Subject: Merge File Issues
Posted by amorrison on Wed, 17 Apr 2024 08:39:28 GMT
View Forum Message <> Reply to Message

Hi,
I recently updated to v6.01.03 and now seem to have an issue with merging data from another file.

The workflow I have is as follows. I have a DataWarrior file that I update with new compounds weekly in DataWarrior I use the advanced stereochemistry flags to indicate chirality. I then export this as an SDF file to calculate properties in another programme, which is then exported as a new SDF file.

As this application does not handle v3000 file format I lose the chiral flags in the structures. When I open the new annotated SDF file (which has the additional compounds) I have been replacing the structures (via merge) with the previously prepared ones in DW with chiral flag annotations. This has worked great until today. Now when I do this all the column filters are lost I can see from the row count they are there but the new compounds I have just added are hidden. Disabling all row filters and resetting all row filters has no effect.

Any help on this one?

Many thanks,

Angus

Subject: Re: Merge File Issues
Posted by nbehrnd on Wed, 17 Apr 2024 10:31:02 GMT
View Forum Message <> Reply to Message

Hello Angus,

can you afford to drop `&` and `or`, the extended stereochemistry of .sdf files (V3000) and subsequently work with the elder syntax of version 2000 instead? The .dwam macro file attached below will export the DW structure array, and exports the subsequent columns as a tag trailing each entry. The file format can be re-read by DW, and hence be serve to merge with an other file, too.

Norwid

File Attachments
1) export_sdf_v2000.dwam, downloaded 583 times

Subject: Re: Merge File Issues

Posted by thomas on Tue, 23 Apr 2024 08:54:26 GMT
View Forum Message <> Reply to Message

Hi Angus,

for testing I constructed two files:
test1.dwar: 'Structure' column with enhanced stereo structures, 'ID' column with unique identifiers
test2.dwar: 'Structure' column with simple stereo structures, 'ID' column with unique identifiers, some property columns

I then merged test1.dwar into test2.dwar and independently test2.dwar into test1.dwar, in both cases such that the enhanced stereo structures would be in the final file. Both attempts worked without an issue.

Could you send me two small test files where the merge problem exists? Ideally with a small macro that contains the merge task? I need to reproduce to fix the issue.

Thanks in advance,

Thomas

Subject: Re: Merge File Issues
Posted by amorrison on Tue, 23 Apr 2024 16:55:54 GMT
View Forum Message <> Reply to Message

Hi Norwid,
Many thanks for the very useful suggestion. I'm trying to keep the chiral annotations (&, or, abs). To be fair, I have not tried exporting it as a v2000 file format and trying the merge in this case.
Best wishes,
Angus

Subject: Re: Merge File Issues
Posted by amorrison on Tue, 23 Apr 2024 17:16:01 GMT
View Forum Message <> Reply to Message

Hi Thomas,
Was not sure how to do this as a macro, so apologies but hopefully you can follow what I am doing. I've attached some files for a small set of compounds from ChEMBL.

Step 1 - Export DataWarrior file CDK2_Dwar.dwar to an sdf file (v3000).
Step 2 - Calculate properties in another program (ACD) and export as a SDF file (CDK2_Export_ACD.sdf)
Step 3 - Open CDK2_Export_ACD.sdf in DataWarrior - Say no to the racemate warning
Step 4 - Merge CDK2_Dwar.dwar in to CDK2_Export_ACD.sdf with Compound_ID as merge key and Structure 'replace with new'

All row filters are then lost and only 3 structures are visible, although the row count is 26. As part of this test I (have just) saved the resulting file and have just reopened it and the structures are now visible although all the row filters are still missing.

Apologies for the long winded explanation but hopefully you get the gist of what I am trying to do.

Best wishes,

Angus

File Attachments
1) CDK2_Dwar.dwar, downloaded 558 times
2) CDK2_Export_ACD.sdf, downloaded 552 times

Subject: Re: Merge File Issues
Posted by thomas on Wed, 24 Apr 2024 11:49:42 GMT
View Forum Message <> Reply to Message

Hi Angus,

many thanks for the files. I could now reproduce the issue, which was caused by an old bug where existing columns were not re-analyzed after merging, when their values updated. Then, DataWarrior used outdated category lists for these columns resulting in Exceptions, when a category filter was applied.

This is now fixed and will be deployed in the next online-update.

Best wishes,

Thomas
