The steps for updating a new line list

The steps for updating a new line list

1. Use code (Jupyter Notebook) to check that the format of the line list is according to the Exomol standards.

   For the .states file (sort by states ID, if there is no lifetime, use the ExoCross program to add the lifetime into the .states file).

   For the .trans file (sort by frequency/wavenumber).

   Code GitHub: https://github.com/Beryl-Jingxin/update_new_linelists.git
2. Upload line list files (the .states, .trans files, .pf files and other related files) to the exoweb server.
3. Create def file (read the paper and use code (Jupyter Notebook) to get information) and upload the .def file to the exoweb server.
4. Add new line list information into the master file (exomol.all) and update the master file to the exoweb server.
5. Update information of line list to the ExoMol website.

The steps for updating new line list files to the exoweb server

1. Access the exoweb server and the file path:

   (1) Access the newt server: ssh0.theory.phys.ucl.ac.uk

   (2) Access the exoweb server from the newt server:

   [username@newt ~]$ ssh username@exoweb.projects.phys.ucl.ac.uk

   [username@exoweb ~]$ cd /mnt/data/exomol/exomol3_data

   In this file path, you can find the folders of all molecules, the file path of the line list files is in this format: molecule/Isotopologue/dataset/line list files.
2. The steps of uploading the new created .def file (small file) to the exoweb server:

   (1) Upload the created .def file to GitHub.

   (2) Open the .def file in the GitHub website and choose the bottom 'Raw'.

   Copy the link which is started with 'https://raw.githubusercontent.com/…'.

   Note: Use the link with 'raw' not the link of this Github webpage.

(3) In the exoweb server, use ‘wget https://raw.githubusercontent.com/…’ to download the created .def file from the GitHub website to the exoweb server.

3. The steps of uploading the new .states, .trans, .pf or their compressed files (large files) to the exoweb server. The following are the detailed steps for downloading files from Google Drive using the command line API, provided that the files are privately shared and require identity authentication.

   (1) Save these large line list files to Google Drive.

   (2) Obtain the file ID:

   Right-click (or directly click) the file or compressed file which needs to be uploaded to the exoweb server and select ‘Get Share Link’. The link takes the form of https://drive.google.com/open?id=XXXX, which XXXX is the file ID and will be used in the following steps.

   (3) Obtain the OAuth token:

   (a) Enter the OAuth 2.0 Playground.

   (b) In Step 1 Select & authorize APIs scroll box, select Drive API V3==, then select https://www.googleapis.com/auth/drive.readonly.

   (c) Click the button Authorize APIs and select Exchange authorization code for tokens. Get Access token and copy this Access token for use in subsequent steps.

   (4) Download the files with the commands:

   In the following commands, use XXXX to replace the file ID. And use the Access token to replace YYYYY. Use the saved file name (including the suffix, such as "myfile.zip") to replace ZZZZZ.

   (5) In the exoweb server (Linux system), open the terminal and enter the following command:

   curl -H "Authorization: Bearer YYYYY" https://www.googleapis.com/drive/v3/files/XXXXX?alt=media -o ZZZZZ

4. Update the existed master file (exomol.all) in the exoweb server (Linux system) directly or upload a new one to replace the old one.

   (1) If update the master file in the exoweb server directly, the steps of editing the master file are:

   vim exomol.all

   Press the ‘Insert’ key to enter edit mode. Press the ‘Esc’ key to exit edit mode. Press ‘Shift’ and ‘:’ together, then enter ‘wq’ to save the file and exit automatically.

   (2) When uploading a new master file to replace the old one, the steps are similar to uploading a new .def file (small file).

For Creating the .def file

If there already exists the .def file for the same molecule and isotopologue but other datasets, do some changes based on the existed .def file.

Where NIST website and Convert Dalton to Kilogram website are:

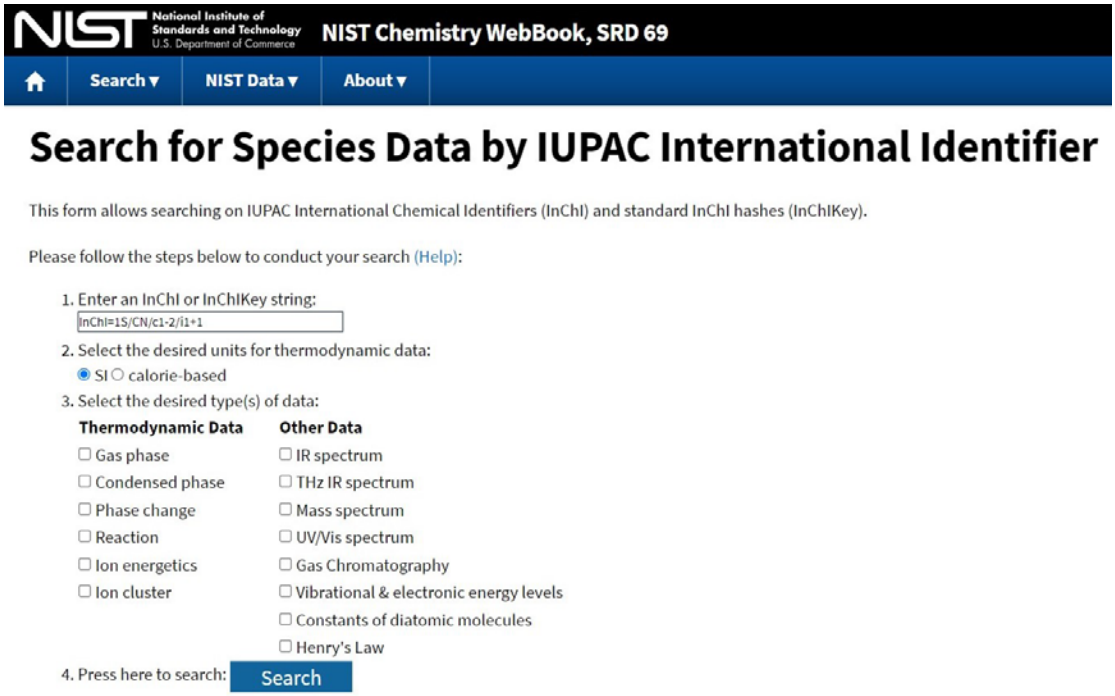
NIST website (IUPAC International Identifier):

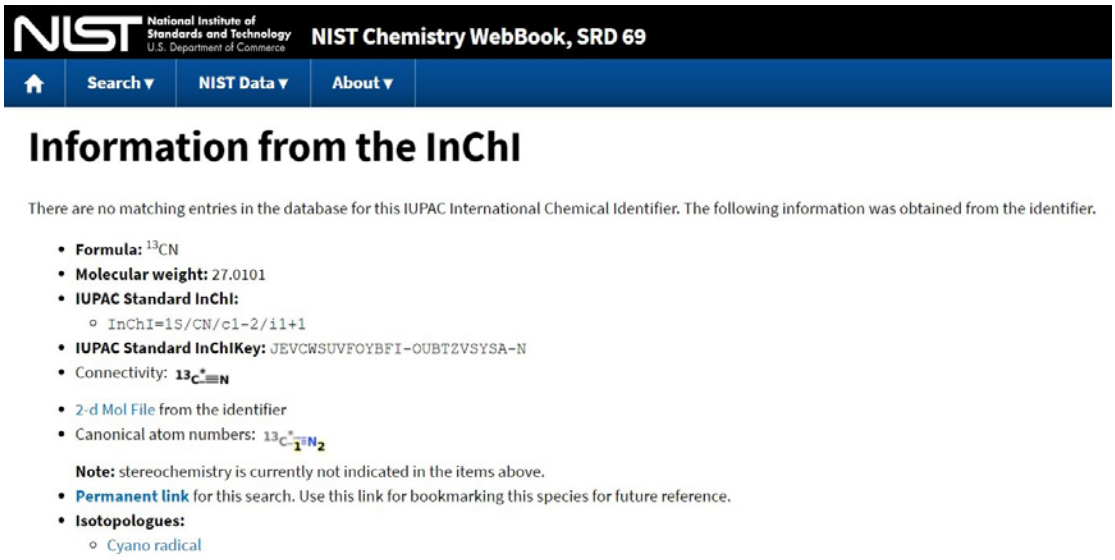
https://webbook.nist.gov/chemistry/inchi-ser/

UniConverters.net website (Convert Dalton to Kilogram):

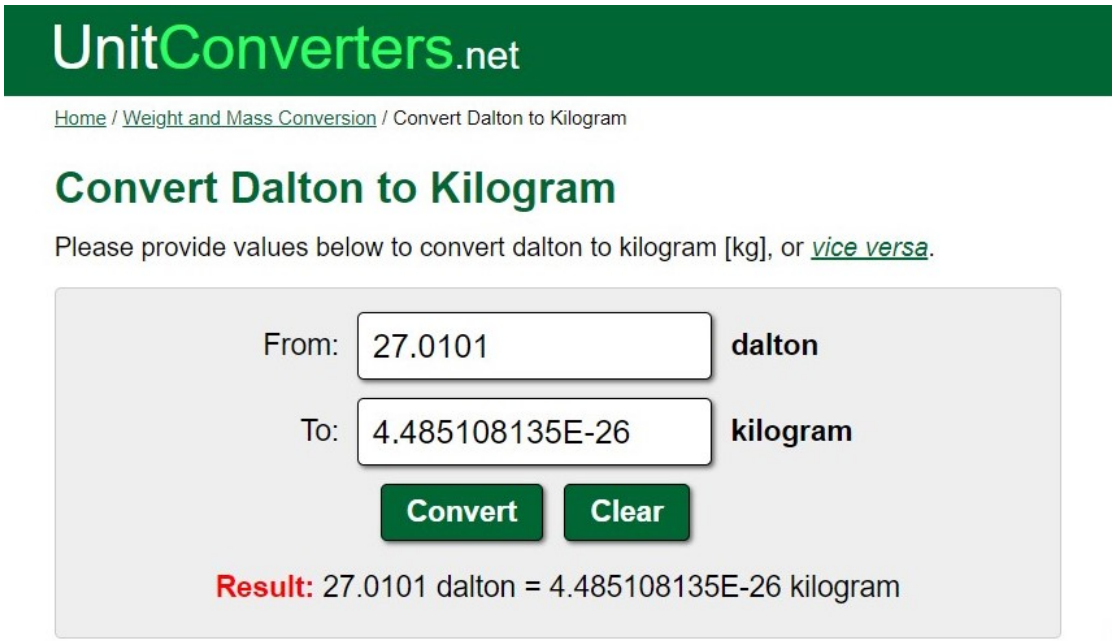
https://www.unitconverters.net/weight-and-mass/dalton-to-kilogram.htm

1. How to get the InChIKey for molecule (in the .def file, it is the InChIKey of isotopologue)?
   (1) You need to enter an InChI string of Isotopologue. The rule of finding InChI could be found from the following example, it is simple enough to find the InChI strings by yourself. For molecule CN:
   (12C)(14N): InChI=1S/CN/c1-2
   (12C)(15N): InChI=1S/CN/c1-2/i2+1
   (13C)(14N): InChI=1S/CN/c1-2/i1+1
   (13C)(15N): InChI=1S/CN/c1-2/i1+1,2+1
   (2) Open the NIST website, search by 'IUPAC Identifier'. Enter the InChI string on the NIST webpage. Then press 'Search'. Then we get the molecular weight (Isotopologue mass in Dalton) and IUPAC Standard InChIKey (InChIKey of molecule, note: it should be InChIKey of isotopologue).





2. Convert the mass from Dalton to Kilogram by Convert Dalton to Kilogram. Enter the mass you get from the NIST website and press ‘Convert’ to get the mass in kilogram.



3. For other information (some values), you can run the Jupyter Notebook and read the paper to get the information of the .states, .trans and .pf files.

For Updating the master file

Find the information from the .def file.

The steps for updating a new line list to the ExoMol website

Site Administration: https://exomol.com/admin/

The steps for updating new line list files on this website are:

1. Chems
    1) Molecules
    2) Isotopologues
2. Data
    1) Add data set
    2) Add link
    3) Add data collections
3. Spectroscopic Models
4. Pages → Activities → ExoMol updates

If you need to add new modules (maybe molecule, isotopologue, data set, link and others), press the top right corner ‘Add xxx’.

All the details when entering the information could be based on other existing samples.