Subject: Re: Importing compounds with assay data from Pubchem into Datawarrior
Posted by thomas on Fri, 23 Jul 2021 13:17:45 GMT
View Forum Message <> Reply to Message

Hi Jon,

unfortunately, I am afraid, I cannot help you much. I never fully explored what is needed to keep a fully updated copy of PubChem's bioactivity on a structure searchable server, but it always seemed to me that it would be beyond the effort, I would be willing to invest. I just tried to download a few data files from the PubChem website (pubchem.ncbi.nlm.nih.gov). Some download button scripts simply didn't do anything (Firefox on Ubuntu), but with most of them I could download either a CSV, SDF or gzipped SDF. All of those I could open successfully in DataWarrior ('.sdf.gz' need the newest dev update). Unless somebody else would volunteer to write the code to retrieve the PubChem bioassay data and merge it with the structures where needed and keep everything updated on a searchable server engine, I won't have the time to provide easy DataWarrior search access.

If you manage to download PubChem data files (csv or sdf) and have trouble reading them into DataWarrior because of limited csv/sdf capabilities, then please let me know. I would do my best to fix these issues on the DataWarrior side.

Thomas