RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inria

KNIME

Au-delà du script Python : les workflows Knime et le rendu graphique

Sommaire

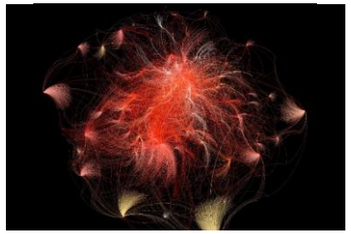
1. Introduction : Inria en 3 diapos
2. Comment exploiter HAL
3. Démonstration

01
Inria en 3 diapos

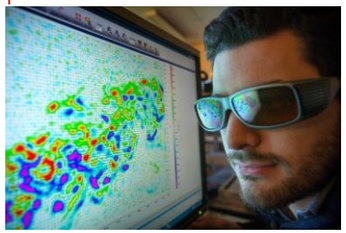
L’excellence scientifique chez Inria

ALGORITHMES & PROGRAMMATION

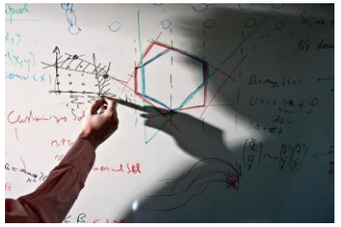
SCIENCE DES DONNÉES & INGÉNIERIE DE LA CONNAISSANCE



MODÉLISATION & SIMULATION



OPTIMISATION & CONTRÔLE



ARCHITECTURES, SYSTÈMES & RÉSEAUX

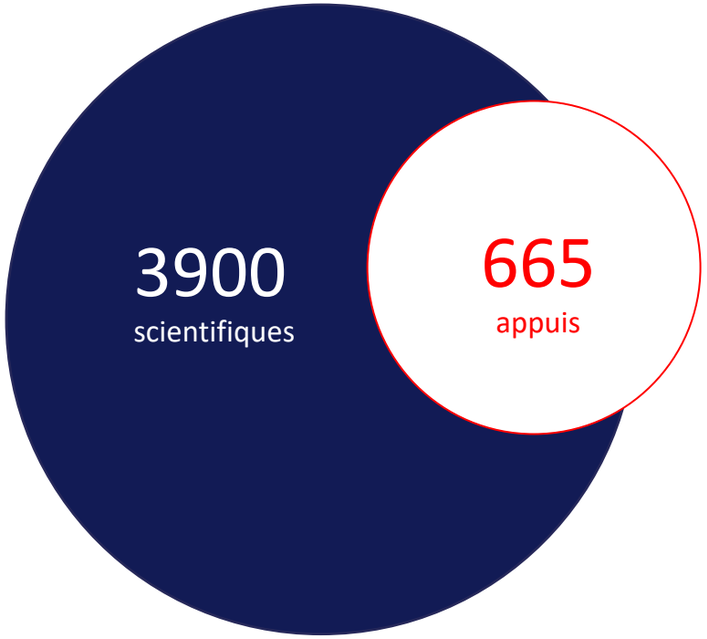
SÉCURITÉ & CONFIDENTIALITÉ

INTERACTION & MULTIMÉDIA

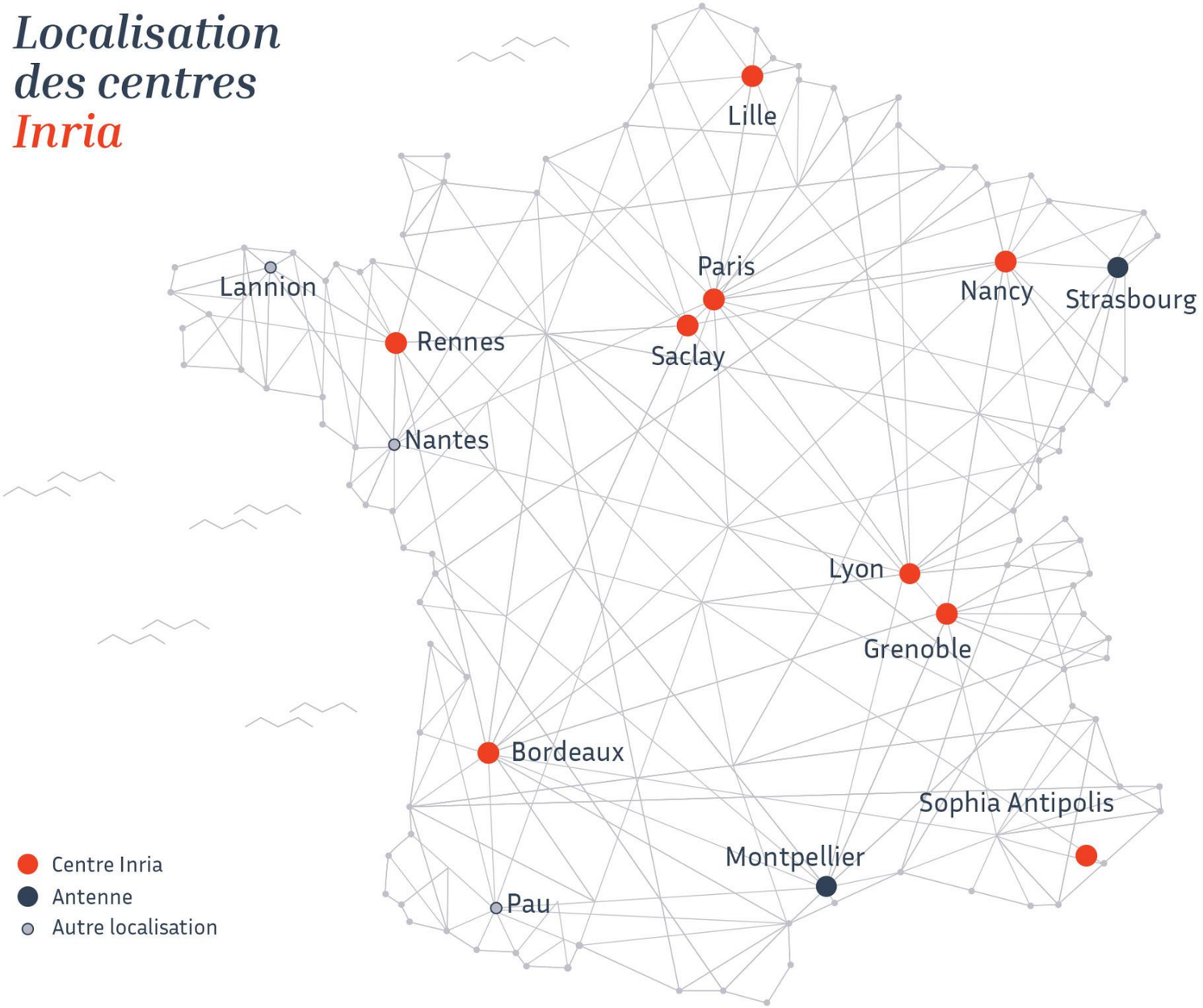
INTELLIGENCE ARTIFICIELLE & SYSTÈMES AUTONOMES

Centres de recherche

4565 collaborateurs

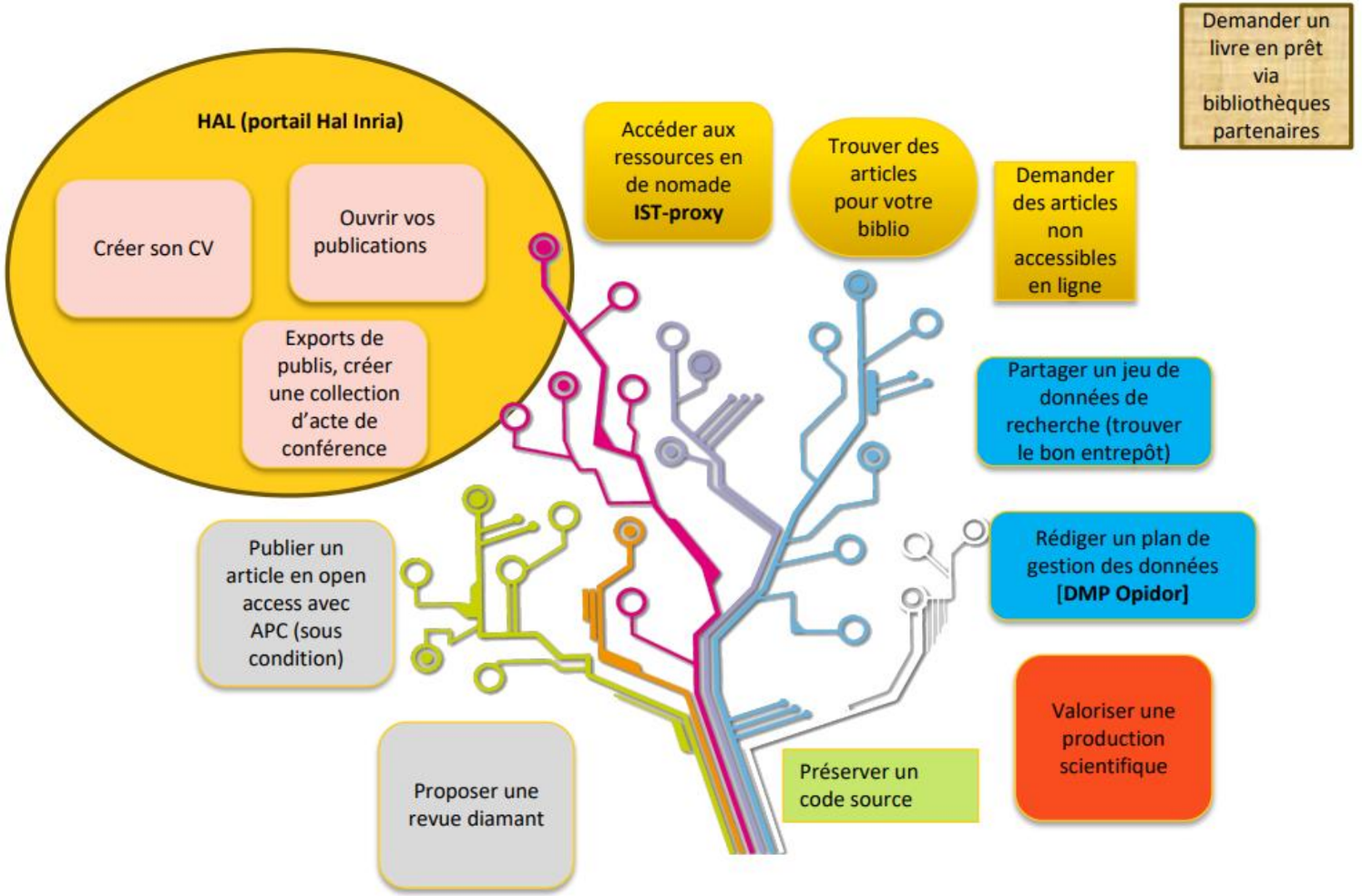


Localisation des centres Inria



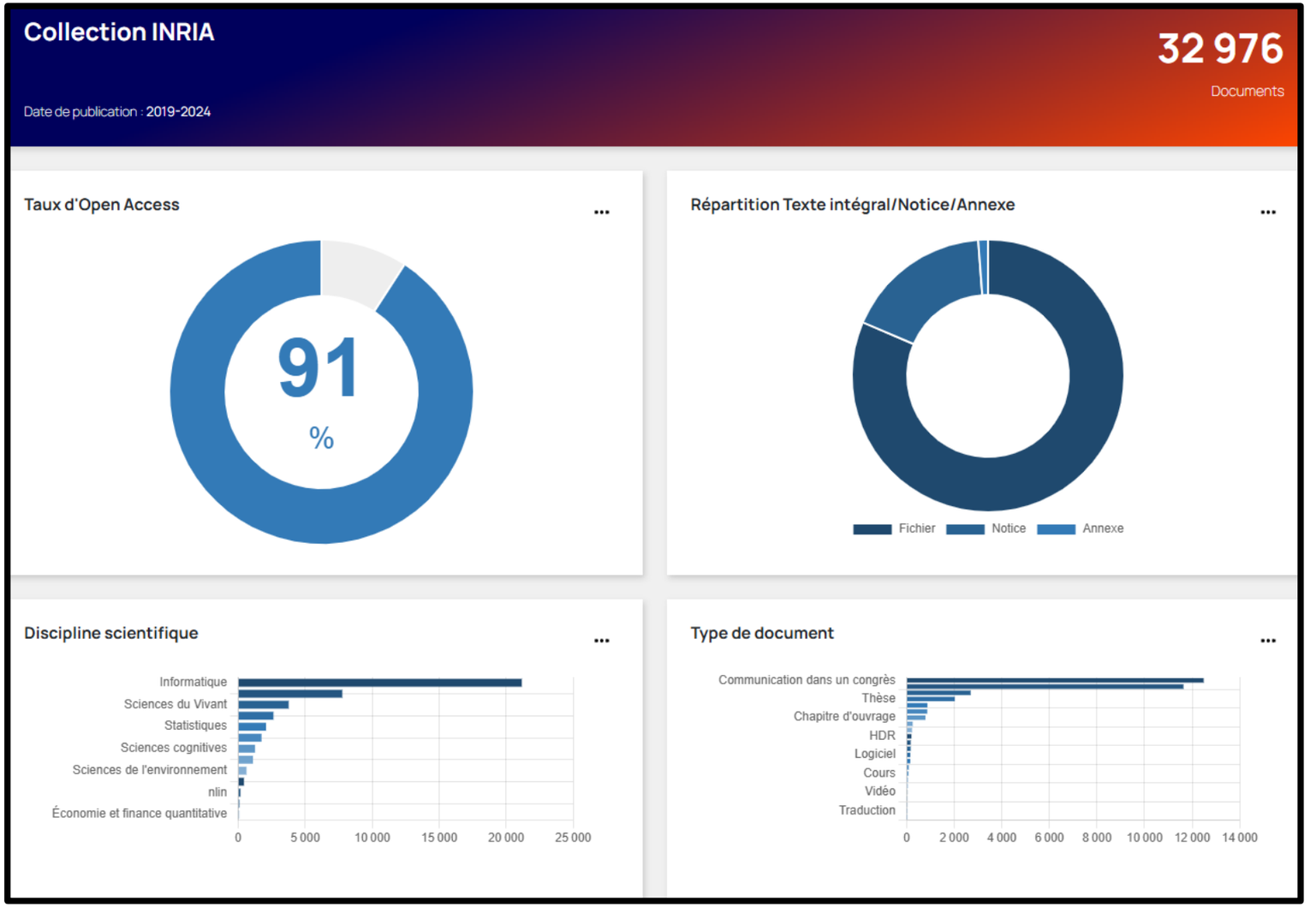
Service Information et édition scientifiques (IES)

6 pôles, dont le Pôle « Archives ouvertes » : 4,5 ETP pour env. 6000 publications par an



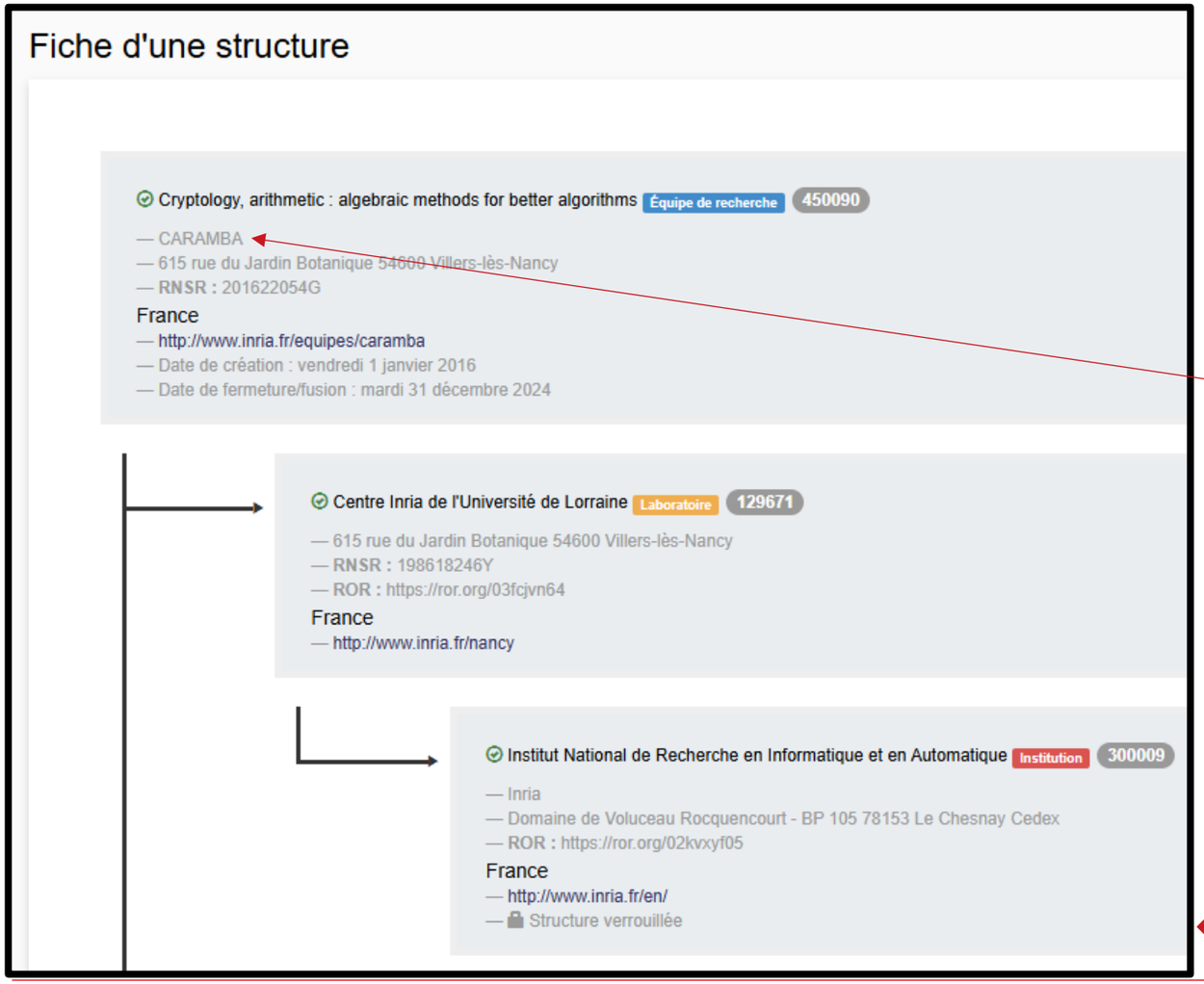
02
Comment exploiter HAL

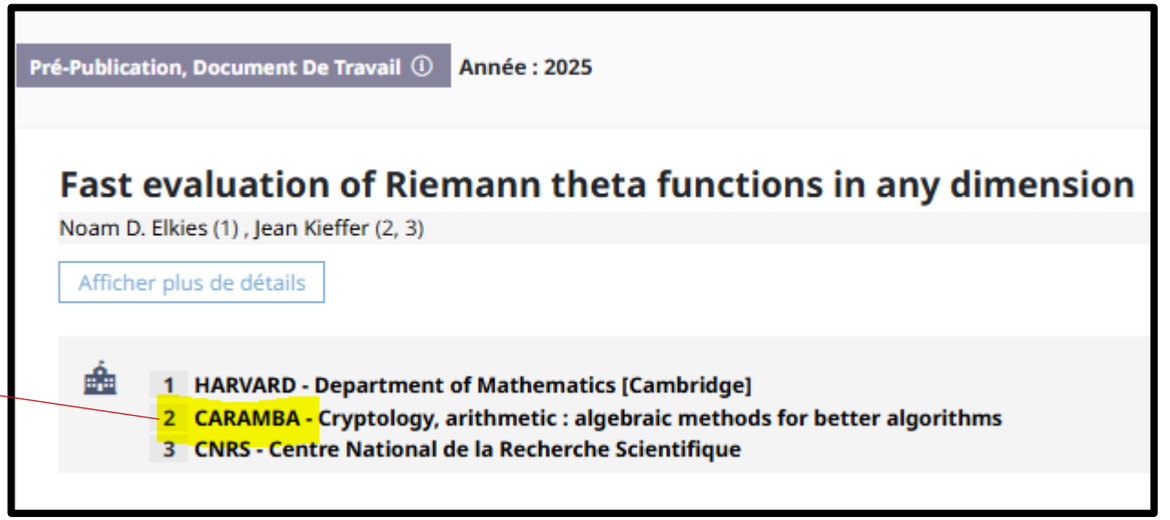
Notre source d’information principale : HAL



Source : Monitor.hal.science

L’identification des publications INRIA : un portail, une collection, des structures hiérarchisées



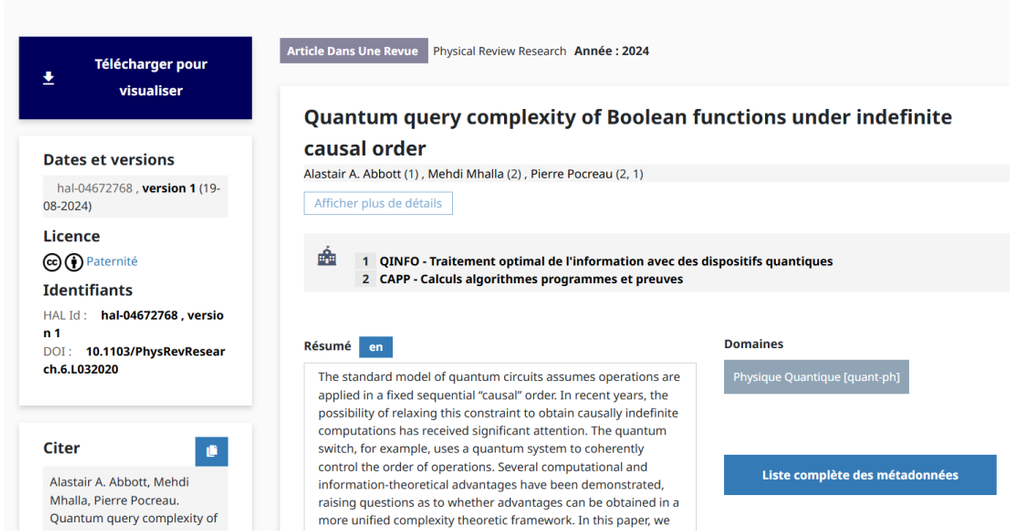


HAL

Aurehal

Exploiter les données de HAL - JSON

Interface utilisateur



Python

API avec sortie en wt=JSON


```
{
  "response": {
    "numFound": 1,
    "start": 0,
    "maxScore": 5.902752,
    "numFoundExact": true,
    "docs": [
      {
        "docid": "4672768",
        "label_s": "Alastair A. Abbott, Mehdi Mhalla, Pierre Pocreau. Quar
(3), pp.L032020. &#x27E8;10.1103/PhysRevResearch.6.L032020&#x27E9;. &#x27E
        "citationRef_s": "<i>Physical Review Research</i>, 2024, 6 (3), pp
href=\"https://dx.doi.org/10.1103/PhysRevResearch.6.L032020\">&#x27E8;10.1
        "citationFull_s": "Alastair A. Abbott, Mehdi Mhalla, Pierre Pocrea
Research</i>, 2024, 6 (3), pp.L032020. <a target=\"_blank\" href=\"https:/
target=\"_blank\" href=\"https://inria.hal.science/hal-04672768v1\">&#x27E
        "label_bibtex": "@article{abbott:hal-04672768,\n  TITLE = {{Quantu
and Mhalla, Mehdi and Pocreau, Pierre},\n  URL = {https://inria.hal.scienc
VOLUME = {6},\n  NUMBER = {3},\n  PAGES = {L032020},\n  YEAR = {2024},\n
04672768v1/file/PhysRevResearch.6.L032020.pdf},\n  HAL_ID = {hal-04672768}
        "label_endnote": "%0 Journal Article\n%T Quantum query complexity
dispositifs quantiques (QINFO)\n%+ Calculs algorithmes programmes et preuv
2643-1564\n%J Physical Review Research\n%I American Physical Society\n%V 6
[physics]/Quantum Physics [quant-ph]Journal articles\n%X The standard mode
the possibility of relax +1684 more",
        "label_coins": "<span class=\"Z3988\" title=\"ctx_ver=Z39.88-
2004&amp;rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Adc&amp;rft.type=journa
04672768&amp;rft.identifier=doi%3A10.1103%2FPhysRevResearch.6.L032020&amp;
;rft.creator=Abbott%2C%20Alastair%20A.&amp;rft.creator=Mhalla%2C%20Mehdi&a
26&amp;rft.source=Physical%20Review%20Research\"></span>",
        "openAccess_bool": true,
        "popularLevel_s": "0",
        "peerReviewing_s": "1",
        "audience_s": "2",
        "licence_s": "http://creativecommons.org/licenses/by/",
        "domainAllCode_s": [
          "phys.qphy"
        ],
        "level0_domain_s": [
          "phys"
        ],
        "domain_s": [
          "0.phys",
          "1.phys.qphy"
```


+ traitements divers

Exploiter les données de HAL – XML-TEI

Interface utilisateur

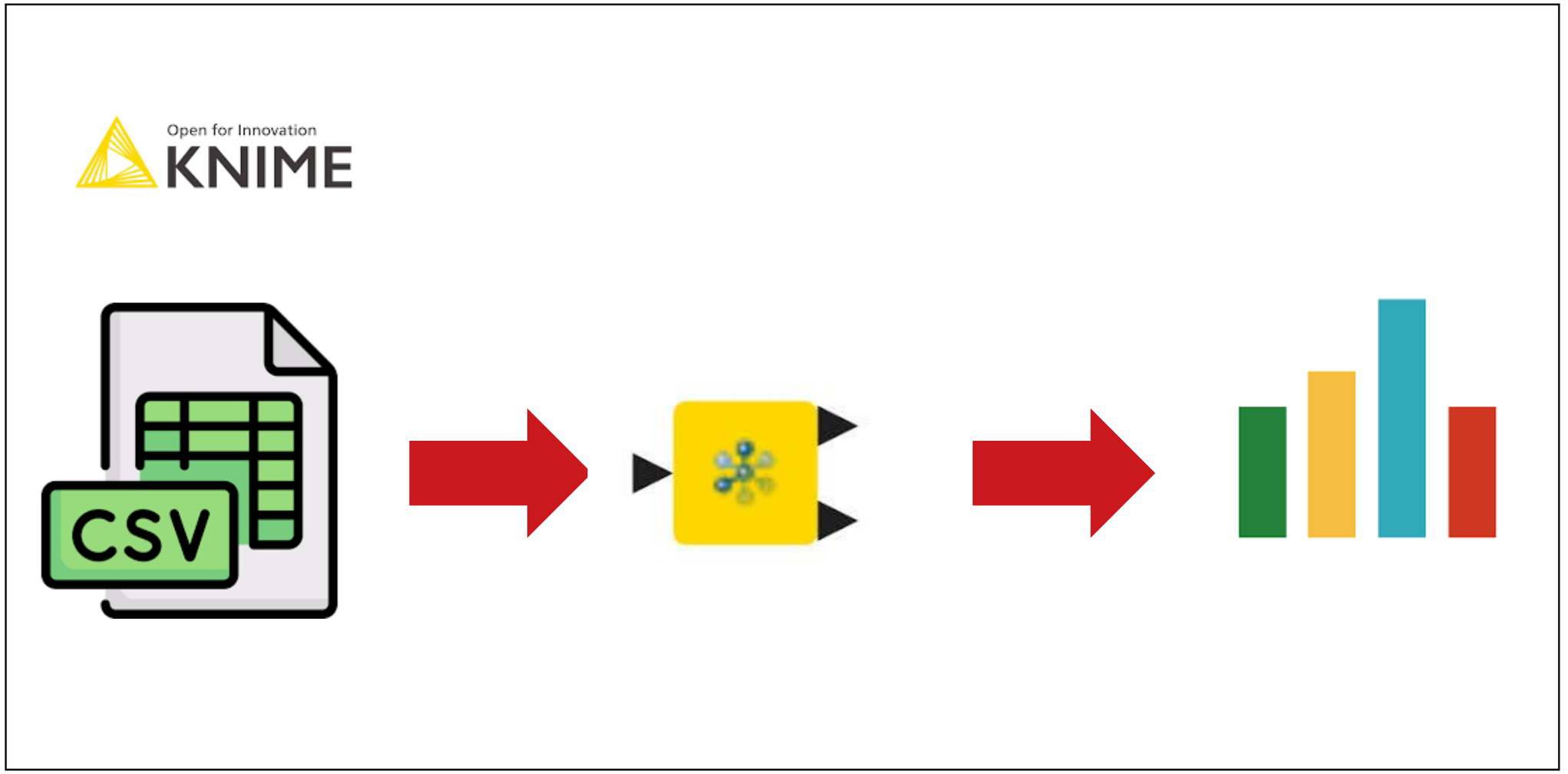
Python

API avec sortie en wt=XML-TEI

```
      <persName>
        <forename>Alastair</forename>
        <surname>Abbott</surname>
      </persName>
      <email type="md5">f083b343b04d7dc69dac1827945e041f</email>
      <email type="domain">inria.fr</email>
    </editor>
    <funder ref="#projanr-75279"/>
    <funder ref="#projanr-69691"/>
    <funder ref="#projanr-57829"/>
    <funder ref="#projanr-42467"/>
  </titleStmt>
  <editionStmt>
    <edition n="v1" type="current">
      <date type="whenSubmitted">2024-08-19 14:38:53</date>
      <date type="whenModified">2025-04-01 17:32:03</date>
      <date type="whenReleased">2024-08-20 08:59:44</date>
      <date type="whenProduced">2024-07-26</date>
      <date type="whenEndEmbargoed">2024-08-19</date>
      <ref type="file" target="https://inria.hal.science/hal-04672768v1/doc
        <date notBefore="2024-08-19"/>
      </ref>
      <ref type="file" subtype="publisherPaid" n="1" target="https://inria.
        <date notBefore="2024-08-19"/>
      </ref>
    </edition>
    <respStmt>
```

+ traitements divers

Exploiter les données de HAL – KNIME



Pour aller plus loin

Pour aller plus loin

- Exemples de workflows : knime.com/example-workflows
- Formations en ligne **:** knime.com/learning-hub
- Mémos KNIME en PJ

Contacts

- Kumar Guh**a : kumar.guha@inria.fr**
- Andréa Nebot **:andrea.nebot@inria.fr**

Merci.

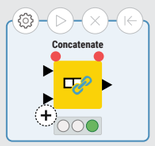
Cheat Sheet: KNIME for Spreadsheet Users

Early access of KNIME Analytics Platform Version 5

Open for Innovation
KNIME

Getting started with KNIME Analytics Platform

• Use the Getting Started Guide to take your first steps with visual workflows at:
www.knime.com/early-access-knime-ap-v5-getting-started
• Learn more about included nodes and explore working examples in the **KNIME Analytics Platform Version 5 Starter Perspective Collection** on **KNIME Community Hub**.



Add comment

Node Action Bar: Interact directly with the node to, e.g., configure, execute, cancel or reset a node.
Configure: Open the configuration dialog.
Execute: Executes the node.
Cancel: Cancels the execution of the node.
Reset: Resets the node.
Node Labels: Double click Add comment below the node to add a comment/label.
Dynamic ports: Additional input ports can be added by clicking the plus on the left side of the node.

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.

Error: The node has encountered an error during execution.

VISUALIZATION

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.

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.

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.

FILTERING

Row Filter — Filters rows in or out of the input table according to a filtering rule. The filtering rule can match a value in a selected column or numbers in a numerical range.

Column Filter — Filters columns in or out of the input table. Columns to be filtered can be manually chosen, selected according to their data type, or based on a wildcard or regex expression matching their name.

Top k Row Filter — Sorts the input table according to a defined sorting criteria and keeps only the first k rows. In the Advanced Settings tab, the output order can be specified.

Table Cropper — Crops the input table based on the chosen row and column range. The row range is defined via row number, the column range either via column name or column number.

VALUE CREATION

Math Formula — Implements a number of math operations across multiple input columns. The math operations can be applied to multiple columns with the Math Formula (Multi Column) node.

Column Renamer — Renames selected columns according to the column name defined in the dialog. Column names must remain unique!

String Replacer — Replaces values in a selected string column if they match a defined pattern.

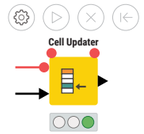
Cell Updater — Updates a single cell of the input table with the value of the specified flow variable. The cell to be updated must be specified via the row number and column name. The output table will be identical to the input table except for the single updated cell.

Table Updater — Updates cells in the top input table with matching cells from the bottom update table. A matching cell must have the same column name and RowID in both tables. Multiple cells of multiple rows and columns can be updated. Additional rows and columns from the update table can be appended to the input table.

FLOW VARIABLES

Flow Variables allow for the parameterization of a workflow. A Flow Variables is a parameter that can assume different values at different execution points in the workflow & overwrite configuration settings in upcoming nodes.

Creating a Flow Variable
1. Use a Configuration or a Widget node to create a Flow Variable at any point in your workflow. 2. Use any of the nodes converting data into Flow Variables.
3. Via the node configuration window in the Flow Variables tab, fill in a blank box with the name of the Flow Variable



Hidden Flow Variable Ports
Each node has two hidden Flow Variable ports to accept incoming Flow Variables & to propagate them to the upcoming nodes. To make these ports visible, hover your cursor over the node. To configure a node's flow variables right-click the node and select **Configure flow variables**.

DATA TYPES & CONVERSIONS

[S] **String**: Sequence of characters, e.g. "This is a string"
[I] **Integer**: Whole real valued number, e.g. -100 or 345
[D] **Double**: Real valued number, e.g. -0.432 or 45.39
Date&Time: A data format for date, time, date&time, or date&time plus time zone.
[B] **Boolean**: Two possible values only, e.g. TRUE and FALSE
[...] **Collection Cell**: Collection of multiple values of either the same or different types e.g., can be a list of values or a set of values. In a set each value occurs only once.
Document/Image: KNIME Analytics Platform supports many more data types like text documents, images, fingerprints, etc.

String to Number — Converts the data type of the selected columns from string to either double or integer. Use the Number to String node for the opposite conversion.

String to Date&Time — Parses the strings in the selected columns according to a date/time format and converts them into Date&Time cells. Four Date&Time forms are supported: only date, only time, date & time, and date & time plus time zone. Use the Date&Time to String node for the opposite conversion.

READ DATA

Excel Reader — Reads content from sheets in Excel files (xlsx, xlsm, xlsb, and xls format). Sheets and cells to be read can be defined in the configuration window.

Google Sheets Reader — Reads data from a Google Sheets spreadsheet after authenticating with the Google Authentication node.

Google Authenticator — Authenticates against Google API services via the "Authenticate" button's pop-up window.

Microsoft Authenticator — Connects to Microsoft Azure and Office 365 cloud services via a number of interactive authentication options.

WRITE DATA

Excel Writer — Writes the input data table into a spreadsheet of an Excel file (xls or xlsx).

Google Sheets Writer — Writes the input data table into a new Google Sheets spreadsheet after authenticating with the Google Authentication node.

Google Sheets Connector — Connects to Google Sheets, given a Google API connection. Depending on the authentication method, the sheet should be either opened with a Google account or shared with a service account.

SharePoint Online Connector — Connects to a SharePoint Online site and allows downstream nodes to access the document libraries as a file system, e.g., to read or write files and folders, or to perform other file system operations. The connection is closed when the Connector node is reset, or the workflow is closed.

DATE&TIME HANDLING

Extract Date&Time Fields — Extracts selected date and time fields from a selected column of type Date&Time and appends their values in new columns.

Date&Time Shift — Shifts a selected date or time with a defined duration or granularity. The shift value can either be a duration column or a numerical column. A positive shift value is added to the selected date/time, a negative value will be subtracted.

Date&Time Difference — Calculates the difference between two Date&Time objects, e.g., from two selected columns, from a selected column and a fixed value, from a selected column and the current execution time, or from one cell and the cell in the previous row for a selected column.

ORCHESTRATION

Email Sender — Sends HTML or plaintext emails from an external SMTP server. Attachments from the filesystem may also be included.

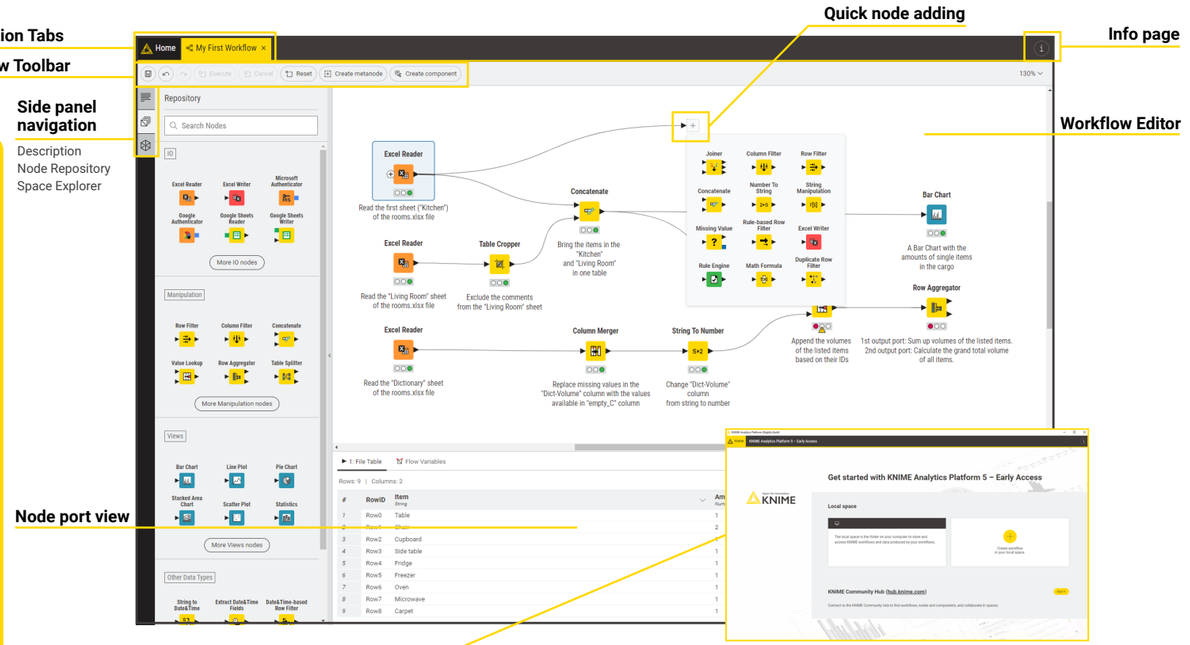
MERGING

Concatenate — Concatenates the rows of all input tables by writing them below each other. Columns with equal names are concatenated. If one input table contains column names that the other table does not, the columns can either be filled with missing values (union) or filtered out (intersection).

Joiner — Joins the columns of the two input tables based on one or multiple joining columns. Allows to select between different joiner modes.

Value Lookup — Adds matching values from a dictionary table to a data table based on a lookup column. When a lookup value matches an entry in the dictionary, the selected cells are added to the data table. Otherwise, missing cells will be inserted.

Column Appender — Combines two or more tables by appending their columns according to the order of input tables. Columns with identical column names will be appended with "(#1)", "(#2)" and so on.



On the entry page you have the option to:
• Create a new workflow in your local space (i.e., the folder on your computer that stores KNIME workflows),
• Open an existing workflow from your local space,
• Connect to the KNIME Community Hub to find workflows, nodes and components, and collaborate in spaces.

DATA AGGREGATION

Pivot — Creates a pivot table by configuring columns for grouping and pivoting. The group columns are turned into unique rows, whereas the pivot values are turned into columns.

Unpivot — Stacks the cells of the selected value columns into one column. The cells of the selected remaining input columns are appended to the corresponding output rows.

Cell Splitter — Splits values in the selected column into two or more substrings, as defined by a delimiter match. A delimiter is a defined character, such as a comma, space, or any other character or character sequence.

Table Splitter — Splits the input table at the row that matches a given condition. The part of the table that occurred before the matching row is forwarded to the top output table, the bottom output table contains the rest of the input table.

Sorter — Sorts the table in ascending or descending order based on the values of one or more columns. Additionally, string-compatible columns can be sorted in alphanumeric instead of lexicographic order.

Cell Extractor — Extracts the value of a single cell from the input table and outputs it as a 1x1 table. The row selection is defined via row number, the column selection either via column name or column number.

Row Aggregator — Aggregates numerical column based on one of the following aggregation functions: Occurrence count, sum, average, minimum, or maximum. Some aggregation functions support weighting. Rows can optionally be grouped by a category column.

Moving Aggregator — Aggregates column values for a defined moving window based on various aggregation functions. The window length is defined in the configuration dialog and can take any number from 2 to the maximum number of rows in the table. The aggregation values are appended as new columns.

Column Combiner — Combines the content of a set of columns row-wise and appends the concatenated string as separate column to the input table.

CLEANING

Missing Value — Defines and applies a strategy to replace missing values in the input table - either globally on all columns, or individually for each column seperately.

Duplicate Row Filter — Detects duplicate rows and applies the selected operation, e.g., removes duplicate rows. Duplicates are rows that have the same value in all selected columns.

Column Merger — Allows to compare values of two columns based on a defined primary and secondary column. The node outputs a new column where the output value for each row will be the value in the primary column if it is not missing, or the value in the secondary column otherwise.

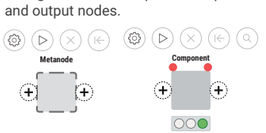
METANODES & COMPONENTS

A Metanode or Component is a node that contains other nodes.

Creating a Metanode or Component
Select all relevant nodes, right-click and select **Create metanode** for a metanode or **Create component** for a component. Right-clicking a metanode or component opens the context menu with a number of options such as expand or configure. To add input or output ports to a metanode or component click the plus on the left side for additional input ports, and the plus on the right side for additional output ports.

Metanodes just collect nodes inside and are an efficient way to clean up your workflow.

Components encapsulate & abstract functionality, can have their own dialog and can have their own sophisticated, interactive views. They can be reused in your own workflows but also shared with others: via KNIME Business Hub or KNIME Community Hub. They can also represent web pages in a Data App deployed to others via KNIME Business Hub. Flow Variables cannot enter or exist a component, unless explicitly configured in the component's input and output nodes.



Resources

• E-Books: KNIME Advanced Luck covers advanced features & more. Practicing Data Science is a collection of data science case studies from past projects. Both available at knime.com/knimepress
• KNIME Blog: Engaging topics, challenges, industry news, & knowledge nuggets at knime.com/blog
• E-Learning Courses: Take our free online self-paced courses to learn about the different steps in a data science project (with exercises & solutions to test your knowledge) at knime.com/knime-self-paced-courses
• KNIME Community Hub: Browse and share workflows, nodes, and components. Add ratings, or comments to other workflows at hub.knime.com
• KNIME Forum: Join our global community & engage in conversations at forum.knime.com
• KNIME Business Hub : For team-based collaboration, automation, management, & deployment check out KNIME Business Hub at knime.com/knime-business-hub

KNIME Press

Open for Innovation
KNIME

Extend your KNIME knowledge with our collection of books from KNIME Press. For beginner and advanced users, through to those interested in specialty topics such as topic detection, data blending, and classic solutions to common use cases using KNIME Analytics Platform - there's something for everyone. Available for download at www.knime.com/knimepress.