JANUS

version 1.0

The Janus software has been developed under a grant by **Umwelt Bundesamt** (**UBA**). UBA holds the intellectual property of the software and of the methodology hereby implemented. UBA releases the software under a GPL3 (GNU General Public License version 3) license.

The software has been developed by the following consortium: Kode Chemoinformatics (Italy), Istituto di Ricerche Farmacologiche Mario Negri (Italy), KnowledgeMiner Software (Germany).

For questions and technical assistance please contact:
chm@kode-solutions.net

Janus is a software tool that allows to identify and prioritize chemicals which may be of concern for the environment and the human health, using a screening approach. Janus provides an assessment based on a conceptual scheme where different QSAR/QSPR models (from the VEGA platform) are integrated. The assessment focuses on the identification as PBT (persistent, bioaccumulative and toxic), CMR (carcinogen, mutagen and repro-toxicant) and ED (endocrine disruptor).

The final outcome is an overall prediction for each of these properties, combined with an estimate of their reliability, and a set of global scores based on such predictions, that allows to prioritize and screen the set of chemicals given as input.

The Janus tool has been developed as a JAVA stand-alone application, based on the JDK 8 platform. The application is a full stand-alone software and does not require any external network connection (no data are transmitted to network servers).

The application is provided with a simple graphical user interface (GUI) that allows the user to load a dataset of molecules, process it, and obtain the results in two main forms (plain text files or an interactive HTML application). Furthermore, a command line interface (CLI) is available, to allow the processing of big datasets; the CLI version can process molecules faster than the GUI, furthermore the GUI could have memory usage problems with big datasets.

The Janus application

Janus is a JAVA application that works on every operating system (Windows/Linux/Mac) with JAVA Runtime Environment (JRE) version 8 installed. The user should check that the latest JRE (freely available at http://www.java.com) is correctly installed in order to run Janus.

The application is provided as a zipped file; indeed no installation is needed. To start the application, the user should unzip the file and move to the "Janus" folder, then run the file Janus-launcher-WIN.bat (on Windows platforms) or Janus-launcher-LINUX.sh (on Linux platforms).

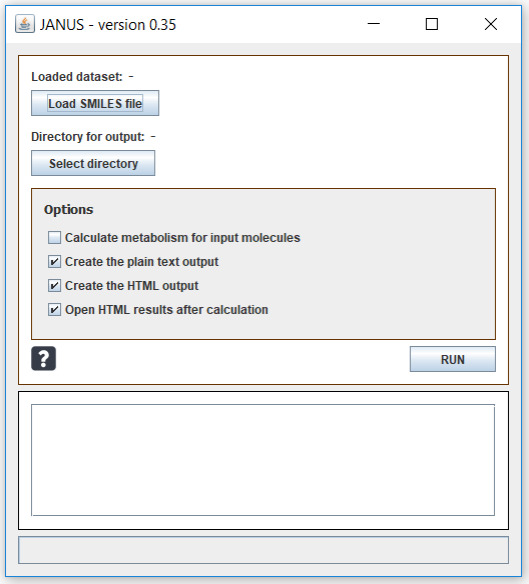
In case of problems with the above-mentioned launchers, it is possible to run the file Janus.jar with the Java Runtime Environment (JRE). On most systems, it is enough to double-click it, otherwise the user can open a command line window (like Command Prompt on Windows systems or BASH shell on Linux), move to the "Janus" folder and type:

java -jar Janus.jar

Janus may encounter problems when processing large datasets due to its memory usage. In such situation, it is possible to use the command line interface of the application. In addition, it is possible to modify and use the launcher files (Janus-launcher-WIN.bat and Janus-launcher-LINUX.sh) to assign more memory to Janus. In the launcher file modify the parameter "-Xmx1024m", where 1024m means that Janus uses 1,024 MB (1 GB), setting a higher number (for instance -Xmx2048m for running the application with 2 GB of memory).

The GUI application

Once started the GUI application, the user will find the main form where to choose the input molecule file, the output directory and the calculation options. Then, the calculation can be started. The following figure shows the main form



With the "Load SMILES file" button, the user should select a local file to be loaded, containing the input molecules as explained in the following section "Input files".

With the "Select directory" button, the user should select a local directory where the calculation output will be saved.

In the options box, the user can select the following options:

- "Calculate metabolism for input molecules": when checked, metabolism rules will be applied to the input molecules, and the found metabolites will be added to the calculation.

- "Create the plain text output": when checked, the results will be saved as plain text files in the selected output directory.

- "Create the HTML output": when checked the results will be saved as browsable HTML files in the selected output directory.

- "Open HTML results after calculation": when checked, the created HTML results will be automatically opened in the web browser.

With the "RUN" button, calculation is started; the log box and a progress bar in the lower part of the form will display the progress of the calculation.

Input files

The input files should be a plain text file containing a set of SMILES string, one on each line. No other information or line header should be added.

Optionally, it is possible to add one or more known experimental values for each molecule. The application will use these values to replace the predicted values that will be calculated. To add experimental values, it is necessary to add after each SMILES one or more strings, separated by the TAB character, in the form [KEY]:[VALUE], where the [KEY] is one of the following keys to define the type of end-point:

Key	**Description**
bcf	Experimental BCF value [log(kg/l)]
logp	Experimental Octanol-Water partition coefficient (LogP) value
p_soil	Experimental persistence in soil (half-life expressed in days)
p_water	Experimental persistence in water (half-life expressed in days)
p_sediment	Experimental persistence in sediment (half-life expressed in days)
tox	Experimental aquatic toxicity value (NOEC expressed in mg/l)
carc	Experimental carcinogenicty value [class 1/0]
muta	Experimental mutagenicity (Ames test) value [class 1/0]
repro	Experimental reproductive toxicity value [class 1/0]
ed	Experimental endocrine disruptor activity value [class 1/0]

and [VALUE] should be set with the experimental value, i.e. the quantitative value or the 1/0 classification in case of qualitative end-point as reported in the above table.

For example, the string:

c1ccccc1 bcf:2.3 carc:1

will load the SMILES c1ccccc1 and associate to this molecule a value of 2.3 log units for BCF and a carcinogenic classification.

Output files

The main output, created in the directory selected by the user, is a HTML browsable page. If the user selected the option, it will be automatically opened in the default browser after the calculation; it is anyway possible to manually go in the results folder and open the *index.html* file.

The main page of the output is a table with the summary of results for all input compounds. The table reports a molecule on each row, that can be corresponding to an input molecule, or a molecule derived by the application as a metabolite of an input molecule (in this case, the "metabolite" label will be shown. For each molecule, a colour-coded classification is reported for the PBT-CMRE properties; for each letter, the colour green means that the predicted value is under the threshold of concern for the property, orange means that the value is of concern, and red that the value is of strong concern. The following columns report the results of each single property, and can be shown or hidden (for better reading of the table) with the check-boxes in the upper-right part of the page. The last three columns contain the final scores (vPvB, SVHC and PBT).

All columns headers can be clicked, to sort the results according to the values in the chosen columns. By default, the compounds are shown with the original input order.

By clicking on the magnifying lens label on the left, the page with the details of the chosen molecule is shown. The page reports the depiction of the compound, the summary of the PBT-CMRE properties, and all the details for each property (including the output of all used VEGA models, and the intermediate values calculated inside the property workflow).

If the user selected the plain text option, the results will be saved as multiple text files in the selected output directory. All files are plain text files, where the TAB character is used to separate the fields. The file *janus_txt_main_results.txt* contains the summary of all calculations and final scores. The set of files named *janus_txt_results_details_XX.txt*, where XX will be one of the labels for the properties (P, B, T, C, M, R or ED) contain the details of the calculations for each separate property.

The Janus CLI

Janus can be run with its command line interface (CLI) to process large datasets more quickly and without any possible problem of memory usage.

The application should be run as a normal Java command line application, with the command:

java -jar Janus.jar -source=[SOURCE FILE] -dest=[DEST DIRECTORY] -metabolism=[yes/no]

The user should set the *[SOURCE FILE]* parameter with the input molecule files (full path and filename) and the *[DEST DIRECTORY]* with the path of the directory where the results will be saved. The *-metabolism* parameter can be used to set whether the metabolism should be calculated for the input molecules or not.

The results will be created only in their plain text form (no HTML is created with the CLI application).