

iemiscdata: USEPA East Palestine, Ohio Norfolk Southern Train 32N Cargo List – Chemical Databases Match

Irucka Embry, E.I.T. (EcoC²S)

2023-09-25

Contents

| | |
|--------------------------------|---|
| Available Chemical Information | 1 |
| EcoC ² S Links | 4 |
| Copyright and License | 5 |

Available Chemical Information

The following provides chemical information from either the `chem_wiki` or `atsdr_tsca_ld50_a` data sets from the `chem.databases` R package created by the author for matching chemical substances in the commodities list from the United States Environmental Protection Agency (US EPA) East Palestine, Ohio Norfolk Southern Train 32N Cargo List. The cargo list is one of the data sets in this `iemiscdata` R package also created by the author.

```
install.load::load_package("iemiscdata", "chem.databases", "data.table", "mgsub",  
  "stringi", "pander")  
# load needed packages using the load_package function from the install.load  
# package (it is assumed that you have already installed these packages)
```

```
data(norfolk_southern_epoh)  
# from iemiscdata package
```

```
data(chem_wiki)  
# from chem.databases package
```

```
data(atsdr_tsca_ld50_a)  
# from chem.databases package
```

```
train_commodities <- unique(norfolk_southern_epoh$Commodity)
```

```
pander(train_commodities)
```

```
POLYPROPYLENE, POLYETHYLENE, residue lube oil, VINYL CHLORIDE, STABILIZED ,  
DIPROPYLENE GLYCOL, PROPYLENE GLYCOL, DIETHYLENE GLYCOL, COMBUSTIBLE
```

```

LIQ., NOS (ETHYLENE GLYCOL MONOBUTYL ETHER) , SEMOLINA, COMBUSTIBLE LIQ., NOS
(ETHYLHEXYL ACRYLATE) , POLYVINYL, PETROLEUM LUBE OIL, POLYPROPYL GLYCOL,
ISOBUTYLENE, BUTYL ACRYLATES, STABILIZED , PETRO OIL, NEC , ADDITIVES, FUEL
, BALLS,CTN,MEDCL , SHEET STEEL, VEGETABLE, FROZEN , BENZENE, PARAFFIN WAX,
FLAKES, POWDER , HYDRAULIC CEMENT, AUTOS PASSENGER and MALT LIQUORS

# remove items that are not needed

train_commodities <- train_commodities[-c(3, 9, 16:20, 23:26)]

# remove words that are not needed

train_commodities <- mgsub(train_commodities, c(", STABILIZED", "COMBUSTIBLE LIQ., NOS",
"LUBE OIL", "WAX"), rep("", 4), fixed = TRUE)

# remove parentheses that are not needed

train_commodities <- mgsub(train_commodities, c("\\(", "\\)"), rep("", 2), fixed = FALSE)

# remove the beginning and trailing white space

train_commodities <- stri_trim_both(train_commodities)

# transform the text to sentence case

train_commodities <- stri_trans_totitle(train_commodities, type = "sentence")

# correct the spelling as needed

train_commodities[4] <- "Dipropylene glycol"

train_commodities[9] <- "Polyvinyl chloride"

train_commodities[13] <- "Butyl acrylate"

pander(train_commodities)

Polypropylene, Polyethylene, Vinyl chloride, Dipropylene glycol, Propylene glycol, Diethylene glycol, Ethylene
glycol monobutyl ether, Ethylhexyl acrylate, Polyvinyl chloride, Petroleum, Polypropyl glycol, Isobutylene,
Butyl acrylate, Benzene and Paraffin

# the exact matched chemical names

pander(chem_wiki[Substance Name` %in% train_commodities])

```

Table 1: Table continues below

| CAS | Substance Name | IUPAC Name | Molecular Formula |
|------------|--------------------|------------|---|
| 9002-86-2 | Polyvinyl chloride | NA | NA |
| 25265-71-8 | Dipropylene glycol | NA | C ₆ H ₁₄ O ₃ |

| CAS | Substance Name | IUPAC Name | Molecular Formula |
|-----------|-------------------|------------------------|-------------------|
| 111-46-6 | Diethylene glycol | 2,2'-Oxydi(ethan-1-ol) | C4H10O3 |
| 9003-07-0 | Polypropylene | NA | NA |
| 141-32-2 | Butyl acrylate | Butyl prop-2-enoate | C7H12O2 |
| 71-43-2 | Benzene | Benzene | C6H6 |
| 75-01-4 | Vinyl chloride | Chloroethene | C2H3Cl |

Table 2: Table continues below

| SMILES |
|--|
| NA |
| C.C.OCCOCCO lp:4:2,7:2,10:2,m:1:5,6,3:8.9 |
| OCCOCCO |
| NA |
| CCCCOC(=O)C=C |
| C1=CC=CC=C1 |
| ClC=C |

Table 3: Table continues below

| InChI |
|---|
| NA |
| NA |
| InChI=1S/C4H10O3/c5-1-3-7-4-2-6/h5-6H,1-4H2 |
| NA |
| InChI=1S/C7H12O2/c1-3-5-6-9-7(8)4-2/h4H,2-3,5-6H2,1H3 |
| InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H |
| InChI=1S/C2H3Cl/c1-2-3/h2H,1H2 |

| InChIKey | Average Mass | Monoisotopic Mass |
|-----------------------------|--------------|-------------------|
| NA | NA | NA |
| NA | 134.2 | 134.1 |
| MTHSVFCYNBDYFN-UHFFFAOYSA-N | 106.1 | 106.1 |
| NA | NA | NA |
| CQEYYJKEWSMYFG-UHFFFAOYSA-N | 128.2 | 128.1 |
| UHOVQNZJYSORNB-UHFFFAOYSA-N | 78.11 | 78.05 |
| BZHJMEDXRYGGRV-UHFFFAOYSA-N | 62.5 | 61.99 |

```
pander(atsdr_tsca_ld50_a[Registry Name` %in% train_commodities])
```

| CAS | Substance Name | Registry Name | SMILES |
|----------|--------------------|---------------------------------|----------|
| 111-76-2 | Ethanol, 2-butoxy- | Ethylene glycol monobutyl ether | CCCCOCCO |
| 57-55-6 | 1,2-Propanediol | Propylene glycol | CC(O)CO |

| CAS | Substance Name | Registry Name | SMILES |
|-----------|-----------------|----------------|----------|
| 71-43-2 | Benzene | Benzene | c1ccccc1 |
| 75-01-4 | Ethene, chloro- | Vinyl chloride | C=CCl |
| 8002-05-9 | Petroleum | Petroleum | NA |

```
# the matched names
```

```
train_match1 <- chem_wiki[`Substance Name` %in% train_commodities]
```

```
train_match2 <- atsd_r_tscs_ld50_a[`Registry Name` %in% train_commodities]
```

```
train_matches1 <- train_match1$"Substance Name"
```

```
train_matches2 <- train_match2$"Registry Name"
```

```
train_match <- unique(c(train_matches1, train_matches2))
```

```
pander(train_match)
```

Polyvinyl chloride, Dipropylene glycol, Diethylene glycol, Polypropylene, Butyl acrylate, Benzene, Vinyl chloride, Ethylene glycol monobutyl ether, Propylene glycol and Petroleum

```
train_match_cas1 <- train_match1$CAS
```

```
train_match_cas2 <- train_match2$CAS
```

```
train_match_cas <- unique(c(train_match_cas1, train_match_cas2))
```

```
pander(train_match_cas)
```

9002-86-2, 25265-71-8, 111-46-6, 9003-07-0, 141-32-2, 71-43-2, 75-01-4, 111-76-2, 57-55-6 and 8002-05-9

```
# remove the matched names
```

```
train_commodities2 <- train_commodities[-which(train_commodities %in% train_match)]
```

```
# these are the chemical substances without an exact match
```

```
pander(train_commodities2)
```

Polyethylene, Ethylhexyl acrylate, Polypropyl glycol, Isobutylene and Paraffin

EcoC²S Links

EcoC²S Home – <https://www.ecoccs.com/>

About EcoC²S – https://www.ecoccs.com/about_ecoc2s.html

Services – <https://www.ecoccs.com/services.html>

1 Stop Shop – https://www.ecoccs.com/other_biz.html

Products – <https://www.questionuniverse.com/products.html>

Media – <https://www.ecoccs.com/media.html>

Resources – <https://www.ecoccs.com/resources.html>

R Trainings and Resources provided by EcoC²S (Irucka Embry, E.I.T.) – <https://www.ecoccs.com/rtraining.html>

Copyright and License

All R code written by Irucka Embry is distributed under the GPL-3 (or later) license, see the [GNU General Public License {GPL} page](#).

All written content originally created by Irucka Embry is copyrighted under the Creative Commons Attribution-ShareAlike 4.0 International License. All other written content retains the copyright of the original author(s).

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).