Chemical Databases under *InstantJChem*

Data organisation

InstantJChem organizes data into a strict hierarchy called a *Data Tree*:

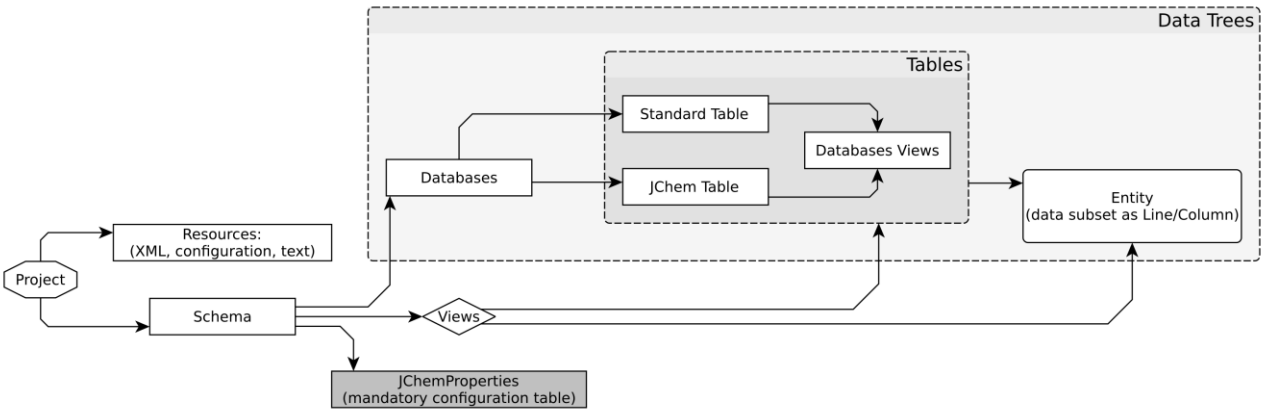

Figure 1: Different levels of data organisation in InstantJChem Refer to the text for explainations of the keywords.

Data are organized in **Projects**, which can contain both generic files and **Schema(s)**. All files placed to the project folder will appear in the graphical user interface as belonging to this project. These could be (*i*) configuration files for specific applications, (*ii*) molecular files used for the data exchange between different applications, and (*iii*) text files for comments. **Schema** could be considered as a layer between the graphical interface and a database management system (Derby, MySQL or Oracle). It contains all information to connect to your databases and manage their security. When creating a new **Schema**, a special table (named JChemProperties by default) is also created into the related database management system to store the JChem configuration.

Therefore, a **Schema** puts together the database management system, the databases and their tables and views. The database can contain two kinds of **tables**: *Standard* tables (containing only standard data types) and *JChem* tables containing molecular information. Molecular information includes structures (as molecular graph), calculated properties, etc.

On the top of these concepts, the software adds some more:

- **Database Views** which are new tables created dynamically by InstantJChem to store the results of queries that the user wants to make permanent;
- **Entities,** subsets of data shaped in the format of a spreadsheet (with lines and columns). They are a layer above the concept of tables in a database. In particular they provide access to the cartridge driving the chemical information into a standard database management system.
- **Data Trees**, a tool providing a tree representation of the data so that different entities/tables and fields can be managed in a unified way graphically.
- Finally, the **Views** are personalized forms or spread sheet like representation to access the data.

How to create a new database with InstantJchem

1. Click on **File**, **New Project**.
2. In the popup menu, select **IJC Project (empty)**, then click the **Next** button
3. In the form, fill the **Project Name**, **Project Location** fields, the **Project Folder** is generated according to the information in preceding fields.
4. Click on **Finish**.

The database is empty. Now it is needed to create at least one **Schema**.

How to create a local database with InstantJChem

1. Click on **File**, **New Project**.
2. In the popup menu, select **IJC Project (with local database)**, then click the **Next** button
3. In the next form, fill the **Project Name**, **Project Location** fields, the **Project Folder** is generated accroding to the information in preceding fields.
4. Click on **Finish**.

The database now contains a new schema.

How to add a Schema to a Project

1. Into the **Projects** window, right click onto the name of your project
2. In the popup menu, select **New Schema**.
3. In the following menu, select **Embedded Derby**, then click the **Next** button -Note: if you want to connect to an existing Oracle or MySQL database, it is needed to select them here.
4. In the following form, give a name in to the **New Schema name** text area, and then click **Finish**.

How to add a New Data Tree to a Schema

1. Into the **Projects** window, develop the content of you project
2. Into the content of your project, right click onto a **Schema**.
3. In the popup menu select **Edit Schema** or **Edit Data Tree**.
4. The right panel display the architecture of the projects: Data trees, Entities, Database Views, Databse Tables, Schema. Each item can be developed and managed.

How to edit a Schema or Data Trees

Into the **Projects** window, develop the content of you project
Into the content of your project, right click onto a **Schema** or into a schema right click on a **Data Tree**.
In the popup menu select **New Data Tree**.
In the next pop up menu, give a name, choose a root entity and a provide a short description.

How to import an SDF file into Schema

1. Into the **Projects** window, develop the content of you project
2. Into the content of your project, right click onto a **Schema**.
3. In the popup menu select **File import**.
4. In the next pop up menu, fill the **File to import** field.
5. Edit the other fields. Pay a particular attention to the **Table details** since it is the interface for many details of the management of chemical structures (mainly absolute stereo,

duplicates filtering, empty structures, fingerprints, standardization). Fields of the table are automatically detected.
6. Click **Next**.
7. Edit the list of fields of the table to be created and loaded from the SDF file, then click **Next**.
8. The SDF is loaded. Examine the monitoring of the import and click **Finish**.

How to delete a project or an element of the project

1. Right click on any element appearing into the Project window
2. Select Delete
3. In the popup window, it is recommended to require deletion of every related files and schema, otherwise *InstantJChem* or the database management system will continue to record part of it that might come across the way in subsequent operations.

How to view a database in a grid

1. Right click on a Data Tree.
2. Click on **New View**.
3. In the popup window, select **Default Grid View**.
4. Click on **Finish**.

How to browse a database in a form

1. Right click on a table.
2. Click on **New View**.
3. In the popup window, select **Empty Form View**.
4. Click on **Finish**. A **Design** mode is activated.
5. In the **Design** mode, add a molecule panel area and as many text area as required to view database fields.
6. Switch to **Browse** mode.

Many elements to the form can be added. It includes also labels, checkboxes for boolean properties, date pane for date fields, list for list fields, multiline text area and a table.

How to perform substructure search query

You need the *Query* window for this task. If it is not displayed, click on **Window** in the the menu bar and click on **Query Builder**.

1. Right click on the main blank area in the **Query** window.
2. Select a field of type molecule on which to perform a query. It generates an interface for molecular structure search.
3. Double click onto the blank area of this interface to open a sketcher and draw your query.
4. When finished click **Set Query.**
5. Select the **Substructure** key word into the rolling menu of the interface.
6. Click **Run Query**.

How to perform similarity search query

1. Right click on the main blank area in the **Query** window.
2. Select a field of type molecule on which to perform a query. It generates an interface for molecular structure search.
3. Double click onto the blank area of this interface to open a sketcher and draw your query.

4. When finished click **Set Query.**
5. Select the **Similarity** key word into the rolling menu of the interface.
6. Click on the **Options** button to access the similarity threshold.
7. Click **Run Query**.

The similarity search is based on the hashed fingerprint computed by InstantJChem during the loading of the structures. Settings of the fingerprints are a property of a chemically oriented entity. Clicking on the corresponding entity in the edition mode (accessed using **Edit Schema** or **Edit Data Tree**), the fingerprint can be checked. However, it can be setup only before the insertion of data: during a file import or prior the manual addition of the first compound in the database.

Note: the similarity search can't be performed on Markush structures.

How to combine queries

Queries are organised in a tree. Each node is a logical word AND/OR. Each leaf is a query. If there is only one query, it belongs to a default AND node.

1. Right Click on a logical node, on the main area in the **Query** window, to change it to OR or to add an OR/AND node to the tree.
2. Right click on an existing query element and in the popup menu click on **Add Field**.
3. Select a field on which to perform a query.
4. Customize the specific generated interface.
5. When finished click **Run Query.**

Note: any query element can be edited with right clicking

How to save a query results

1. When the query result is displayed in the main **Grid View**, click on **Lists**.
2. In the menu, select the **Save as List** option.
3. In the popup menu, select **Permanent** as **New list type** and **All rows** in the **Include in list** area.
4. Click **OK**.

How to combine query results

The **Lists and queries** tab gives access to the history of all queries prepared during a session. Items labelled **Temporary** are destroyed at the end of the session. To keep them for the future work the labels should be changed to **Permanent**.

1. Each query result must be saved in a List.
2. Click on the **Lists and queries** tab near the **Projects** tab.
3. Select some lists.
4. Click on the **Lists** item in the tool bar.
5. Develop the **List Operations** item and select an appropriate operation.

6. Configure the popup menu (check list orders, operation and resulting list saving options) and click **OK**.

How to perform a substructure search using lists of atoms

1. Set up a substructure search.
2. Edit the substructure.
3. In the sketcher, click on the **Periodic System** icon.
4. In the popup menu, click on the **Atom list** button.
5. While this button is activated, activate any combination of element button.
6. When setup, click the **Close** button.
7. Click on all atoms of the structure that should be replaced by the atom list.
8. When finished, click on the **Set Query**.
9. End the setup of the query and click **Run Query**.

How to perform a Chemical Term search

1. Right click on the main blank area in the **Query** window.
2. Select a field of type molecule on which to perform a query. It generates an interface for molecular structure search.
3. Click on the **Chem Term** button.
4. In the popup window select and edit the required **Chemical Term** (for instance in the **Favourites** list, the **Bioavailability**).
5. Click **Run Query**.

How to create an empty JChem table

1. Double click on the name of an existing Schema to open the connection.
2. Right click on the Schema and select in the menu **New Data Tree and structure entity**.
3. Right click on the created data tree and select **Edit Data Tree**.
4. The interface displays a data tree management view. Right click on the data tree and in the popup menu click on **New Standard Field** or the **New Chemical Terms Field**.
5. In the popup form that opens, select the field type
6. Configure the fields. Key properties are: the name, whether it is required or not, a default value.

How to create an empty Standard table

1. Double click on the name of an existing Schema to open the connection.
2. Right click on the Schema and select in the menu **New Data Tree and standard entity**.
3. Right click on the created data tree and select **Edit Data Tree**.
4. The interface displays a data tree management view. Right click on the data tree just created and in the popup menu click on the **New Standard Field**.

5. In the popup form that opens, select the field type
6. Configure the fields. Key properties are: the name, whether it is required or not, a default value.

How to add an item to existing table

1. Select a view of a database table. For instance select the **Grid View**.
2. Click on the **New row** icon.
3. In the dialog popup menu, fill all required informations: draw structure, give values for numerical and text fields.
4. Click **Add**.
5. If needed edit a new item and click **Add** again when it is setup
6. When finished, click **Close**.

How to create a relationship between tables (Many to One)

1. Open the **Edit Data Tree** interface. Both tables to be linked should be already present at this point.
2. The table being on the “Many” side, should be added a **Required** integer field to contain the table keys of the “One” side table.
3. Display the Entities and click on the **New Relationship** icon.
4. In the popup menu select, **New Simple Relationship**.
5. In the Basic tab, give **Name** to the new relationship, then select **Many to one** as the type of the relationship.
6. Tick the **Create DB constraints**.
7. Select the **From** table, that should be the “Many” side, and the relevant **Field** created before.
8. Select the **To** table, that should be the “One” side, and the relevant **Field** that should be the table's key.
9. In the **DB Constraints** tab, select the relevant rules in case of changes or deletion of entries in one or the other table. For instance select **Restrict** as the **On Delete** rule. Note: once a rule has been set, it can be hardly changed without deleting and recreating the relationship.
10. Click **Finish.**

Note that any relationship between tables in a database must be designed before loading the database. Otherwise, many complications are to be expected.