KNIME Quickstart Guide

Table of Contents

License

Starting with Version 2.1, KNIME is released under the [GNU General Public License](#), Version 3 (including certain additional permissions according to Sec. 7 of the GPL). It is also available – through the dual licensing model – under customized licenses. If you wish to receive KNIME under a different license than the GPL, please [contact us](#) to discuss licensing arrangements.

Versions

	Linux	Windows	Mac
KNIME (32bit)	✓	✓	
KNIME (64bit)	✓	✓	✓
KNIME Developer Version (32bit)	✓	✓	
KNIME Developer Version (64bit)	✓	✓	✓

Installation

Download one of the above versions, unzip it to any directory for which you have write permissions. On Windows click the *knime.exe* file, on Linux the *knime and on Mac the knime.app* in order to start KNIME.

Welcome Screen / Additional features

When KNIME is started the first time a welcome screen opens. From here you can

1) **Open KNIME workbench**: opens the KNIME workbench to immediately start exploring KNIME, build own workflows and explore your data.

2) **Get additional nodes**: In addition to the ready-to-start basic KNIME installation there are additional plug-ins for KNIME e.g. an R and Weka integration, or the integration of the Chemistry Development Kit with additional nodes for the processing of chemical structures, compounds, etc. You can download these features also later from within KNIME (File, Update KNIME...)

Workbench overview

The KNIME Workbench is organized as follows:

KNIME

File Edit View Node Search Run Help

100%

Workflow Projects

KNIME_project

Workflow Projects

Each workflow refers to a workflow project. All projects are displayed here. Import and export of workflows is supported. Status (closed, idle, executing and executed) is indicated by an icon.

Favorite Nodes

♥ Personal favorite nodes
☆ Most frequently used nodes
🕐 Last used nodes

Favorite Nodes

Manage your favorite, most frequently and last used nodes. Nodes are added to favorites by dragging them from Node Repository into favorite nodes category.

Node Repository

⊞ IO
⊞ Database
⊞ Data Manipulation
⊞ Data Views
⊞ Statistics
⊞ Mining
⊞ Meta

Node Repository

Find all KNIME nodes here, ordered by categories. Help for selected nodes is displayed in the Node Description. Drag them onto the editor in order to add them to the workflow.

2: KNIME_project

Workflow Editor

Here the workflows are assembled by dragging nodes onto this editor, connecting, configuring and executing them.

Node Description

Node Description

Provides help about the selected node, its dialog options, views, expected input data, and resulting output.

Outline

Outline

Overview over the workflow and navigation help for large workflows.

Console

KNIME Console

Console

Status information, warning and error messages are logged here. This information is also written to a log file. Change level of detail under Preferences/KNIME/KNIME GUI.

Building a workflow

A workflow is built by dragging nodes from the *Node Repository* onto the *Workflow Editor* and connecting them. Nodes are the basic processing units of a workflow. Each node has a number of input- and/or output ports. Data (or a model) is transferred over a connection from an out-port to the in-port of another node.

Node Status

When a node is dragged onto the workflow editor the status light shows red, which means that the node has to be configured in order to be able to be executed. A node is configured by right clicking it, choosing „Configure", and adjusting the necessary settings in the node's dialog.
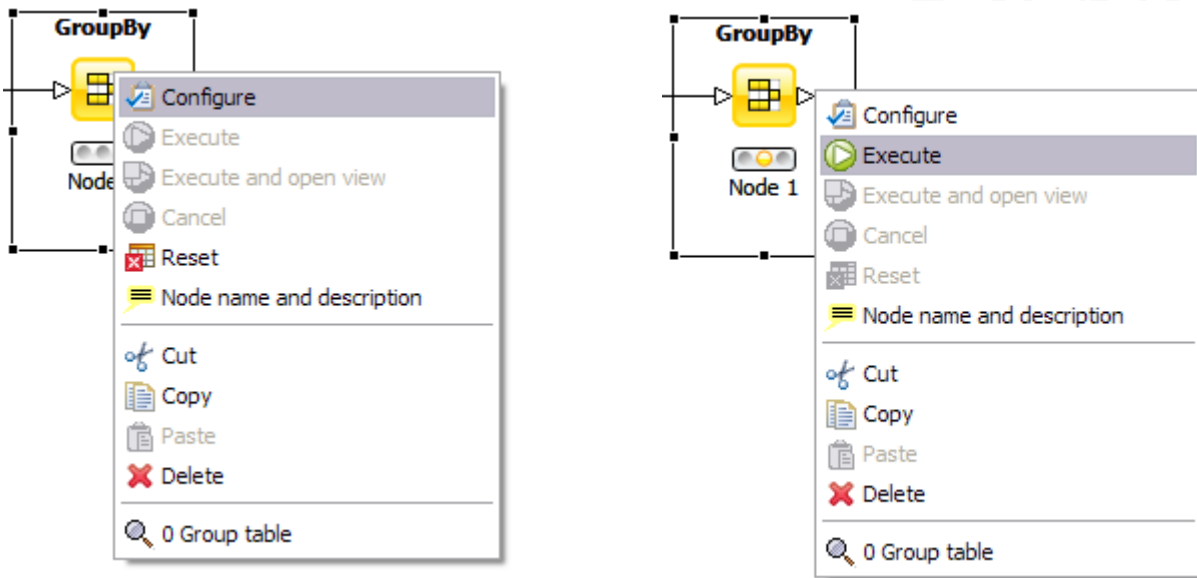
When the dialog is closed by pressing the „OK" button, the node is configured and the status light changes to yellow: the node is ready to be executed. Right-click on the node again shows an enabled „Execute" option; pressing it will execute the node and the result of this node will be available at the out-port. After a successful execution the stauts light of the node is green.

The result(s) can be inspected by exploring the out-port view(s): the last entries in the context menu open them.

Ports

Ports on the left are input ports, where the data from the out-port of the predecessor node is provided. Ports on the right are out-ports. The result of the node's operation on the data is provided at the out-port to successor nodes. A tooltip gives information about the output of the node, further information can be found in the node description.
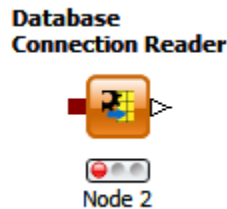
Nodes are typed such that only ports of the same type can be connected.

Data Port: The most common type is the data port (a white triangle) which transfers flat data tables from node to node.
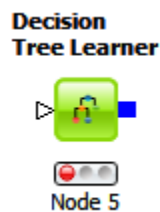

**Database Port:** Nodes executing commands inside a database can be recognized by their

database ports (brown square):

Database Connection Reader

Node 2

PMML Ports: Data Mining nodes learn a model which is passed to the referring predictor node via a blue squared PMML port.

Decision Tree Learner

Node 5

Other Ports: Whenever a node provides data which does not fit a flat data table structure a general purpose port for structured data is used (dark cyan square). Port s which are neither data, database, PMML, or ports for structured data are displayed as „unknown" types (gray square).

Color Manager   **R Learner (Local)**

Node 6   Node 7
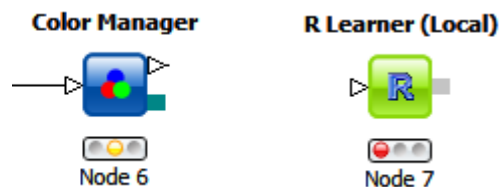
Uninstalling

KNIME is uninstalled from your system by simply deleting the installation directory. Per default the workspace is also in this directory. If you have chosen a different location for the workspace be sure to delete this directory as well.

Example Flow

We now want to take you step-by-step through the process of building a small, simple workflow: we read in data from an ASCII file, assign color to it, cluster the data and display the data in a table and a scatter plot. After we execute this flow we will examine the data

model that has been built. We assume you have just started KNIME with an empty workflow.

Adding Nodes

In the Node Repository expand the "IO" and the contained "Read" category as depicted below (left picture) and drag&drop the *File Reader* icon into the *Workflow Editor* window.
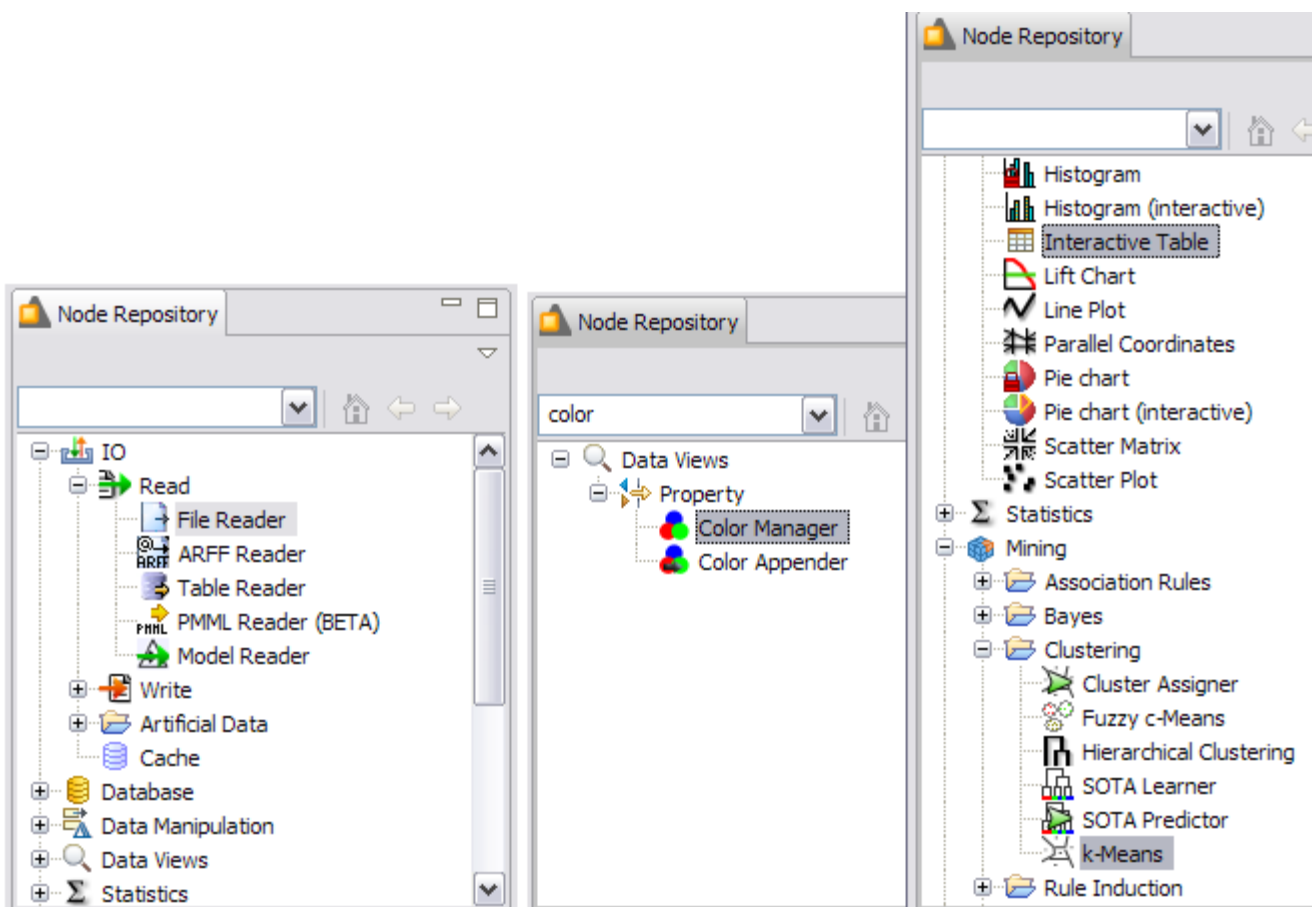

The next node for now will be the *K-Means* clustering algorithm. Expand the *Mining* category followed by the *Clustering* category, and then drag the *K-Means* node to the flow (picture on the right).

In the search box of the *Node Repository* enter "color" and press "Enter". This limits the nodes shown to the ones with "color" in their name (see figure above in the middle). Pull the *Color Manger* node onto the workflow (this node will define the color in the data views later). To see again all nodes in the repository, press ESC or Backspace in the search field of the *Node Repository*. Now, drag the *Interactive Table* and the *Scatter Plot* from the

Data Views category to the *Workflow Editor* and position it to the right of the *Color Manager* node.

Connecting Nodes

Now you need to connect the nodes in order to get the data flowing. Click an output port and drag the connection to an appropriate input port. Complete the flow as pictured below:
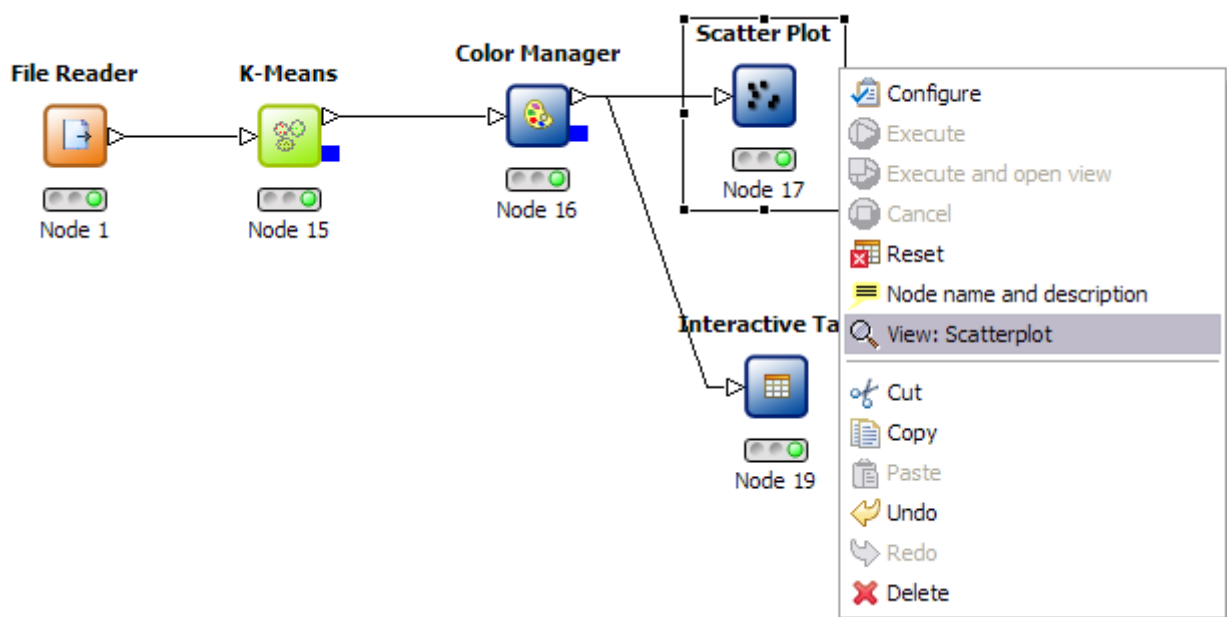

Of course, your nodes will not show a green status, as long as they are not configured and executed.

Configuring Nodes

Fully connected nodes showing a red status icon need to be configured. Start with the *File Reader*, right-click it and select "Configure" from the menu. Navigate to the "IrisDataSet" directory located in the KNIME installation directory. Select the data.all file from this location. The *File Reader*'s preview table shows a sample of the data.
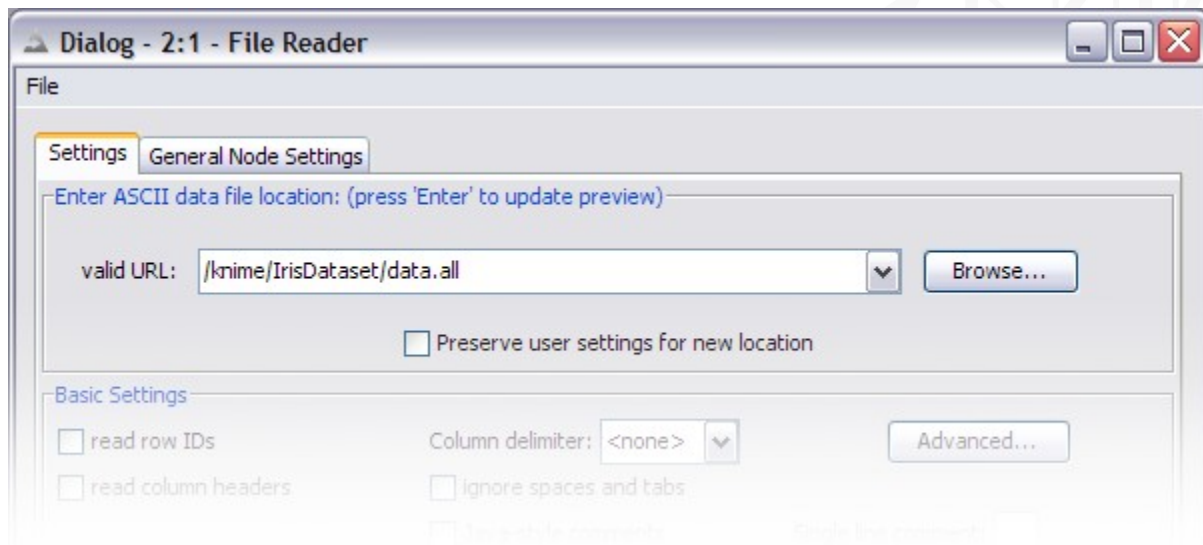
Press OK to close the dialog of the *File Reader* node. Once the node has been configured correctly, it switches to yellow (meaning ready for execution). After that, the *K-Means* node will immediately turn yellow, since its default settings will be applied. To be sure, that the default settings fit your needs, open the dialog and inspect the default settings.

In order to configure the *Color Manager* node you must first execute the *K-Means* node. After execution all nominal values and ranges of all attributes are known: this meta information is propagated to the successor nodes. The *Color Manager* needs this data before it can be configured. Once the *K-Means* node is executed, open the configuration dialog of the *Color Manger* node.
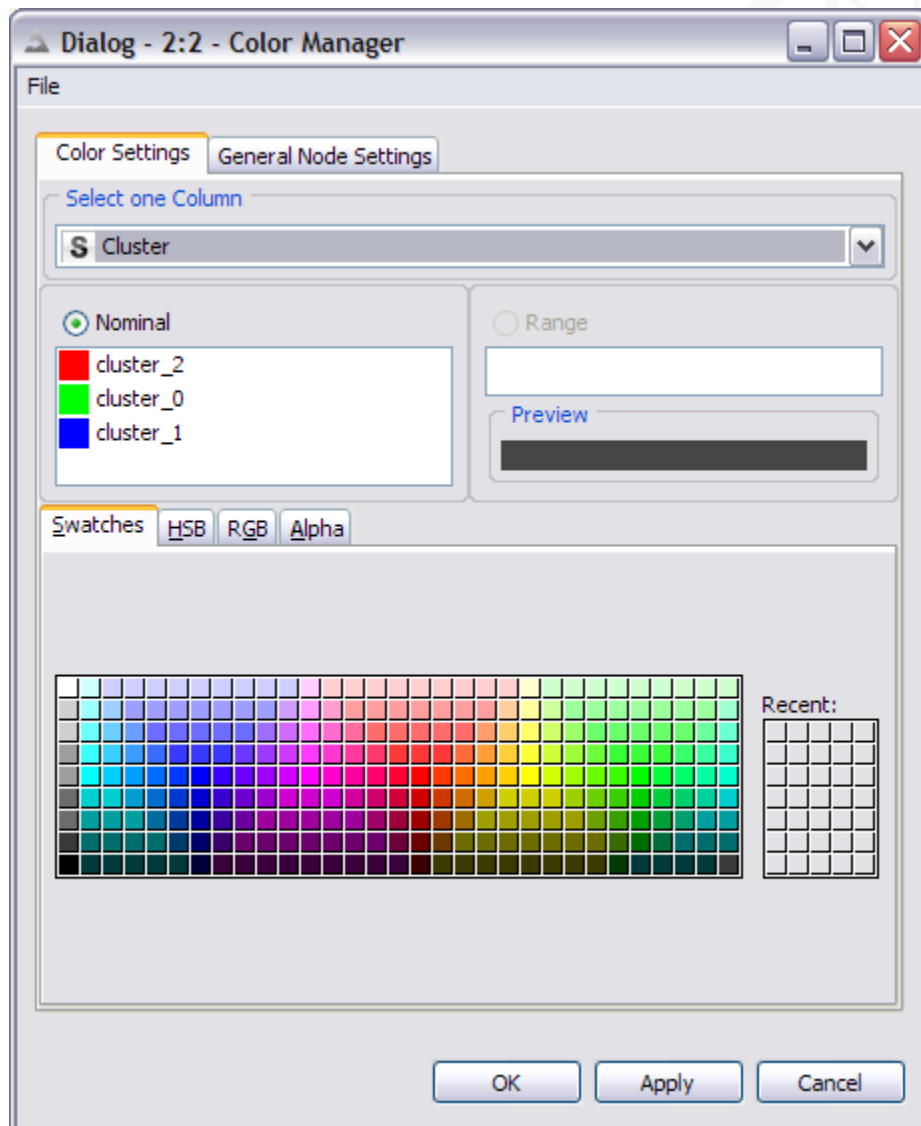
Executing Nodes

Now execute the *Scatter Plot* node: the workbench will execute all predecessor nodes for you. In a larger, more complex flow you could select multiple nodes and trigger execution for all of them. The workflow manager will execute the nodes as needed, if possible in parallel.

Inspecting the Results

In order to examine the data and the results, open the nodes' views. In our example, the *K-Means*, the *Interactive Table* and the *Scatter Plot* have views. Open them from the nodes'

© Copyright KNIME GmbH

context menus.

Hiliting
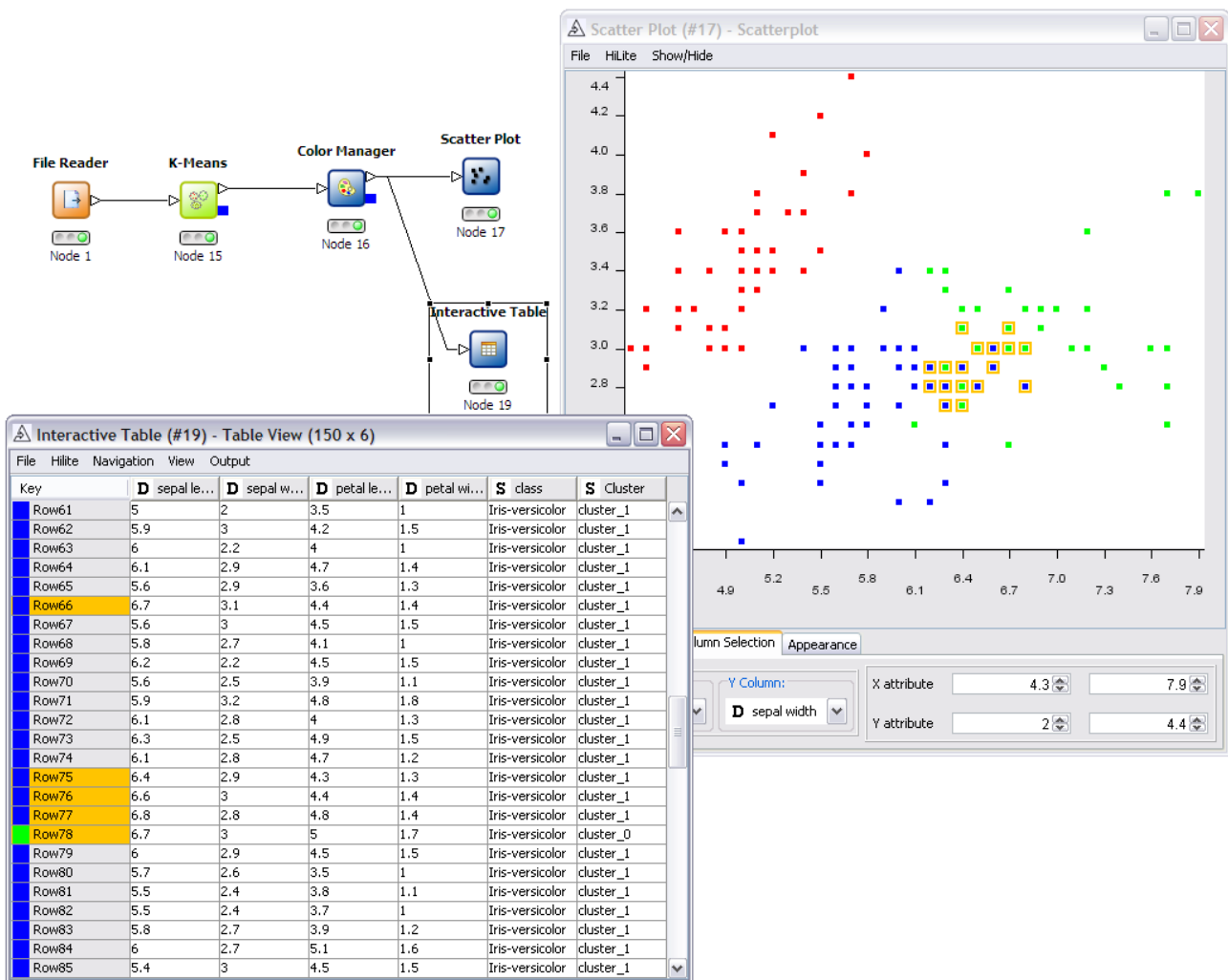
Select some points in the scatter plot and choose "Hilite Selected" from the "Hilite" menu. The hilited points are marked with an orange border. You will also see the hilited points in the table view. The propagation of the hilite status works for all views in all branches of the flow displaying the same data.

Embark on Your Own Voyage of Discovery!

Now this was just a very simple example to get you started. There is a lot more to discover. Play with it! We tried to keep it simple and intuitive. We would love to receive your
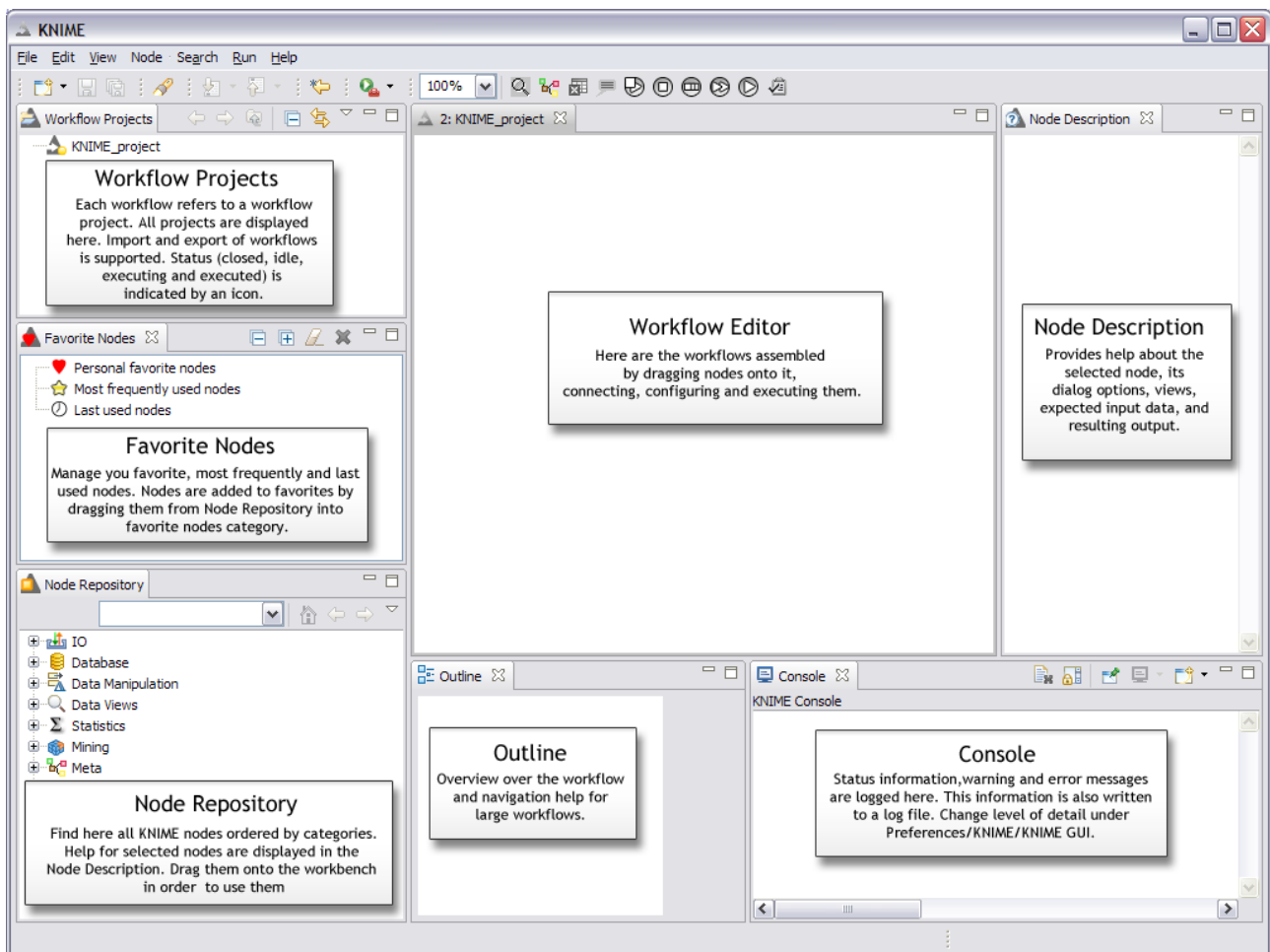
feedback and find out what you liked and what you did not like; things you find awkward or things that did not seem to work.

KNIME Workbench User Guide

Description of Available Views

In the following the KNIME workbench and its features are described in more detail. When KNIME is initially opened it starts with the following arrangement of views :

Workflow Projects

All KNIME workflows are displayed in the Workflow Projects view. The status of the

© Copyright KNIME GmbH

workflow is indicated by an icon showing whether the workflow is closed, idle, executing or if execution is complete.

Favorite Nodes

The Favorite Nodes view displays your favorite, most frequently used and last used nodes. A node is added to your favorites by dragging it from the node repository into the personal favorite nodes category. Whenever a node is dragged onto the workflow editor, the last used and most frequently used categories are updated.

The favorite nodes view comes with the following actions in the menu bar of the view:

	Collapses all expanded categories
	Expands all categories
	Clears the last used and most frequently used categories
	Removes the selected node from your favorites

The number of nodes in the most frequently and last used categories is per default restricted to ten nodes. This number can be adjusted in preferences. Select "File/Preferences.."/KNIME/KNIME GUI to set different values for the maximum size of frequently used nodes and maximum number of last used nodes.

Node Repository

The node repository contains all KNIME nodes ordered in categories. A category can contain another category, for example, the Read category is a subcategory of the IO category.

Nodes are added from the repository to the workflow editor by dragging them to the workflow editor.

Selecting a category displays all contained nodes in the node description view; selecting a node displays the help for this node.

If you know the name of a node you can enter parts of the name into the search box of the

node repository. As you type, all nodes are filtered immediately to those that contain the entered text in their names:
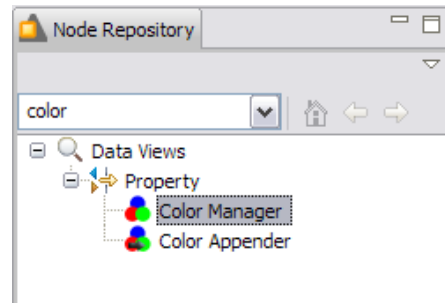
Outline

The outline view provides an overview over the whole workflow even if only a small part is visible in the workflow editor (marked in gray in the outline view). The outline view can also be used for navigation: the gray rectangle can be moved with the mouse, which causes the editor to scroll so that the visible part matches the gray rectangle.

© Copyright KNIME GmbH
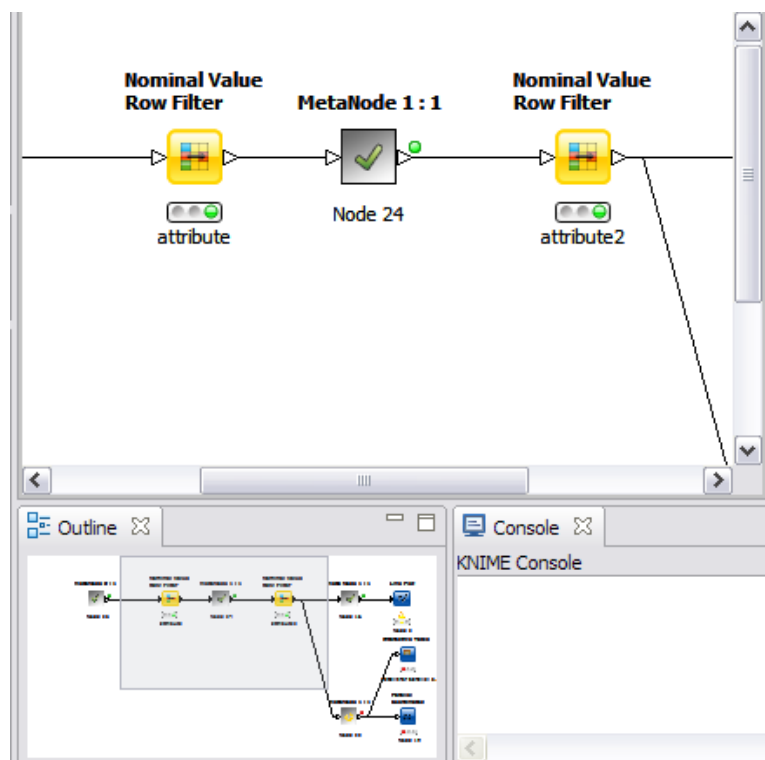
Console

The console view prints out error and warning messages in order to give you a clue of what is going on under the hood. The same information (with a DEBUG detail level is written to a log file, which is located at {workspace}/.metadata/knime/knime.log).

If you want to change the level of detail (either of the log file or of the console view) go to File/Preferences.../KNIME for the level of detail of the log file or to File/Preferences.../KNIME/KNIME GUI for the level of detail of the console view.</p>

You can choose between:

- **DEBUG**: Debug messages mainly used for development. It is not recommended to use this for the console view since it slows down KNIME.
- **INFO**: Logs information messages. Not really important but also not completely useless!
- **WARNING**: If a node fails in configure a warning message is also issued. Warning messages are not fatal; usually the workflow can continue to be executed but they denote that something worth to knowing about has taken place. Default and recommended level for the console view.
- **ERROR**: Only issued when something fatal has happened, i.e. the workflow can no longer be executed.

Node Description

The node description displays information about the selected node (or the nodes contained in a selected category). In particular, it explains the dialog options, the available views, the expected input data and resulting output data. Under Linux there are some issues with this view, since it needs the system's web browser.

" KNIME/Eclipse tries to find a Mozilla-based browser automatically, if the environment variable MOZILLA_FIVE_HOME is not set. The knime.sh should note which browser it is using in this case. You can try to explicitly set MOZILLA_FIVE_HOME to the firefox directory and if this doesn't help you can also try passing "-Dorg.eclipse.swt.browser.XULRunnerPath=..." to knime.sh. There is a known problem with Firefox 3 (and xulrunner >= 1.9) for which there is no workaround other than using an older version. This may also cause you some trouble." See also the linked Eclipse bug

report https://bugs.eclipse.org/bugs/show_bug.cgi?id=236724.

In order to provide a full text search, the node descriptions are also integrated in the Eclipse help. Select Help/Help Contents from the menu in order to open the Eclipse built-in help. There is a KNIME category, which has a Node Descriptions submenu. In the search field you can perform a full text search across all the node descriptions. If, for example, you type "cluster", all node descriptions containing the word cluster are displayed:
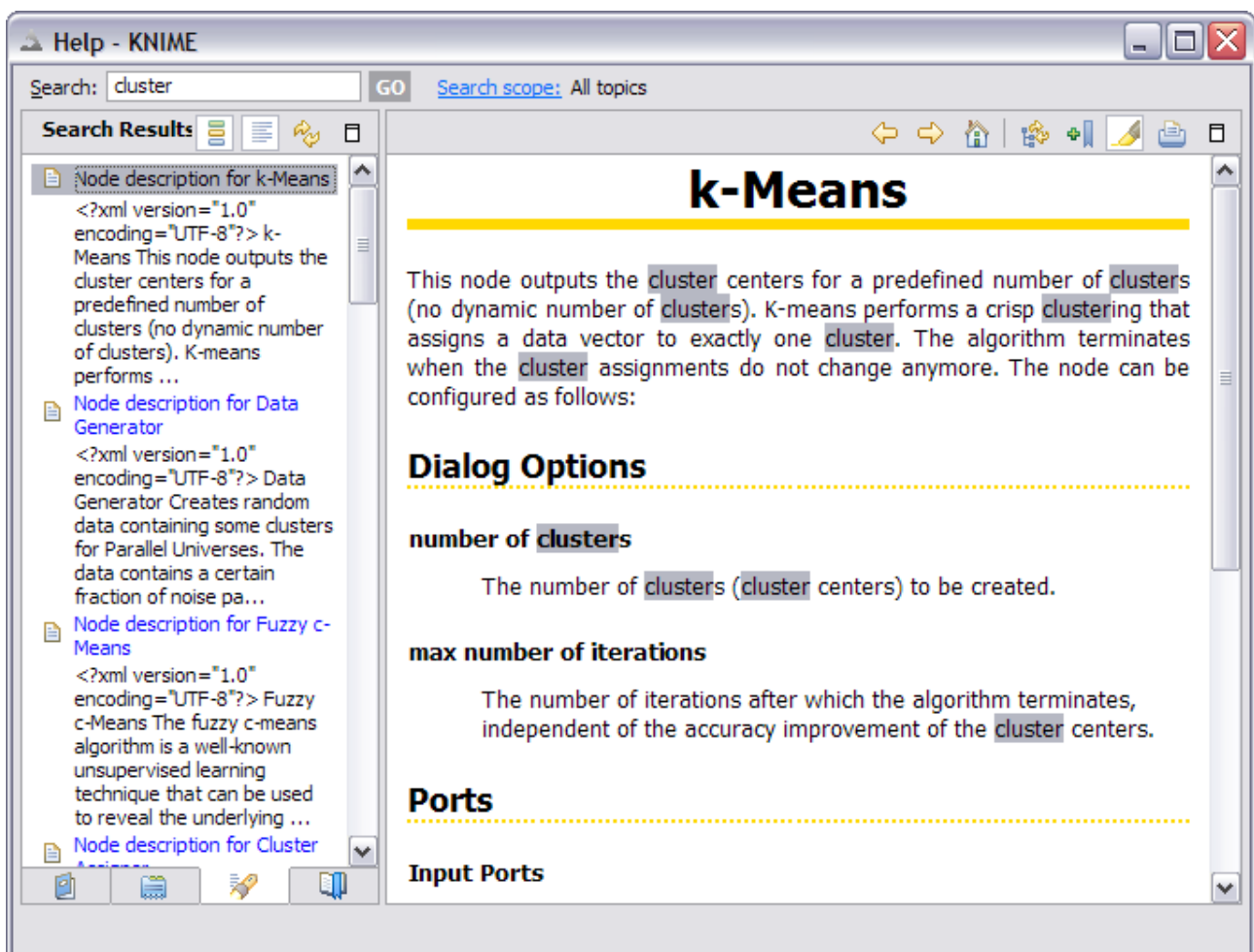
Preferences

The preferences are opened with File/Preferences... The KNIME-related preferences are separated into three categories:
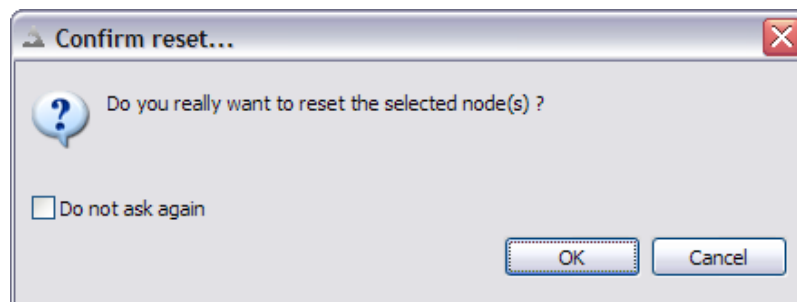
KNIME

Preferences of KNIME which also apply to KNIME if started in batch mode

- **Log file Log Level**: Level of detail for the log file. Default value is DEBUG, which means that information for developers is also logged. Sending this log file to us if you encounter any unexpected behavior may give us a hint at what caused the problem.

- **Maximum working threads for all nodes**: The KNIME workflow manager tries to optimize the execution time of all nodes, for example, by distributing separate branches of the workflow to several threads. It boils down to running nodes in parallel wherever possible. And here you can enter how many threads should be used for parallelization. By default it is twice the number of CPU's. This has proven to be a good amount.

- **Directory for temporary files**: KNIME needs to store some temporary files (data of executed but not yet saved workflows) somewhere. This is where you can specify the location.

KNIME GUI

Preferences related to the graphical user interface of KNIME.

- **Console View Log Level**: Level of detail for the log messages displayed in the console view. Usually WARNING is enough. DEBUG slows down performance and is mostly useful for development.

- **Confirm Node Reset**: Check or uncheck whether you want a confirmation dialog to pop up when you reset an already executed node. If you checked the "Do not ask again" checkbox in this type of dialog, go to preferences to make them reappear.

- **Confirm Node/Connection Deletion**: Same as above but for confirmation of

deleting nodes and/or connections.

- **Confirm reconnection of already connected nodes**: As of KNIME 2.0 it is possible to drag a connection to an already connected port. The connection is replaced if the node is configured, but if it is executed (and thus will be reset by replacing the connection) a confirmation dialog appears. This confirmation dialog can also be turned on or off via this preference.
- **Maximum size for most frequently used nodes**: The amount of nodes maintained by the most frequently used nodes category of the Favorite Nodes view.
- **Maximum size of last used nodes**: The amount of nodes maintained by the last used nodes category of the Favorite Nodes view.

Master Key

KNIME does not store any passwords (e.g. for databases) in plain tests but encrypts them using a master key.

Workflow Editor

The workflow editor is used to assemble workflows, configure and execute nodes, inspect the results and explore your data. This section describes the interactions possible within the editor.

Node Options

Configure

When a node is dragged to the workflow editor or is connected, it usually shows the red status light indicating that it needs to be configured, i.e. the dialog has to be opened. This can be done by either double-clicking the node or by right-clicking the node to open the context menu. The first entry of the context menu is "Configure", which opens the dialog. If the node is selected you can also choose the related button from the toolbar above the editor. The button looks like the icon next to the context menu entry.

Execute

In the next step, you probably want to execute the node, i.e. you want the node to actually

perform its task on the data. To achieve this right-click the node in order to open the context menu and select "Execute". You can also choose the related button from the toolbar. The button looks like the icon next to the context menu entry.

It is not necessary to execute every single node: if you execute the last node of connected but not yet executed nodes, all predecessor nodes will be executed before the last node is executed.

Execute All

In the toolbar above the editor there is also a button to execute all not yet executed nodes on the workflow.

This also works if a node in the flow is lit with the red status light due to missing information in the predecessor node. When the predecessor node is executed and the node with the red status light can apply its settings it is executed as well as its successors. The underlying workflow manager also tries to execute branches of the workflow in parallel.

Execute and Open View

The node context menu also contains the "Execute and open view" option. This executes the node and

immediately opens the view. If a node has more than one views only the first view is opened.

Open View

A node can have no, one or several views. Each view appears as an entry in the node's context menu. Select it in order to open the related view. A view that is opened before the node has been executed, is updated as soon as the node is executed. You can open the view of a node several times, e.g. if you want to compare different columns in a scatter plot. A view is automatically reset if the node is reset.

Open Out-port View

If a node does not have a view but you are interested in the result of the node's operation on the data, you can inspect the data. It is available at the node's out-port. At the bottom of the context menu there is an entry for each outport of the node. Each one opens the

referring out-port view. Note, that the out-port view does not support any interaction or hiliting. If you want to hilite data or see hilited data you have to connect the out-port to the Interactive Table node.

Reset

You can reset a node by choosing the reset option from the context menu. The node returns from the executed state (green status light) to configured state (yellow status light). If the node is selected you can also choose the related button from the toolbar above the editor.

The button looks like the icon next to context menu entry.

Cancel

If a node is currently executing you can cancel the execution by selecting the "Cancel" option from the context menu or the related button (same icon as in the context menu) from the toolbar.

Cancel All

The toolbar also contains a "Cancel All" button, which cancels the execution of all running nodes.

Enter Custom Node Name

When a node is dragged to the workflow it has a name such as "Node 1" or similar by default below the status light. You can change this name to better describe what the node is actually doing, e.g. "filter values > 10". This can be done by selecting the node and then clicking on the name: the name becomes editable. Press "Return" to apply your changes.

Enter Custom Node Description

In the context menu you will also find the "Node name and description" option. Selecting this opens a dialog to enter a new name for the node. In addition you can enter a more detailed description or notes about the node. This action is also available via a button in the toolbar.

Connections

You can connect two nodes by dragging the mouse from the out-port of one node to the in-port of another node. Loops are not

permitted.

If a node is already connected you can replace the existing connection by dragging a new connection onto it. If the node is already connected you will be asked to confirm the resulting reset of the target node. You can also drag the end of an existing connection to a new in-port (either of the same node or to a different node).

Import/Export of workflows

Import of Workflows

You can import a workflow either from a different workspace or from a zip file, e.g. if the workflow was exported from KNIME. The import wizard is either opened from the menu "File/Import KNIME workflow..." or by opening the context menu in the workflow projects view and selecting "Import KNIME workflow...".
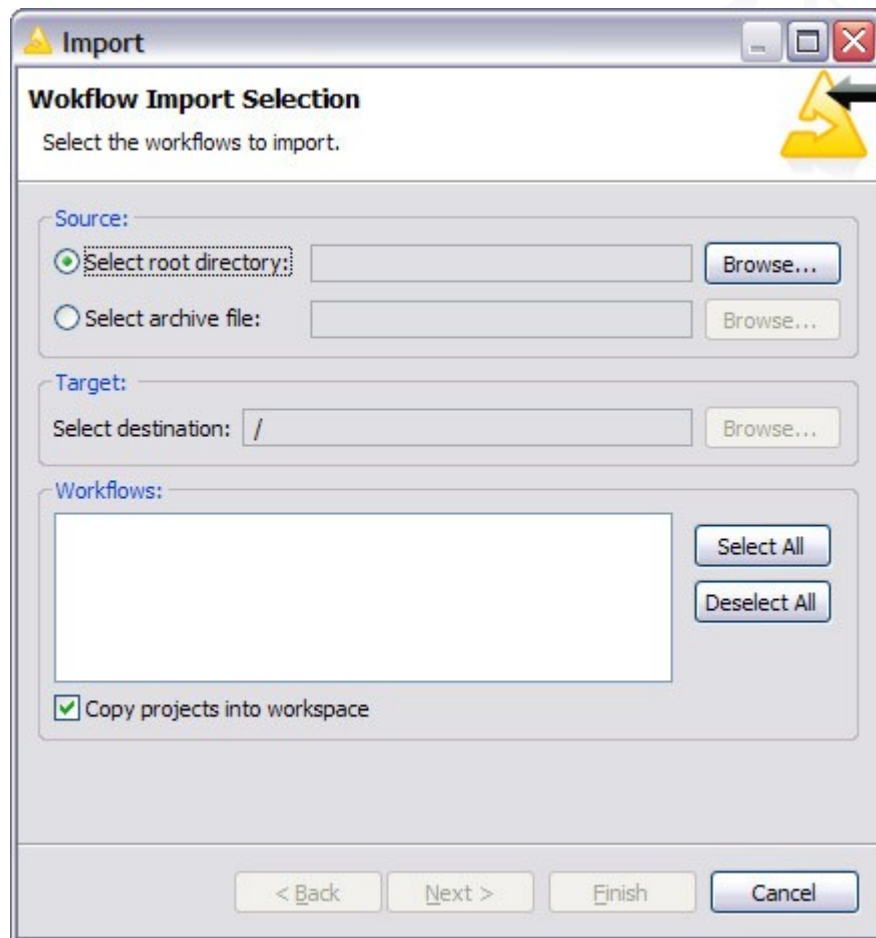
Select the root directory if you want to import workflows from another workspace. Select the archive file option if you want to browse to the zipped workflow. Select the workflows you want to import. If a workflow with the same name already exists in your current workspace you can rename the imported one on the next page of the wizard. By clicking OK the project is imported to your workspace. If you unchecked the "Copy projects into workspace", changes to that workflow will also apply to the workflow in the source location.

Export of Workflows

The export workflow action is also available via the menu (File/Export KNIME workflow...") or via the context menu of the workflow projects view. Both open the export workflow wizard. Select the workflow you want to export. If you right-clicked a workflow to open the export wizard this workflow is pre-selected. In the second field browse to the target location or enter the path leading to the export location.
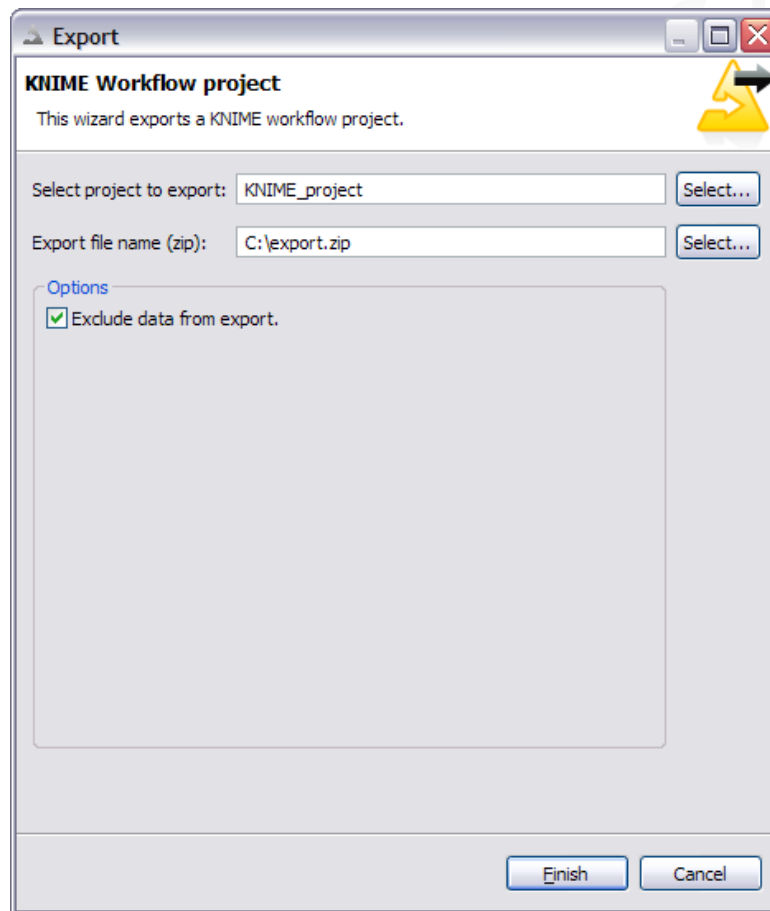
The option to exclude data from being exported is activated by default. If checked, only the structure of the workflow is exported, which will result in a configured but non-executed workflow if it is re-imported. If you explicitly share the data (for example if the other person does not have access to a database) you can uncheck this option.

Using Meta Nodes

Meta nodes are nodes that contain subworkflows, i.e. in the workflow they look like a single node, although they can contain many nodes and even more meta nodes. They are created with the help of the meta node wizard. You can open the meta node wizard by either selecting "Node/Add Meta Node"
from the menu or by clicking the button with the meta node icon in the toolbar (workflow editor must be active).

Create Pre-defined Meta Node

To create a pre-defined meta node, select one and click "Finish". The one you have selected is added to the workflow.
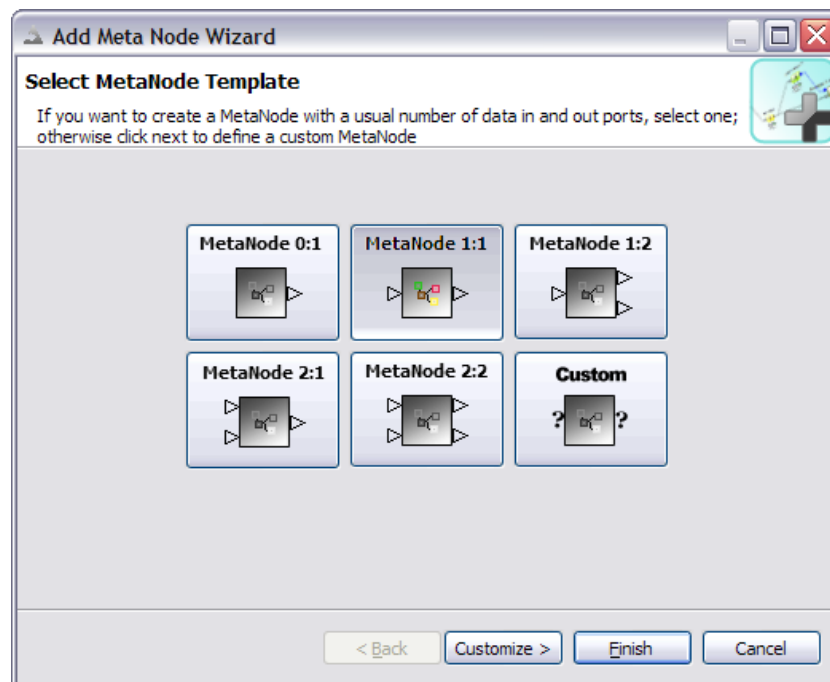
Create Customized Meta Node

If you need either a different number of in- or out-ports or want to have different port types you can select one of the pre-defined meta nodes as a template and then click "Customize" to access the next page of the wizard.
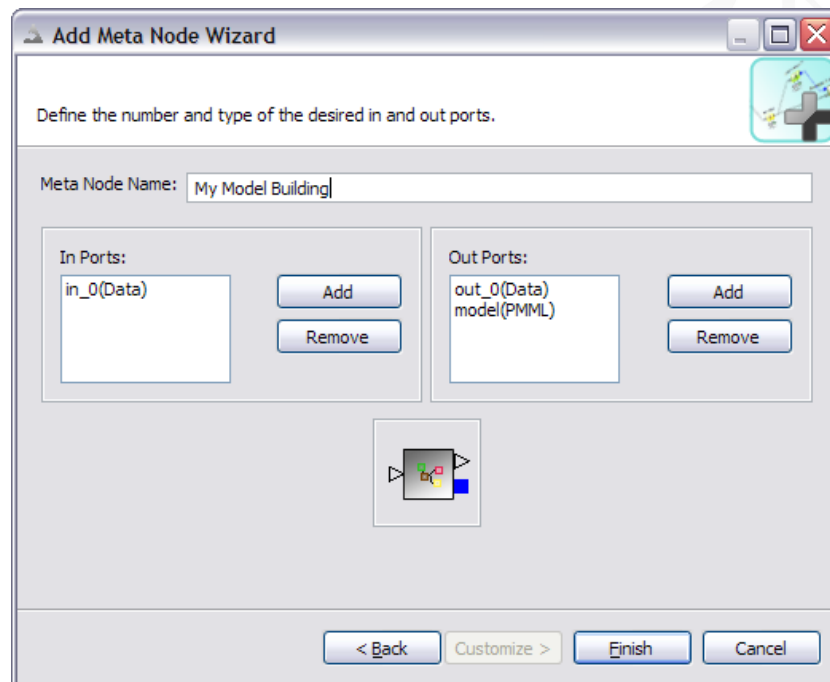
On this page you can add or remove in- and out-ports of the template. An icon at the bottom immediately gives you a preview of the node. When you add a port to the template you can choose the type for the port:
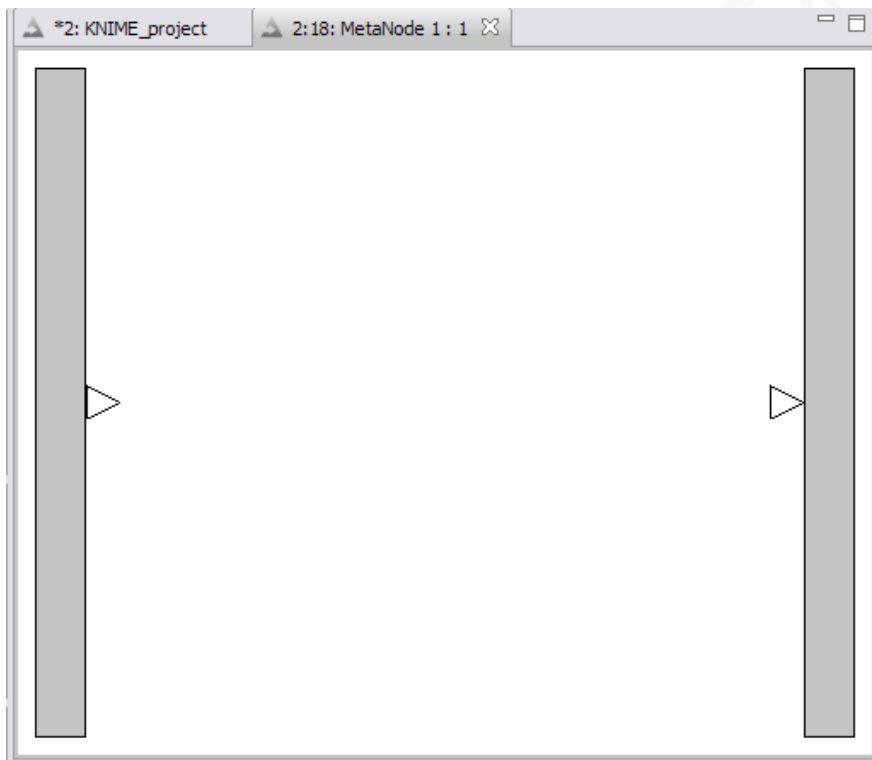
- data,
- data base, or
- data mining port (PMML).

Once the node fits your needs, click "Finish" in order to add it to the workflow.

Inside a Meta Node

In order to open a meta node you can either double-click it or choose "Open Subworkflow editor" from its context menu.

Depending on the amount of in- and out-ports the inside of a meta node looks similar to the picture below:

The in- or out-ports are fixed to the so-called workflow port bars, which can be moved and resized. The data connected to the in-port from outside appears inside the meta node editor in the in-port. And vice versa: the data connected to the inner out-port appears in the outer out-port.

Meta Nodes From Outside

Meta nodes look different to normal nodes. The background icon is not rounded and has a dark gray background. There is no status light and no progress.

States of Meta Nodes

A meta node does not have as many states as a node. The states of a meta node are the same as the states of a workflow. A meta node can be

- **idle/configured**: If there is at least one node inside the meta node that is neither executed nor executing.
- **executing**: If at least one node is executing
- **executed**: If all contained nodes are executed

The state of a meta node is displayed by an icon in the meta node (seen from outside).

Out-Ports of Meta Nodes

In addition to the state of the meta node the out-ports of a meta node also have states. A small decorator icon displays the state of the out-port. If a port is not connected or connected to an idle node neither specs nor data are available. This is indicated by a red icon. If a port is connected to a configured node, some specs are available. This is indicated by a yellow icon. If a port is connected to an executed node, specs and data are available. This is indicated by a green icon.