

US EPA TOXCAST DATA RELEASE JULY 2020

Database & tcpl R Package

This file describes the contents of the July 2020 ToxCast data release. The database/R package contents of the download contain the following files:

- [1] "Data_processing.html"
- [2] "Data_retrieval.html"
- [3] "Introduction_Appendices.html"
- [4] "prod_internal_invitrodb_v3_3.zip"
- [5] "README.pdf"
- [6] "tcpl_2.1.0.tar.gz"

Loading the Database

System Requirements:

- MySQL 5.7 (either on your laptop/desktop or on a remote server)
 - *Note: if the MySQL 5.7 server is on a remote server, you still need to download and install MySQL 5.7 on your laptop/desktop to use client tools to connect to the server*
 - Download: <https://dev.mysql.com/downloads/mysql/5.7.html>
- Minimum 200GB free disk space
 - *Note: prod_internal_invitrodb_v3_3.zip unzips an 80GB .sql file*
- Minimum 4GB RAM (8GB+ recommended)

Instructions

1. Unzip prod_internal_invitrodb_v3_3.zip
2. Open "Command Prompt" with *Run as administrator*
3. Run the following command to log into the server (will prompt for password):

```
mysql -h my_host_name -u my_user_name -p
```

4. Create the database schema that will hold the invitrodb data. Run the following command:

```
CREATE DATABASE invitrodb;
```

5. Exit from the MySQL server by running \q
6. Running the following Windows Command Prompt command now that the database is created (this will load the SQL file:

```
mysql -h my_host_name -u my_user_name -p invitrodb < prod_internal_invitrodb_v3_3.zip
```

7. Keep the command prompt window open until the command finishes loading the file (approx.. 8 hours)

Additional Notes & Tips

Because this is a large database, we recommend minimizing use of other applications while loading this if it is your personal laptop/desktop. If you are on Windows, it will be helpful to monitor to the mysql process in the Windows Task Manager (CTRL+ALT DELETE, click "Task Manager"), to periodically monitor RAM and

CPU usage. If you are a more advanced user, we also recommend running the process in the background. See here for instructions on how to do this in Windows: <https://docs.microsoft.com/en-us/windows-server/administration/windows-commands/start>.

For questions or concerns, please contact Jason Brown at: brown.jason@epa.gov.

TCPL package installation

To install the package on Windows the user will first need to properly install R tools, enabling R to compile the source package. The package relies on R ($\geq 3.1.0$), and the following packages: data.table ($\geq 1.9.4$), DBI, MASS, numDeriv, RColorBrewer. To interact with the MySQL database, the user also needs the “RMySQL” package. The “RMySQL” package requires a local MySQL server, although it does not require the server to be running. More information about installing RMySQL on Windows can be found online.

The “tcpl_2.1.tar.gz” file is the beta version of the R package used to process all of the data in this release. The package can be installed using the following R command:

```
install.packages("tcpl_2.1.0.tar.gz", type = "source", repos = NULL)
```

The html files are the R package vignette, detailing the processing, and showing some examples of how to use the package. The appendices vignette also includes an appendix listing the database field names.

For questions or concerns, please contact Jason Brown at: brown.jason@epa.gov.