Subject: Converting COD2020 to SDF
Posted by ghutchis on Thu, 28 Jan 2021 01:46:45 GMT
View Forum Message <> Reply to Message

I saw there was a new update to the COD data. Since I'm working with SD files mostly, I tried to export the 3D geometries.

But on my MacBook running Big Sur, it hangs with 3D export. A window pops up, but it's empty and never updates.

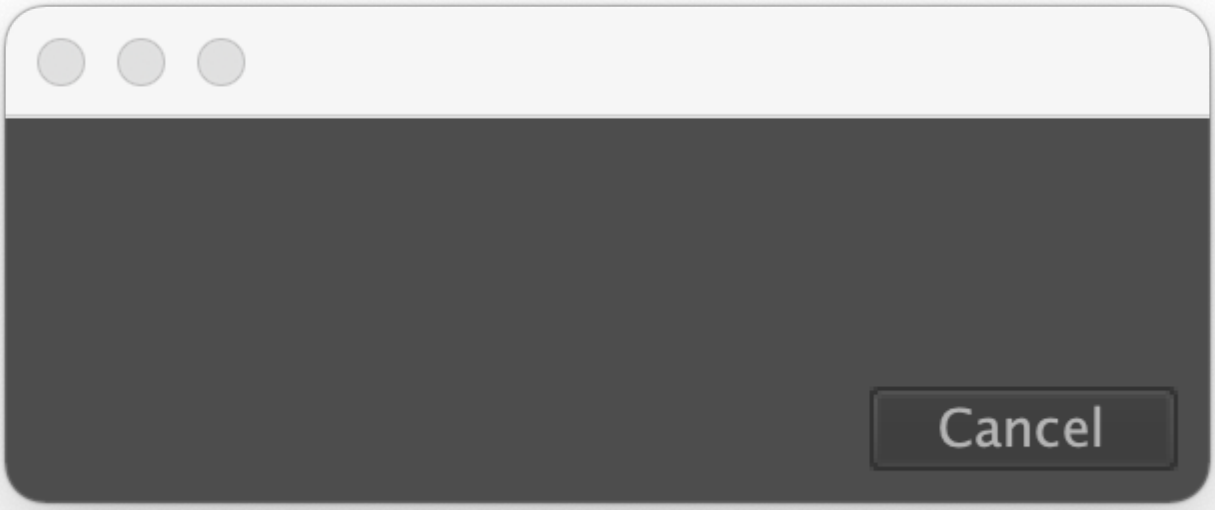
Export to 2D works great - no matter the options I pick in SD export.

I'm happy to help debug - I periodically get "Uncaught Exception: null" errors with the COD2020 file open.

I see a few errors in the system.log but these seem rendering-related:
Jan 27 20:19:11 Mercury DataWarrior[17344]: getattrlist failed for /System/Library/Extensions/AppleIntelKBLGraphicsGLDriver.bundle/Contents/MacOS/AppleIntelKBLGraphicsGLDriver: #2: No such file or directory
Jan 27 20:19:11 Mercury DataWarrior[17344]: getattrlist failed for /System/Library/Frameworks/OpenGL.framework/Resources//GLRendererFloat.bundle/GLRendererFloat: #2: No such file or directory
Jan 27 20:21:17 Mercury com.apple.xpc.launchd[1]: Coalition Cache Hit: app<application.org.openmolecules.datawarrior.122966090.12296611 6(503) > [4385]
Jan 27 20:21:18 Mercury DataWarrior[17395]: getattrlist failed for /System/Library/Extensions/AppleIntelKBLGraphicsGLDriver.bundle/Contents/MacOS/AppleIntelKBLGraphicsGLDriver: #2: No such file or directory
Jan 27 20:21:18 Mercury DataWarrior[17395]: getattrlist failed for /System/Library/Frameworks/OpenGL.framework/Resources//GLRendererFloat.bundle/GLRendererFloat: #2: No such file or directory

File Attachments

1) Screen Shot 2021-01-27 at 8.31.15 PM.png, downloaded 1099 times



Subject: Re: Converting COD2020 to SDF
Posted by thomas on Thu, 28 Jan 2021 08:58:04 GMT
View Forum Message <> Reply to Message

I remember that there was a bug sometime ago with the SD-file export, which should be fixed now. I have just confirmed on my MacBookPro running Catalina, that importing the July 2020 COD file and exporting as SD-file works without obvious problems.

I assume that you are not running the official 5.2.1 version. An official DataWarrior update is due end of February, but for the meantime I suggest that you download the current development version via openmolecules.org/datawarrior/dw521x.zip, unpack the zip file and replace the original datawarrior.jar file in /Applications/DataWarrior.app/Contents/Java with the new one. You may also put the new macro and worldfactbook into their folders. That should fix this and other problems and adds new functionality.

Please let me know, if you discover other issues. Especially, Mac specific specific problems sometimes slip my attention.

Subject: Re: Converting COD2020 to SDF
Posted by nbehrnd on Thu, 28 Jan 2021 14:36:13 GMT
View Forum Message <> Reply to Message

In case only a few structures are needed, and you are comfortable with the terminal / CLI you

might consider the following:

Each COD dataset has its COD ID, a number which may be seen in DataWarrior and which may be used to
access the dataset e.g. on COD's text-based form
https://www.crystallography.net/cod/search.html.

The entry «cif» on COD's listed results directs you to the structure's model data set, which you may store on your computer. Among the cod-tools (https://wiki.crystallography.net/cod-tools/) which you may obtain as an archive, or as a bundle package (e.g., Linux Debian, or Ubuntu) is the tool codcif2sdf for the CLI / terminal, which offers the conversion of COD's .cif into the .sdf file. To run successfully, you need OpenBabel which equally is freely available.

The output may be redirected into a permanent record, e.g. calling

codcif2sdf 1505213.cif > example.sdf

Both a COD cif as well its conversion into the .sdf of a typical entry are attached below.

In the past, the conversion a COD .cif into a .sdf only with OpenBabel often was not as good as with codcif2sdf. Alternatively, use a visual program capable to read and write both file formats (e.g., Jmol, or CCDC's Mercury)

Norwid

File Attachments

1) 1505213.cif, downloaded 844 times
2) example.sdf, downloaded 839 times

Subject: Re: Converting COD2020 to SDF
Posted by ghutchis on Thu, 28 Jan 2021 16:00:08 GMT
View Forum Message <> Reply to Message

On my laptop at home, I had the most recent version - but I replaced the JAR and everything seems to be running smoothly.

I'm aware of the codcif2sdf tools - I'm the maintainer of Open Babel.. I also remember that Thomas mentioned to me that he has done some additional sanity checks in hiss version.

Perhaps it's a separate question - how is the DWAR compiled from COD? Do you take the codcif2sdf SDF and pull them together into the DWAR or something else?

(Prompted by a few people, I'm working to put up a resource that rsync's from COD and provides the SDF.

Subject: Re: Converting COD2020 to SDF
Posted by nbehrnd on Thu, 28 Jan 2021 21:56:31 GMT
View Forum Message <> Reply to Message

Possibly codcif2sdf is part of the tools used for at least two reasons: Using only Openbabel (including release 3.1.0) would not retain the bibliographic information in the .sdf written which show up in DW's display.

An other reason is sometimes both crystallographic motif as well as symmetry of the unit cell are needed to complete the molecule's appearance as one would draw it in a notebook. By choice of the unit cell this may look like the molecule were broken. And it is the reconstruction of intramolecular atom connectivity where OpenBabel may more frequently encounter problems, than codcif2sdf. As an illustration, I attach both COD's .cif about a simple triazine, as well as the rewritten form as .sdf, once by OpenBabel, once by codcif2sdf.

Once the individual entries are rewritten for the organic chemist's eye, they may be stacked into a .sdf of multiple models and read as-such by DataWarrior. Possibly Thomas or/and the maintainer's of COD developed some template scripts to automate the whole process including formatting, and checks for plausibility and consistency. Given the scope and coverage of COD and TCOD often complementary to the ones e.g., by CCDC and ICSD, additional mirrors most likely are welcomed enthusiastically.

Norwid

TCOD, the sibling of COD mentioned: https://www.crystallography.net/tcod/

File Attachments

1) 7050214.cif, downloaded 792 times
2) triazene_obabel.sdf, downloaded 803 times
3) triazene_cod.sdf, downloaded 781 times

Subject: Re: Converting COD2020 to SDF
Posted by thomas on Fri, 29 Jan 2021 08:36:13 GMT
View Forum Message <> Reply to Message

Some years ago, when we needed statistical data from crystallographic data for our conformer generation and forcefield algorithms in OpenChemLib, and after getting a veto from Colin Groom to use the CSD for that purpose, we looked for alternatives. As a result we invited one of the COD maintainers (Antanas Vaitkus) to spend parts of his PhD time at Actelion in Switzerland to work on the cif2sdf conversion and especially to improve the calculation of bonds from atom coordinates, which has some issues especially with organo-metallic structures. He did a marvelous job improving the bond calculation logic by looking into lots of original papers, added validation code to produc warnings and errors. After returning to Vilnius he established the conversion as a regularly occurring process that not only creates an SD-file, but also creates a dwar file in their SVN repository.

We download the dwar, remove fishy structures, apply some minor changes, add the

organic/metal-organic/inorganic classification, apply a template, and put it on this website for download.

Thomas

Subject: Re: Converting COD2020 to SDF
Posted by nbehrnd on Fri, 29 Jan 2021 18:59:25 GMT
View Forum Message <> Reply to Message

Coincidental side note: Crystallographic data as basis to set up force fields is a centre of gravity of Hofmann's chapter about small organic molecules in «Data Mining in Crystallography» within the series «Structure and Bonding»:

https://link.springer.com/chapter/10.1007/978-3-642-04759-6_ 4

Norwid
