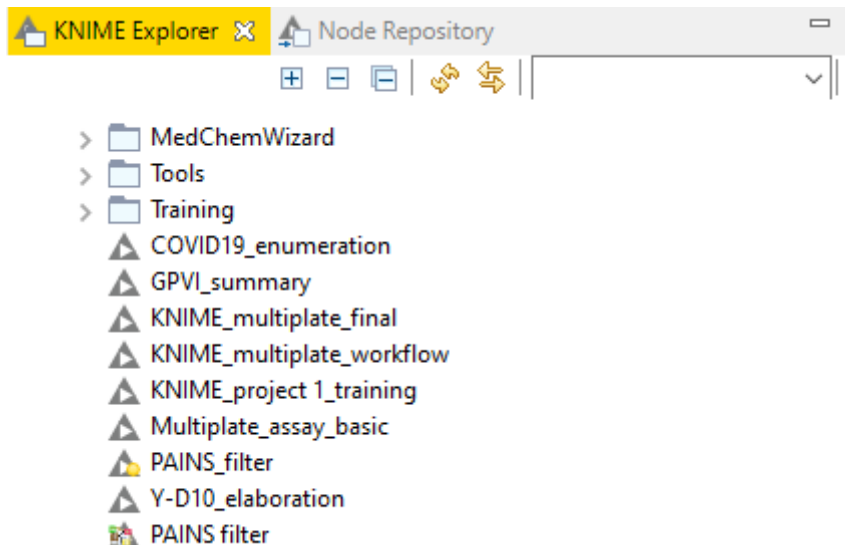
1. File > Import KNIME workflow
2. Browse to select workflow of interest > open > finish
3. Double click workflow on KNIME explorer

Using the PAINS filter

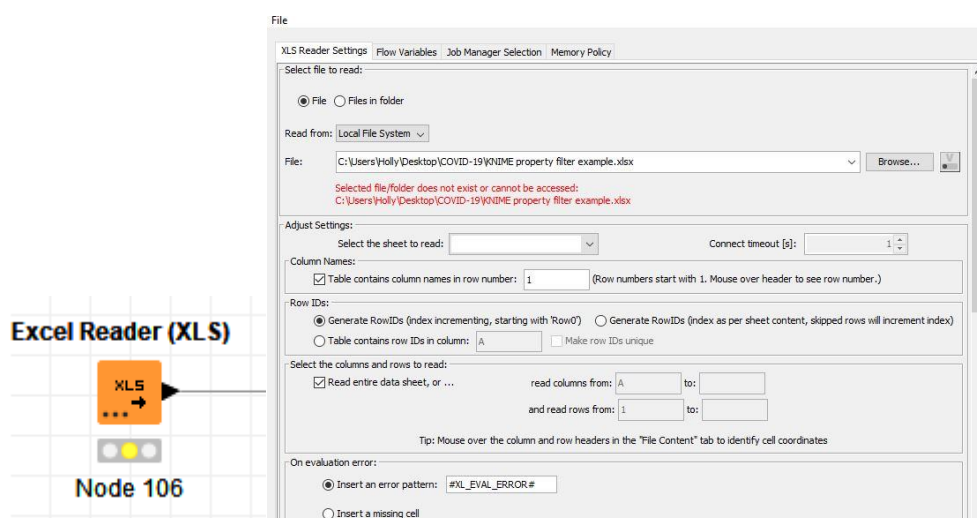
1. Import workflow:



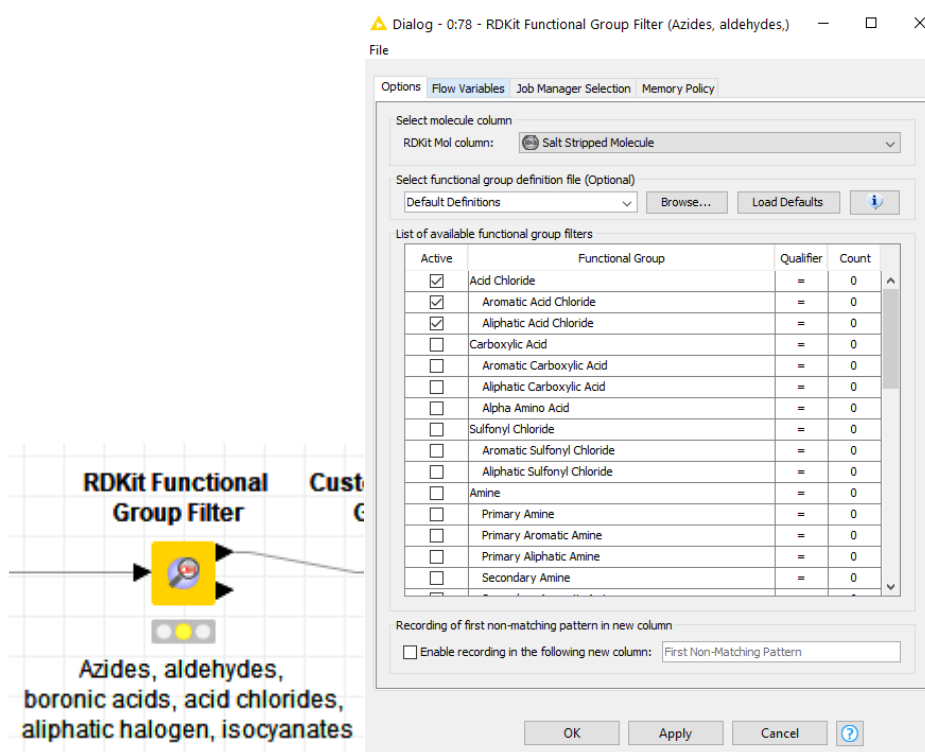
2. Double click workflow link in the 'KNIME explorer':



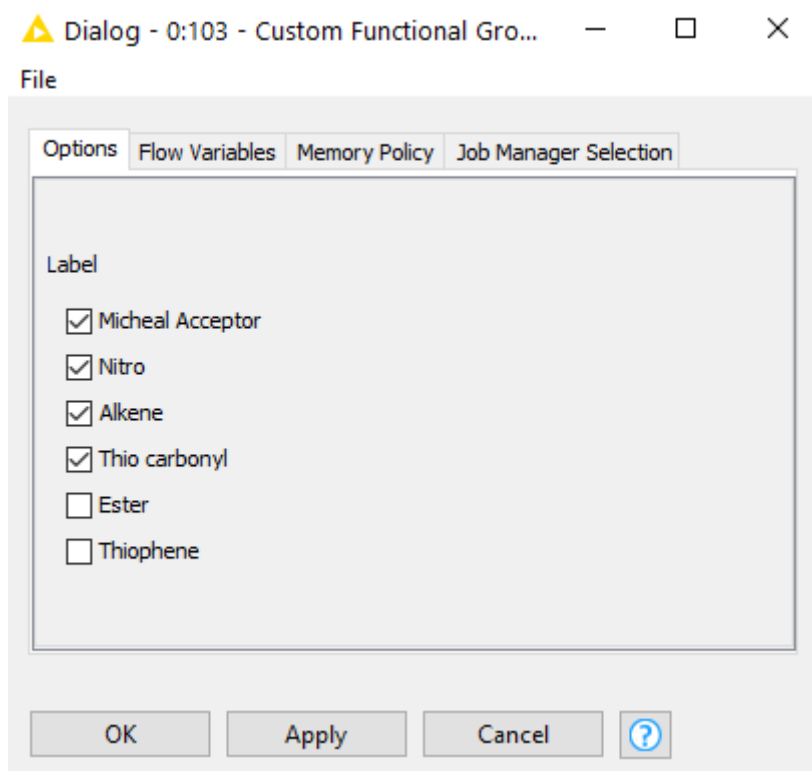
3. Double click on Excel reader node:



- a. 'Browse' to the spreadsheet of interest
 - b. If you scroll down you can see a preview of what the spreadsheet will look like. Make sure this looks reasonable. Ideally the input spreadsheet should have one sheet consisting of the compound ID and SMILES.
4. Double click on the 'RDKit functional group filter' and select the functional groups to REMOVE then click okay. The pre-selected functional groups are detailed below the node.



5. Double click the 'custom functional group filter' and select other functional groups that you wish to remove then click okay:



6. Double click on the property filter and set the parameters then click okay:

Dialog - 0:100 - Property Filter

File

Options Flow Variables Memory Policy Job Manager Selection

LogP Max Change

Maximum Molecular Weight

Maximum Number of Heavy Atoms

Max TPSA Change

Maximum Number of Rotatable Bonds

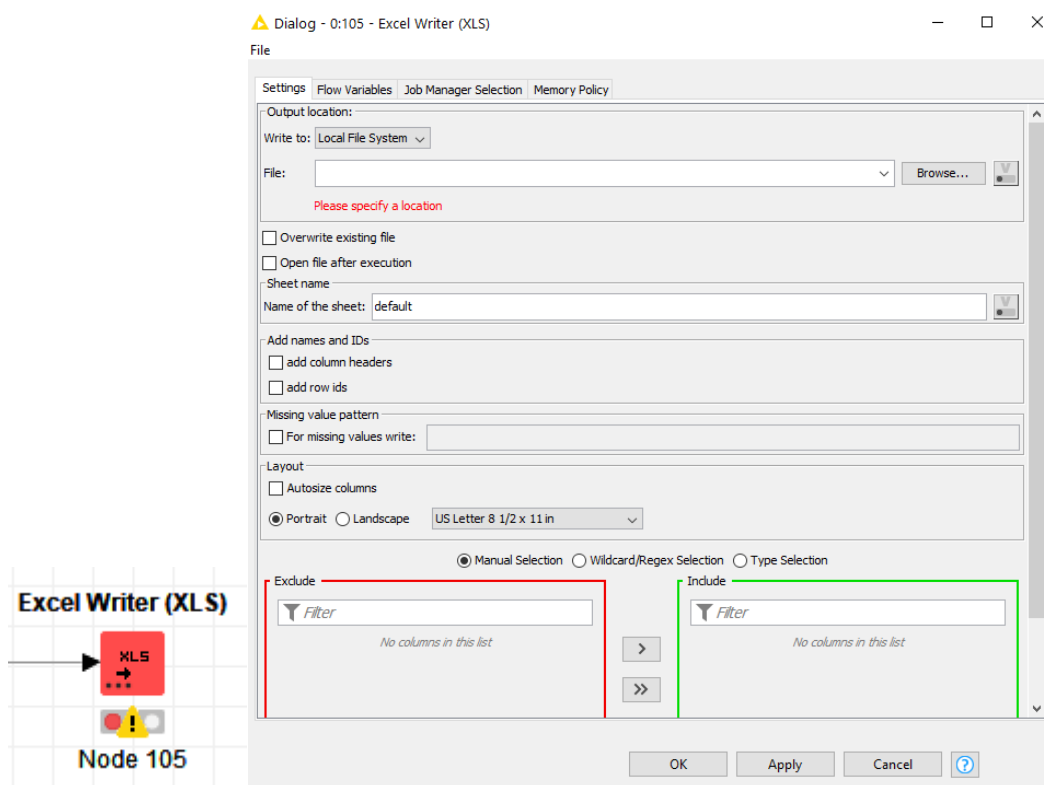
Maximum Number of H Bond Donors

Maximum Number of H Bond Acceptors

Max Fraction Sp3 Change

OK Apply Cancel ?

7. Right click on the property filter and click 'execute'
 - a. Alternatively can select the node and then press fn > f7
 - b. The traffic light underneath the node should now be green
8. Double click on the Excel writer, and specify the save location, click okay then run this node



Advanced KNIME Knowledge

Metanodes

Components (GUIs)

Flow Variables

Useful Chemistry KNIME Knowledge

Glossary

Node	A node is the smallest programming unit in KNIME. Each node serves a dedicated task, from very simple ones - like changing the name of a data column - to very complex one - like training a random forest.

Training Resources

KNIME offers a number of **FREE** online training courses: <https://www.knime.com/knime-introductory-course>

KNIME also has three short one page 'cheat sheets' which are VERY helpful for new users: <https://www.knime.com/learning/cheatsheets> . See below for a sneak preview:

Cheat Sheet: Building a KNIME Workflow for Beginners



Getting started with KNIME Analytics Platform

- Read through the installation guide at [knime.com/installation](https://www.knime.com/installation)
- Check out the 7 things you should do after installing KNIME Analytics Platform at [knime.com/blog/seven-things](https://www.knime.com/blog/seven-things)
- Take the E-Learning Course at [knime.com/knime-introductory-course](https://www.knime.com/knime-introductory-course)
- Browse workflows, nodes, and components at hub.knime.com

Understanding the traffic light system:

- Not configured: Node is not yet configured and cannot be executed with its current settings
- Configured: Node has been correctly configured and may be executed at any time
- Executed: Node has been successfully executed and results can be viewed and used in downstream nodes

EXPLORE

Scatter Plot: Represents input data rows as points in a two dimensional plot. Input dimensions (columns) on the x-y axis plot and graphical properties can be changed in the configuration window or interactively in the node view.

Line Plot: Plots numerical values in data columns (y-axis) against values in a reference column (x-axis). Data points are connected via colored lines. If the reference column on the x-axis contains sorted time values, the line plot graphically represents the evolution of a time series.

Data Explorer: Provides an interactive view to summarize the statistics of the input data via statistical measures and histograms - for both numerical and nominal columns.

Sunburst Chart: Displays categorical columns through a hierarchy of rings. Each ring is sliced according to the nominal values in the corresponding column and to the selected hierarchy. This is a powerful chart for multivariate analysis.

Color Manager: Assigns a color property to each input row based on the row's value in a selected column. This color property affects the graphical representation in the upcoming views.

Box Plot: Visualizes numeric columns using the quartile statistics. Watch out for the points at the end of the whiskers - they might mark outliers!

Stacked Area Chart: Plots multiple numerical data columns on top of each other using the previous line as the base reference. The areas in between lines are colored for easier comparison. This chart is commonly used to visualize trending topics.

Pie Chart: Visualizes one aggregated metric for different data partitions with colored slices on a circle where the areas are proportional to the metric values. The partitions are defined by a categorical column.

Bar Chart: Visualizes one or more aggregated metrics for different data partitions with rectangular bars where the heights are proportional to the metric values. The partitions are defined by a categorical column.

ANALYZE

Decision Tree: The Learner node trains a C4.5 decision tree. The configuration window includes options for pruning, early stopping, information measures, splitting values, and more. Both the Learner and the Predictor node provide an interactive view where the decision tree is displayed together with the input data propagation.

k-Means: Implements the k-Means clustering algorithm. Number of clusters must be set prior to node execution. This node builds the clusters. The Cluster Assigner node finds the closest cluster and assigns it to the input data row. Being an unsupervised algorithm, this node pair doesn't follow the classic Learner - Predictor scheme.

Logistic Regression: The Learner node trains a logistic regression model to predict categorical target values. The configuration window includes options for solver, input feature choice, regularization functions to avoid overfitting, & more.

Scorer: Calculates a number of performance measures such as accuracy, F1 score, or Cohen's Kappa, to quantify the quality of a classifier.

Numerical Scorer: Calculates a number of numerical error measures, such as root mean squared error, mean absolute error, or R², to quantify the quality of a numerical predictor model.

ROC Curve: Displays the Receiver Operating Characteristic (ROC) curve of a classifier working on a binary class problem. One of the two classes is arbitrarily chosen as the positive class and the ROC curve is built on the probabilities/scores produced for that class on the input data set.

Integrations to many open source data analytics tools are also available. Some use the KNIME node GUI (W2O, Weka, Keras, Spark MLlib). Others offer nodes with a...

READ

File Reader: Reads all text files, particularly character separated files, such as CSV files. The File Reader is the workhorse for reading text data.

Excel Reader (XLS): Reads content from sheets in Excel files (XLS, XLSX). Sheet and cells to be read can be defined in the configuration window.

Table Creator: Allows users to manually create a data table in its configuration window as a data sheet. Data cells can be copied and pasted in the sheet. Perfect for generating small data sets.

Table Reader: Reads data from a table file. Table files are organized using a KNIME proprietary format, including the full file structure and are optimized for space and speed - providing maximum performance with minimum configuration!

Google Sheets Reader: Reads data from a Google Sheet file. Authentication occurs on the Google site. Google credentials are not saved within the KNIME workflow.

knime:// protocol: References a file path relative to some key location of the current KNIME installation like `knime://knime.workflow/`, `~/knime` or `knime://knime-server.mountpoint/` `path/to/filename`

Workflow Diagram: Read (File Reader) → Transform (Data Explorer) → Analyze (Learner/Predictor) → Deploy (Scorer/ROC Curve)

Legend:
● Learner Nodes: Supervised algorithms in KNIME Analytics Platform have a Learner node to train a model on a previously labelled training set.
● Predictor Nodes: Used for applying models. The two inputs are the trained model and the data to process. The output contains the original data and the model predictions.

Version History