

US EPA TOXCAST DATA RELEASE OCTOBER 2015

Database & tcpl R Package

This file describes the contents of the October 2015 ToxCast data release. The database/R package contents of the download contain the following files:

- [1] "prod_external_invitrodb_v2.all.zip"
- [2] "prod_external_invitrodb_v2.bare.sql"
- [3] "Pipeline_Overview.pdf"
- [4] "tcpl_1.0.tar.gz"

To utilize the database, the user must have write access to a MySQL server, allowing them to install the database contained in the "prod_external_invitrodb_v2.all.zip" file. The "prod_external_invitrodb_v2.all.zip" is a zipped sql file containing the entire invitrodb_v2 database, whereas the "prod_external_invitrodb_v2.bare.sql" contains just the database schema and tcpl methods pre-loaded. To avoid errors while calling large queries we recommend the MySQL server setting, 'max_allowed_packet', is set to at least 1 GB.

The "tcpl_1.0.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_1.0.tar.gz", type = "source", repos = NULL)
```

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 (>= 3.1.0), and the following packages: data.table (>=1.9.4), DBI, RSQLite, 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 "Pipeline_Overview.pdf" file is the R package vignette, detailing the processing, and showing some examples of how to use the package. The vignette also includes an appendix listing the database field names.

For questions or concerns, please contact Monica Linnenbrink at: linnenbrink.monica@epa.gov.