Global Pillage: Downloading Chemical Data Sets for Arthor and SmallWorld
Version: May 2022

BindingDB
Download Page: https://www.bindingdb.org/bind/chemsearch/marvin/SDFdownload.jsp?all_download=yes
Download URL:
https://www.bindingdb.org/bind/downloads/BindingDB_All_terse_2D_2021m3.sdf.zip

ChEBI
Download Page: https://www.ebi.ac.uk/chebi/downloadsForward.do
Download URL: ftp://ftp.ebi.ac.uk/pub/databases/chebi/SDF/ChEBI_lite_3star.sdf.gz
obabel ChEBI_lite_3star.sdf.gz --title "" --append "ChEBI_ID" -O ChEBI_lite_3star.smi

ChEMBL
Option 1
https://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/latest/chembl_28_chemreps.txt.gz
zcat chembl_28_chemreps.txt.gz | tail +2 | awk '{ print $2 "\t" $1 }' > chembl_28.smi
Configuration URL: https://www.ebi.ac.uk/chembl/compound_report_card/%s/
Option 2
https://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/latest/chembl_28.sdf.gz
obabel chembl_28.sdf.gz -O chembl_28.smi
Configuration URL: https://www.ebi.ac.uk/chembl/compound_report_card/%s/

ChemSpace [registration required]
Download Page: https://chem-space.com/compounds

Coconut (Natural Products)
Download Page: https://coconut.naturalproducts.net/download
Download URL: https://coconut.naturalproducts.net/download/absolutesmiles

Crystallography Open Database (COD)
Download URL: http://www.crystallography.net/cod/smi/allcod.smi
Configuration URL: http://www.crystallography.net/cod/%s.html

DSSTox/CompTox (EPA)
Download Page: https://comptox.epa.gov/dashboard/downloads
Download URL:
ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard/DSSTox_SDF_File_20160720.sdf.gz
Download URL: https://figshare.com/projects/NCCT_Chemistry_Dashboard_Data/32198

eMolecules
Download Page: https://www.emolecules.com/info/products-data-downloads.html
Download URL: https://downloads.emolecules.com/free/2021-04-01/version.smi.gz
zcat version.smi.gz | awk '{ print $1 "\t" $2 }' > eMolecules20210401.smi

Enamine REAL [registration required]
Download Page: https://enamine.net/compound-collections/real-compounds/real-database#

FragranceDB (SuperScent + Flavornet)
Download Page: https://gdb.unibe.ch/downloads/
Download URL: http://130.92.106.217:8080/FRAG_SETS/FragranceDB.smi

GDB-13
Download Page: https://gdb.unibe.ch/downloads/
Download URL: http://gdbtools.unibe.ch:8080/cdn/gdb13.tgz

LIPID MAPS Structure Database (LMSD)
Download Page: https://www.lipidmaps.org/data/structure/download.php
Download URL: https://www.lipidmaps.org/files/?file=LMSD&ext=sdf.zip

LOTUS (Natural Products)
Download Page: https://lotus.naturalproducts.net/download
Download URL: https://lotus.naturalproducts.net/download

Mcule Purchasable
Download Page: https://mcule.com/database/
Download Page: https://downloads.emolecules.com/free/
Download URL:
https://mcule.s3.amazonaws.com/database/mcule_purchasable_full_210414.smi.gz
gunzip mcule_purchasable_full_210414.smi.gz
Configuration URL: https://mcule.com/%s

MolPort [registration required]
Download Page: https://www.molport.com/shop/database-download

NCI DTP Database
Download Page: https://cactus.nci.nih.gov/download/nci/
Download URL: https://cactus.nci.nih.gov/download/nci/NCI-Open_2012-05-01.sdf.gz
Download URL: https://cactus.nci.nih.gov/download/nci/NCI-Open_09-03.sdf.gz
Download URL: https://cactus.nci.nih.gov/download/nci/NCI_aug00_2D.sdz
Download URL: https://cactus.nci.nih.gov/download/nci/NCISMA99.sdz

PDB Ligand Expo
Download Page: http://ligand-expo.rcsb.org/ld-download.html
Download URL: http://ligand-expo.rcsb.org/dictionaries/Components-smiles-stereo-oe.smi

PubChem Compound
Download Page: https://pubchemdocs.ncbi.nlm.nih.gov/downloads
Download URL: https://ftp.ncbi.nlm.nih.gov/pubchem/Compound/Extras/CID-SMILES.gz
zcat CID-SMILES.gz | awk '{ print $2 "\t" $1 }' > pubchem_compound.smi

PubChem Substance
Download Page: https://pubchemdocs.ncbi.nlm.nih.gov/downloads
Download URL: https://ftp.ncbi.nlm.nih.gov/pubchem/Substance/CURRENT-Full/SDF/

SAVI Database (NCI)
Download Page: https://cactus.nci.nih.gov/download/savi_download/
Download URL:

SureChEMBL
Download URL: https://ftp.ebi.ac.uk/pub/databases/chembl/SureChEMBL/data/

TasteDB (SuperSweet + BitterDB)
Download Page: https://gdb.unibe.ch/downloads/
Download URL: http://130.92.106.217:8080/FRAG_SETS/TasteDB.smi

USPTO Reactions
Download Page: https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873
Download URL: https://ndownloader.figshare.com/files/8664370
Download URL: https://ndownloader.figshare.com/files/8664379

VEHICLe
Download Page: http://ftp.ebi.ac.uk/pub/databases/chembl/VEHICLe/
Download URL: http://ftp.ebi.ac.uk/pub/databases/chembl/VEHICLe/VEHICLe.csv
tail +2 VEHICLe.csv | awk -F, '{ print $2 "\t" $1 }' > VEHICLe.smi

ZINC
Download Page: https://zinc20.docking.org/tranches/home/
Configuration URL: https://zinc15.docking.org/substances/%s/

Appendix: Pre-requisites
On RedHat/CentOS: sudo yum install openbabel