Documentation for RSEI Geographic Microdata (RSEI-GM)

February 2023

Contents

Documentation for RSEI Geographic Microdata (RSEI-GM)

The Risk-Screening Environmental Indicators (RSEI) Geographic Microdata (GM) (RSEI-GM)[1] are unique datasets that provide detailed air and water modeling results at various levels of aggregation, spatial geographies, and time periods for data users. RSEI Geographic Microdata allow for a flexible ability to analyze RSEI model outputs and results from a receptor-based perspective of potentially impacted geographic areas. In contrast to the suite of RSEI results that are distributed in tools such as EPA's EasyRSEI dashboard and in EPA's Envirofacts data warehouse,[2] RSEI Microdata are not aggregated and assigned to the facility level, but rather include values resulting from where the chemical releases and potential impacts may occur.

RSEI Microdata are made available in a variety of data and file formats to meet analytical needs. The most detailed files are disaggregated grid cell files, meaning values are presented for each modeled chemical release and each impacted 810 meter by 810 meter grid cell. These disaggregated files are very large in size (up to 120 gigabyte (GB) per year for the entire United States) and can be difficult to download and manipulate from a technical perspective. To make data analysis easier for users, EPA also produces Microdata files at less granular levels, including census block groups, census tracts, and zone improvement plan (ZIP) codes, and produces aggregated files that sum up RSEI results into single values for each geographic unit. In the RSEI Microdata for air, the modeling results available include chemical concentrations, toxicity-weighted concentrations, and RSEI Scores. In the RSEI Microdata for water, the modeling results available include chemical concentrations and toxicity-weighted concentrations for each stream segment or flowline downstream of a water discharge.

More information about downloading the RSEI Microdata can be found on the RSEI website.

Beginning with the 2018 Toxics Release Inventory (TRI) reporting year, RSEI Microdata datasets are typically produced on a two-year alternating schedule. In even TRI reporting years (2018, 2020, etc.), expanded Microdata datasets are produced that include RSEI model results for all TRI reporting years (1988-current) at different geographies (i.e., grid cell, block group, census tract, and ZIP code) for different chemical sets (e.g., all chemicals, core chemicals/original industries, 2001 core chemicals). In odd TRI reporting years (2019, 2021, etc.), Microdata are only produced for the three most recent TRI reporting years. Non-routine or *ad hoc* Microdata datasets may also be produced in odd years and will be made available to the public.

[1] The use of the word "Microdata" throughout this document should be interpreted to be synonymous and interchangeable with the words "RSEI-GM" and "Geographic Microdata".

[2] In order to make the RSEI data small enough to work with in a desktop database application, EasyRSEI adds up potential impacts from certain waste management activities involving TRI chemicals and attributes them to the facilities of origin. These facility-level results are also presented in Envirofacts.

Each time a new RSEI version is released, the most recent three years of grid-cell level Microdata will be available on the RSEI file transfer protocol (FTP) site, accessible through ftp://newftp.epa.gov/RSEI/Current_Version or https://gaftp.epa.gov/rsei/Current_Version. Previous versions will also be maintained on the ftp site.

Microdata files for different geographies, or for earlier years in the case of the even-year expanded Microdata datasets, are available upon request; use the Contact Us form on the RSEI website to send an email with your request.

RSEI Geographic Microdata are summarized differently than the RSEI facility-level data

RSEI facility-level results are distributed in a compact form in tools like EasyRSEI and in applications like EPA's Envirofacts data warehouse. The RSEI facility-level data use the same underlying modeling results as the Microdata, but then sum and assign RSEI result values resulting from each modeled chemical waste management activity (e.g., releases to the environment, transfers off site) back to the original waste-generating facility. For example, when ranking counties or states in EasyRSEI, you are actually ranking the sum of facility-level RSEI results resulting from all waste-originating facilities who are located in each county or state. Being that the chemical release is the ultimate unit of analysis for the RSEI model,[3] potential impacts resulting from chemical waste management activities may occur elsewhere in another county or state. The RSEI facility-level results do not account for where these potential impacts may be occurring geographically, but rather attribute the potential impacts to where the chemical waste originated from.

To highlight this dichotomy, using **Figure 1** below as a hypothetical example, a facility located in the state of Delaware impacts air quality across the borders in the nearby states of Pennsylvania and Maryland. In EasyRSEI and other applications that present RSEI facility-level results, the RSEI pounds-based, hazard-based, and risk-related (score) result values associated with this facility will all be attributed to the state of Delaware in a query that looked at RSEI results by state. In a separate query that now looked at the RSEI Microdata (not the RSEI facility-level results) by state, the RSEI-modeled result values resulting from this facility would be split among the three states (Delaware, Pennsylvania, and Maryland) according to where the potential impacts may ultimately be occurring.

[3] More specifically, it is the chemical release pathway; for instance, water releases can be split into different pathways for fish consumption and drinking water exposure. Chemicals in wastes may also be transferred off site for further waste management (such as for disposal or further treatment) and may ultimately be released to the environment at the off-site location.

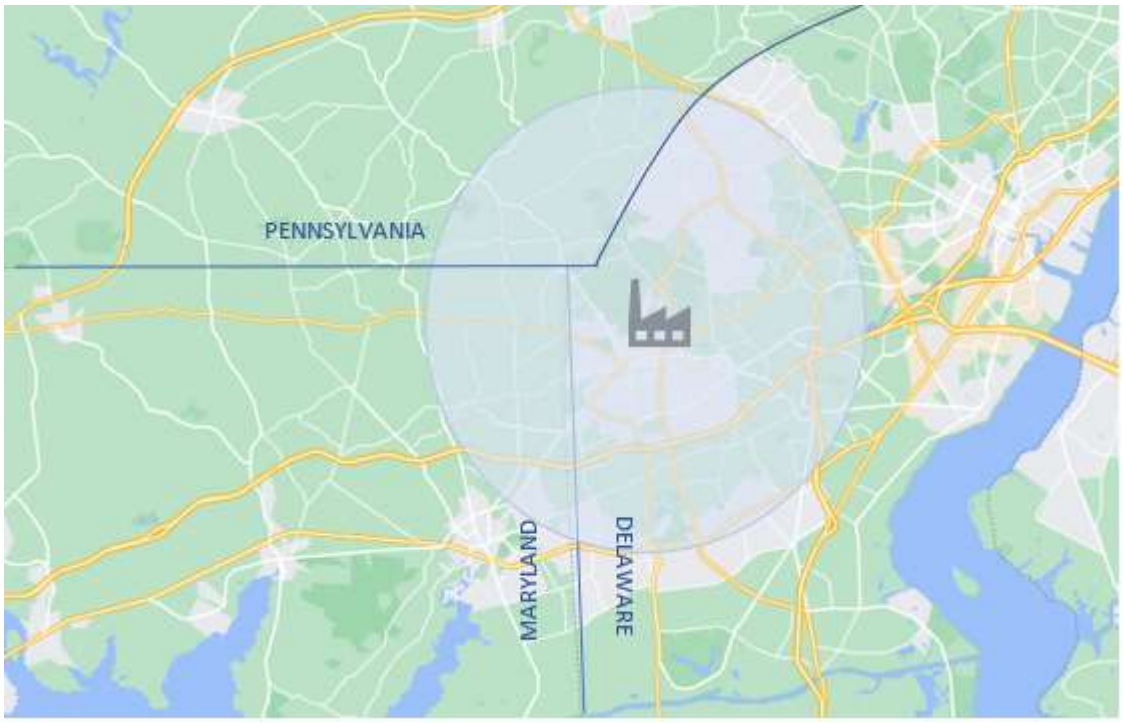

Figure 1. Hypothetical scenario of an emissions plume crossing state boundaries (not to scale)

The RSEI facility-level and the RSEI Microdata datasets are both based on the same underlying modeling, and as a result, total RSEI-modeled result values will be the same at the national-scale level in either dataset. The facility-level dataset is appropriate for prioritizing chemicals, facilities, industries, and pollution prevention opportunities or for screening for situations that may warrant further investigation. Ranking or looking at trends for geographic areas can be useful as long as users understand the nature of the RSEI facility-based data.

With the Microdata, geographic-based analyses are more intuitive; when you rank states, you are ranking the potential impacts that may occur within the geographic confines of each state, regardless of where the chemical releases or generated wastes originate. Impacts resulting from on-site or off-site management of chemical waste are attributed to the geographic areas located around the facilities that ultimately release the chemical into the environment. The Microdata are provided for looking at small-scale geographic areas, and users can examine the potential impacts that environmental releases of toxic chemicals from multiple facilities may have on a particular area from a relative risk-related perspective.

Microdata users should familiarize themselves with RSEI methods and data

As with any screening-level model, RSEI is subject to the limitations, caveats, and assumptions of the underlying data sources and models that it incorporates. You should carefully consider and contextualize the strengths, weaknesses, and meaning that the RSEI methodology and results may have on any analysis or investigation. RSEI relies primarily on TRI-reported data, which only captures information regarding certain toxic chemical waste management activities from facilities subject to TRI reporting requirements. RSEI does not provide information regarding many other sources of environmental pollution or potential risk, including mobile sources, small industrial or commercial facilities, and contaminated hazardous waste (Superfund) sites. The RSEI website contains more description about RSEI, TRI, and other important environmental data sources and useful resources in gathering more information.

Outliers should be checked

The accuracy and validity of TRI data submitted to EPA is entirely the responsibility of the facilities that submit the data. Given the importance of the TRI to the public and other users of the data, EPA is committed to helping facilities submit TRI data of optimal quality, and has implemented activities to support this goal. The data in the TRI database are subject to ongoing data quality review and corrective actions by both EPA and by TRI-reporting facilities. The RSEI model uses the same TRI dataset used for EPA's annual TRI National Analysis, which represents data contained in the TRI database at a given point in time in the autumn season following the July TRI reporting deadline for toxic chemical release reporting forms. Data received by EPA too late to be included in this dataset used for the TRI National Analysis and the RSEI model update are made available to the public the following spring season (after the TRI database has been updated and refreshed) through Envirofacts and other online tools and data sources made available by EPA for accessing and analyzing TRI data. Depending on timing, these online tools and data sources may contain more current TRI data that should be checked for any revisions to anomalously high RSEI result values seen in RSEI data products.

In addition, some high toxicity-weighted concentrations in RSEI are driven by reporting of TRI chemical categories, like certain glycol ethers or various metal compounds such as lead compounds or nickel compounds. Because specific data on the exact identities of these compounds are not reported to TRI, RSEI models a worst-case scenario and assumes that the most toxic form of the chemical or chemical compound is being released for the majority of scenarios. Exceptions to this include polycyclic aromatic compounds (PACs) and chromium and chromium compounds, for which RSEI assumes a more likely release scenario and toxicity profile based on EPA's National Emissions Inventory (NEI) data.[4] In all of these cases, additional investigation needs to be done to determine the exact identities and forms of the chemicals

[4] The RSEI toxicity topic page provides more information. Also see https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei to learn more about NEI data.

being released to the environment before any conclusions about potential impacts or risks can be drawn.

Adjustments must be made for time-series analysis

When using the Microdata over more than one year, users should take care to ensure that the same set of TRI reporting requirements are in force for the whole time period of the analysis. The Chemical table (Chemical.csv, available in the Public Release Data tables or through the RSEI data dictionary) has a series of fields (e.g., Core01ChemicalFlag) to help with picking the appropriate chemical set for a given time-series analysis; the year in the field name is the beginning TRI reporting year for which reporting requirements have not changed over time. For instance, the Core98ChemicalFlag field is "1" (or "True") for any TRI-listed chemical whose reporting requirements have been constant over the TRI reporting years from 1998-present. Users should limit their time series to those chemicals whose requirements have not changed over the period of time being considered; otherwise differences in reporting may skew results (e.g., increases in number of reportable chemicals may erroneously imply increases in potential impacts). Microdata files for "Core" (1988-current) and "Core01" (2001-current) are produced every other year in the extended Microdata set. The following table shows the core fields available in the chemical table, along with the time period to which each flag pertains.

Core Flags Available in the Chemical Table*

Field Name	Time Period
Core88ChemicalFlag	1988-current
Core95ChemicalFlag	1995-current
Core98ChemicalFlag	1998-current
Core00ChemicalFlag	2000-current
Core01ChemicalFlag	2001-current
Core11ChemicalFlag	2011-current
Core12ChemicalFlag	2012-current
*The RSEI core chemical lists are defined by the lists found on the TRI Explorer website here, under "Chemical".	

Users should also be aware that the following seven industry groups were added to list of facilities subject to the TRI reporting requirements for reporting year 1998:

- Metal mining
- Coal mining
- Electric utilities
- Commercial hazardous waste treatment

- Chemicals and allied products-wholesale
- Petroleum bulk terminals and plants-wholesale
- Solvent recovery services

If a time-series analysis includes 1998, those seven industry groups are recommended to be removed prior to analysis. The RSEI Facility table (Facility.csv) also has a True/False field called NewIndustryFlag that is "1" (or "True") for facilities in the seven industry groups added for reporting year 1998. More information on the TRI industry expansion.

The TRI Program's Factors to Consider When Using Toxics Release Inventory Data contains additional important information to consider when accessing and analyzing TRI data over the years.

Air and water Microdata are distributed separately

RSEI water Microdata are produced with different spatial geography than the RSEI air Microdata. Instead of the RSEI grid cell that is used in the air Microdata, the water Microdata use the flowline, or stream segment, as the basic unit of analysis. Before using the water Microdata, users should familiarize themselves with the RSEI water modeling methodology and be aware that TRI data do not include all toxic chemical discharges to water (see the RSEI website for more information on the TRI reporting universe and where to find other sources of environmental release data). The RSEI modeled concentrations for each flowline reflect annual chemical water releases discharged upstream, assumed to be released to surface waters at a constant rate over the year, and do not generally reflect the current state of impairment of the receiving stream or water body.

The RSEI water Microdata is indexed by flowline Common ID (COMID), which is a unique identifier assigned by the NHDPlus data set, Version 2. NHDPlus is based on the U.S. Geological Survey (USGS)'s medium resolution National Hydrography Dataset (NHD), and adds important information like stream flow and velocity estimates that are necessary for water modeling.

Microdata file types

There are different kinds of files that contain Microdata results, and over time additional files will likely be produced to help users get the information they need in the easiest way possible.

Universes. There are three possible universes of chemicals and facilities provided:

- **Core**- limited to the set of TRI-listed toxic chemicals whose reporting requirements have not changed over the entire TRI reporting time period and to the set of original industries subject to TRI reporting requirements.[5] This is the smallest universe of chemicals and facilities.
- **Core01**- limited to the set of TRI-listed toxic chemicals whose reporting requirements have not changed since TRI reporting year 2001. Includes all reporting facilities.
- **All chemicals**- includes all TRI-listed toxic chemicals reported to TRI and all reporting facilities. Any Microdata file with no core specification in the file name contains all TRI-listed chemicals and reporting facilities.

EPA's TRI program maintains a list of regulatory activities with respect to changes in TRI chemical and facility reporting requirements, which can be found through the TRI website here: https://www.epa.gov/toxics-release-inventory-tri-program/tri-laws-and-regulatory-activities Also, in the RSEI data dictionary tables, the Chemical table has a field FirstReportingYear which indicates the first TRI reporting year for a given chemical (not the calendar year in which the given chemical was added to the TRI list), and fields like Core01ChemicalFlag which is "1" or "True" if that chemical is on that particular core chemical list. Using these flags, users can extract any other core chemical subset from the set of all TRI-listed chemicals.

Aggregations. The Microdata can be presented at two different levels:

- **Disaggregated**. This is the most detailed and granular level of data available, where each chemical release and corresponding RSEI result values are presented separately, so that users can see what chemicals and facilities are affecting geographic spatial units (grid cell, census tract, etc.).
- **Aggregated.** In aggregated files, chemical releases and corresponding RSEI result values (e.g., toxicity-weighted concentrations and risk-related (score) results) are summed together for each geographic spatial unit.[6]

Geography. Microdata are created for the following geographic spatial units:

Air Microdata

- **Grid cell**. The 810 meter by 810 meter basic unit of the RSEI grid cell system, defined by an x,y address. Section 5.1 of the RSEI methodology document contains more detailed information on the technical specifications for the RSEI grid, as well as equations for translating latitude and longitude coordinates into a locational (x,y) address. RSEI grid

[5] Facilities in North American Industry Classification System (NAICS) codes that correspond to the seven industry groups added in TRI reporting year 1998 are not included.

[6] Chemical concentrations are not summed in the aggregated Microdata files, therefore RSEI-modeled chemical concentrations are not available at this level of aggregation.

shapefiles that define the grid are available for download on the RSEI FTP site for use in mapping.

- **Block group**. The 2010 U.S. census boundaries are used.
- **Census tract**. The 2010 U.S. census boundaries are used.
- **ZIP code.** 2017 ZIP code boundaries are used.

Water Microdata

- **Flowline.** Stream segments as defined in the NHDPlus Version 2 dataset. Each flowline is indexed by a unique identifier called "COMID".

Coverage.

- **Grids.** The aggregated grid cell Microdata is split into separate grids. One file contains only the conterminous U.S.; other files contain Alaska, Hawaii and island areas. Only grid cells with non-zero concentrations are included; if a grid cell is not affected by any releases, it will not be in the file. The grid codes are as follows:
  - 14 = Conterminous US
  - 24 = Alaska
  - 34 = Hawaii
  - 44 = Puerto Rico/Virgin Islands
  - 54 = Guam/Marianas Islands
  - 64 = American Samoa
- **National.** Most files are for the entire U.S., including island areas. As with the aggregated grid cell files, if a geographic unit is not affected by any releases, it will not be included in the file.
- **State.** State-level shapefiles are provided that include all RSEI grid cells that define the state, regardless of RSEI modeled concentrations. Users can use the state-level shapefile as a template set of x,y addresses to then extract state-specific Microdata results from other grid cell level files (e.g., for other TRI reporting years).

File Type. There are two basic kinds of file types produced:

- **Comma-delimited text file.** See field descriptions in Appendix 1 at the end of this document for more specifics.
- **Shapefiles.** Each zip file contains a comma-delimited text file as well as the set of files that define the shapefile. Note that some of these files are very large and may not be viewable in some entry-level geographic information system (GIS) software programs.

The files are generally zipped in gzip (file extension ".gz"), or Winzip (file extension ".zip") formats, both of which can be unzipped using Windows Explorer or a command line. The following sections describe each type of Microdata file and how they are named. Field names and descriptions for each type of file can be found in Appendix 1.

File Descriptions and Naming Conventions

Grid Cell Microdata files

Disaggregated Microdata files
Approximate Size: 110 GB, 25 GB when zipped
Description: These are the Microdata files that contain the most disaggregated data possible. For each 810m by 810m grid cell, the file includes RSEI result values such as risk-related results (scores), concentrations, and toxicity-weighted concentrations for each modeled chemical release. There may be multiple records for any one particular grid cell. As an example, if two releases of the same chemical impact the same grid cell, there will be separate records for each release.[7]

Naming: These annual files are named as microXXXX_YYYY, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, micro2014_2010 is from the TRI RY 2014 RSEI update, and contains data for chemicals released in 2010.

See the Disaggregated Grid Cell Microdata table in Appendix 1 for field names and descriptions.

Aggregated Microdata files
Approximate Size: 700 MB, 200 MB when zipped (for GC 14, the conterminous U.S.)
Description: Aggregated Microdata files use the same data as the disaggregated files, but sum the RSEI result values over all the chemical releases affecting each grid cell. Because the values are summed, unweighted concentrations are not available.

Naming: These annual files are named aggmicroXXXX_YYYY_gcZZ, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, YYYY is the TRI reporting year data contained in the file, and ZZ is the grid code. For example, aggmicro2017_2010_gc14 is from the TRI RY 2017 RSEI update, contains data for chemicals released in 2010, for grid code 14 (the conterminous United States).

See the Aggregated Grid Cell Microdata table in Appendix 1 for field names and descriptions.

State Shapefiles
Approximate Size: less than 100 MB when all files zipped
Description: These zip files contain shapefiles with aggregated grid cell-level data for individual states for the most recent data year only. The data are presented in a shapefile format for use with GIS programs. Due to the size of the national-scale grid cell Microdata files, a set of state-specific files are made available. For this set only, RSEI grid cells with no impacts are included; in this way, each state file includes every grid cell within the state, and it can be merged with the

[7] Note that this may either occur from releases from different facilities or from multiple types of releases (e.g., stack air emissions, fugitive air emissions) from the same facility.

national-scale files to extract data for one state from other grid cell-level files. For state shapefiles, a comma-delimited text file is not included.

Naming: These files are named ss_aggmicroYYYY.zip, where ss is the 2-letter state abbreviation and YYYY is the TRI reporting year (RY) data freeze used for the RSEI model and also the TRI RY data contained in the file. The state-level shapefiles are only produced for the most recent TRI reporting year for each expanded Microdata datasets.

See the State Aggregated Grid Cell Microdata Shapefiles table in Appendix 1 for field names and descriptions.

Block Group Microdata files

Disaggregated Block Group Microdata
Approximate Size: 15 GB, 3 GB when zipped
Description: These files are the same as the grid cell disaggregated Microdata files, but instead of being presented at the grid cell level, the values are averaged and presented for census block groups. For each census-defined block group, the file includes RSEI result values such as risk-related results (scores), concentrations, and toxicity-weighted concentrations for each modeled chemical release. There may be multiple records for any one block group; for instance, if two releases for the same chemical (either from different facilities or one from a stack release and one from a fugitive release from the same facility) affect the same block group, there will be separate records for each release.

Naming: These annual files are named as censusmicroblockgroupXXXX_YYYY, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, censusmicroblockgroup2017_2010 is from the TRI RY 2017 RSEI update, and contains data for chemicals released in 2010.

See the Disaggregated Block Group Microdata table in Appendix 1 for field names and descriptions.

Block Group Shapefiles
Approximate Size: less than 1 GB when all files zipped
Description: These zip files contain the same data as the disaggregated block group files, but sum the chemical releases and corresponding RSEI result values over each block group, so that there is only one record in each table for each affected block group. The data are presented in a shapefile format for use with GIS programs. For user convenience, a standard database file (dbf) and a comma-delimited text file (provided as a comma-separated values (csv) file) are included in each zip file; that text file is the same as the text file for the aggregated Microdata. These files do not contain chemical- or facility-specific information, except the number of facilities and chemicals affecting each block group. These files also contain percentile fields to help in mapping.

Naming: These annual files are named as censusmicroblockgroupXXXX_YYYY_aggregated, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, censusmicroblockgroup2017_2010_aggregated is from the TRI RY 2017 RSEI update, and contains data for chemicals released in 2010.

See the Block Group Shapefiles (dbf file) table in Appendix 1 for field names and descriptions.

Aggregated Block Group Microdata
Approximate Size: 20 MB, 7 MB when zipped
See block group shapefiles description above; the aggregated block group data are available in the zip shapefile or as a separate comma-delimited csv text file. The fields differ slightly between the dbf and csv files.

See the Block Group Aggregated Data (csv file) table in Appendix 1 for field names and descriptions.

Census Tract Microdata files

Disaggregated Census Tract Microdata
Approximate Size: 5 GB, 1 GB when zipped
Description: These files are the same as the grid cell disaggregated Microdata files, but instead of being presented at the grid cell level, the values are averaged and presented over census tracts. For each census-defined tract, the file includes RSEI result values such as risk-related results (scores), concentrations, and toxicity-weighted concentrations for each modeled chemical release. There may be multiple records for any one census tract; for instance, if two releases for the same chemical (either from different facilities or one from a stack release and one from a fugitive release from the same facility) affect the same census tract, there will be separate records for each release.

Naming: These annual files are named as censusmicrotractsXXXX_YYYY, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, censusmicrotracts2017_2010 is from the TRI RY 2017 RSEI update, and contains data for chemicals released in 2010.

See the Disaggregated Census Tract Microdata table in Appendix 1 for field names and descriptions.

Census Tract Shapefiles
Approximate Size: less than 500MB when all files zipped
Description: These zip files contain the same data as the disaggregated census tract files, but sum the chemical releases and corresponding RSEI result values over each census tract, so that there is only one record in each file for each affected census tract. The data are presented in a shapefile format for use with GIS programs. For user convenience, a standard database file

(dbf) and a comma-delimited text file (provided as a comma-separated values (csv) file) are included in each zip file; that text file is the same as the text file for the aggregated Microdata. These files do not contain chemical- or facility-specific information, except the number of facilities and chemicals affecting each census tract. These files also contain percentile fields to help in mapping.

Naming: These annual files are named as censusmicrotractsXXXX_YYYY_aggregated, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, censusmicrotracts2017_2010_aggregated is from the TRI RY 2017 RSEI update, and contains data for chemicals released in 2010.

See the Census Tract Shapefiles (dbf file) table in Appendix 1 for field names and descriptions.

Aggregated Census Tract Microdata
Approximate Size: 6 MB, 3 MB when zipped
See census tract shapefiles description above; the aggregated census tract data are available in the zip shapefile or as a separate comma-delimited csv text file. The fields differ slightly between the dbf and csv files.

See the Census Tract Aggregated Data (csv file) table in Appendix 1 for field names and descriptions.

ZIP Code Microdata files

Disaggregated ZIP Code Microdata
Approximate Size: 1.2 GB, 200 MB when zipped
Description: These files are the same as the grid cell disaggregated Microdata files, but instead of being presented at the grid cell level, the values are averaged and presented over ZIP codes. For each ZIP code, the file includes RSEI result values such as risk-related results (scores), concentrations, and toxicity-weighted concentrations for each modeled chemical release. There may be multiple records for any one ZIP code; for instance, if two releases for the same chemical (either from different facilities or one from a stack release and one from a fugitive release from the same facility) affect the same ZIP code, there will be separate records for each release.

Naming: These annual files are named as censusmicrozipcodeXXXX_YYYY, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, censusmicrozipcode2017_2010 is from the TRI RY 2017 RSEI update, and contains data for chemicals released in 2010.

See the Disaggregated ZIP Code Microdata table in Appendix 1 for field names and descriptions.

ZIP Code Shapefiles
Approximate Size: less than 500MB when all files zipped
Description: These zip files contain the same data as the disaggregated ZIP code files, but sum the chemical releases and corresponding RSEI result values over each ZIP code, so that there is only one set of records for each affected ZIP code. The data are presented in a shapefile format for use with GIS programs. For user convenience, a standard database file (dbf) and a comma-delimited text file (provided as a comma-separated values (csv) file) are included in each zip file; that text file is the same as the text file for the aggregated Microdata. These files do not contain chemical- or facility-specific information, except the number of facilities and chemicals affecting each ZIP code. These files also contain percentile fields to help in mapping.

Naming: These annual files are named as censusmicrozipcodeXXXX_YYYY_aggregated, where XXXX is the TRI reporting year (RY) data freeze used for the RSEI model, and YYYY is the TRI reporting year data contained in the file. For example, censusmicrozipcode2017_2010_aggregated is from the TRI RY 2017 RSEI update, and contains data for chemicals released in 2010.

See the ZIP Code Shapefiles (dbf file) table in Appendix 1 for field names and descriptions.

Aggregated ZIP Code Microdata
Approximate Size: 2 MB, 1 MB when zipped
See ZIP code shapefiles description above, the aggregated ZIP code data are available in the zip shapefile or as a separate comma-delimited csv text file. The fields differ slightly between the dbf and csv files.

See the ZIP Code Aggregated Data (csv file) table in Appendix 1 for field names and descriptions.

Water Microdata files

Water Microdata files come in two basic forms: yearly shapefiles containing toxicity-weighted concentrations for each affected flowline, and all-years discharge-based files which contain modeled concentrations for individual discharges at each downstream flowline. In both cases, on-site discharges from TRI reporters are distributed separately from discharges from publicly-owned treatment works (POTWs) resulting from off-site transfers from TRI reporters.

Annual Aggregated Shapefiles for Water Microdata
Approximate Size: 40 MB, 25 MB when zipped
Description: For these files, the individual toxicity-weighted concentrations from each discharge are summed for each flowline. The data are presented in a shapefile format for use with GIS programs. For user convenience, a standard database file (dbf) and a comma-delimited text file (provided as a comma-separated values (csv) file) with toxicity-weighted concentrations are included in each zip file; note that the text file has additional fields with coordinates for three points on each flowline. These files do not contain chemical- or facility-

specific information. The data are indexed by the flowline identifier (COMID), and the reach code associated with each flowline is also provided.

Naming: These annual files are named as NHDMicroResults_Zsite_YYYY, where Z is "On" or "Off" depending on whether the file contains results for on-site discharges or from discharges resulting from off-site transfers to POTWs, and YYYY is the TRI reporting year data contained in the file. For example, NHDMicroResults_Onsite_2015 is for on-site discharges, and contains data for chemicals released in 2015.

See the Water Microdata: Annual Aggregated Shapefiles table in Appendix 1 for field names and descriptions.

All Years Discharges for Water Microdata
Approximate Size: 5 GB, 1 GB when zipped
Description: These text files contain data for individual discharges, including concentration, toxicity-weighted concentrations, downstream travel time, and year of release. Each record is for a single discharge (defined by the ReleaseNumber) and single downstream flowline (defined by the COMID). The reach code is also provided. Information on the discharging facility and the chemicals discharged can be found by linking the ReleaseNumber contained in this file to the Public Release Data files. See the section on linking RSEI tables for details.

Naming: These files are named as NHDMicroResults_conc_aggZsite, where Z is "On" or "Off" depending on whether the file contains results for on-site discharges or from discharges resulting from off-site transfers to POTWs. For example, NHDMicroResults_conc_aggOffsite is for off-site discharges, and contains data for chemicals released in all years (1988-current).

See the Water Microdata: All Years Discharges table in Appendix 1 for field names and descriptions.

Census crosswalk files

Each set of census crosswalk files links the RSEI grid cell geography to a different U.S. decennial census year. There is one crosswalk for each area and decennial census year (i.e., 1990, 2000, 2010). Crosswalk files are named by area (Alaska, Con(terminous) U.S., etc.). The last three fields in each crosswalk file contain percent values that can be used to adjust the census block or grid cell contents when performing the crosswalk. The percent of the census block that is within the grid cell (PCT_B_C) and the percent of the grid cell that is within the census block (PCT_C_B) are area-weighted and can be used for metrics that do not involve population, such as concentration and toxicity-weighted concentration. The percent of the grid cell's population that is within the census block (PCT_PC_B) is population-weighted, and can be used to crosswalk fields that involve population, such as risk-related (score) results. Note that the "PCT_PC_B" field is not available for the territories (VI, PR, GU, AS, MP). The Northern Mariana Islands are in the Guam file and the Virgin Islands are in the Puerto Rico file. There are no crosswalks for Puerto Rico, the Virgin Islands, Mariana Islands, Guam, or American Samoa for

1990. For these areas, the RSEI model uses 2000 census block boundaries and scales each grid cell's population by the overall ratio of 1990/2000 population for each area.

If census geographies larger than block level are desired, users are advised to use the RSEI Microdata geography files which are already crosswalked from grid cells to census block groups, census tracts, and ZIP codes.

See the Census Crosswalk table in Appendix 1 for field names and descriptions.

Public Release Data files (data tables)

These data tables contain TRI reporting data and other data used in RSEI modeling, including input parameters such as facility information and chemical toxicity. **Figure 2** below shows the relationships between the main Public Release Data tables. The RSEI Data Dictionary provides field descriptions and data tables that are made available in comma-separated values (csv) and spreadsheet (xlsx) formats for the most recent RSEI model version. The Public Release Data files for current and older RSEI model versions can be found through the RSEI FTP site.

Grid Shapefiles

RSEI grid shapefiles define the RSEI grid and can be used for mapping purposes. These files do not contain any RSEI modeled results. New grid shapefiles were posted on the RSEI FTP site in early 2017. The shapes are the same as previous to 2017; however, the fields and format are different, and now additional files for grid cell sizes other than 810 meter x 810 meter are available. Shapefiles of the RSEI grid geography are available by grid region, in two versions: polygon and center point. The RSEI Data Dictionary provides additional descriptions and attribute data for the grid shapefiles.

Linking RSEI tables

The Microdata tables must be used with the RSEI Public Release Data files (data tables) of the same RSEI model version.

Disaggregated Microdata data files provide information on each modeled chemical release affecting each spatial geography.[8] To save space, the facility, chemical, media, and release details are stored in the RSEI Public Release Data tables. Each Public Release Data table links into the Microdata tables using a key field (such as ChemicalNumber, FacilityNumber, etc.), as illustrated in **Figure 2**, below. Field descriptions for these tables can be found in the RSEI Data Dictionary. As an example, if a record in the disaggregated grid cell Microdata had a very high toxicity-weighted concentration, you could find its ChemicalNumber (Field Number 5 in the Disaggregated Grid Cell Microdata table in Appendix 1) and look it up in the Chemical table and

[8] Aggregated Microdata data files combine release information and RSEI modeled results for each spatial geography, and thus cannot be linked back to the RSEI data tables.

find its chemical name, toxicity information, and when it was added to the TRI list of toxic chemicals. Similarly, you could look up the FacilityNumber in the Facility table, and find the facility name, location, and facility modeling parameters.

The air Microdata contains both on-site and off-site releases in the same file. If you are interested in facility information for a given record in a disaggregated Microdata file, the FacilityNumber links to the RSEI Facility table if the release is an on-site release from a TRI-reporting facility (the value in the Media field will be 1 or 2); if the release is from an off-site facility, the value in the Media field will be 6, 750, or 754 and the FacilityNumber in the Microdata file links to the Off-site table. If you want information on the originating facility for an off-site transfer (the TRI reporting facility that transferred the waste to, for instance, an off-site incineration facility), use the ReleaseNumber in the Microdata file to link to the Releases table. The FacilityNumber for the record in the Releases table links to the Facility table.[9]

It is very important to use the Public Release Data tables for the same RSEI model version as the Microdata tables. For instance, you must use the RSEI model version V2.3.10 (TRI reporting year 2020) Public Release Data tables with the RSEI model version V2.3.10 (TRI reporting year 2020) Microdata. The key field values change with each RSEI model version and update, so if you use a different RSEI model version's tables to do analysis with the Microdata, you will get erroneous results.

[9] Each record in the Releases table includes the reporting facility (FacilityNumber) and the off-site facility, if applicable (OffsiteNumber). The OffsiteNumber field in the Releases table is the same as the FacilityNumber field in the Offsite table.

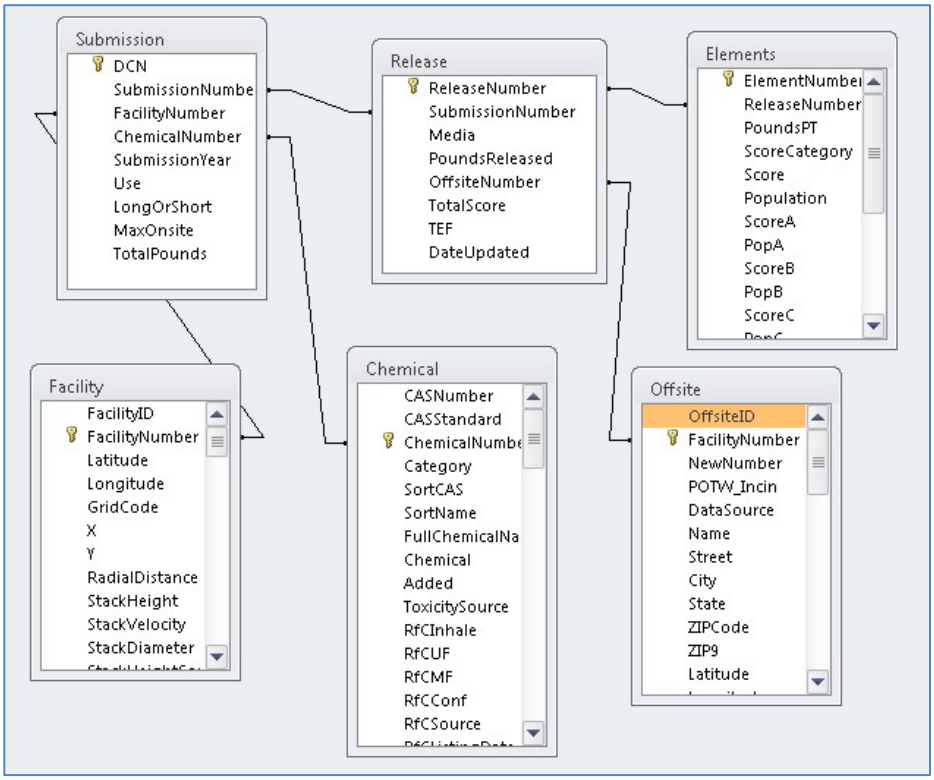

Figure 2. Example relationships between RSEI Public Release Data tables

Recent RSEI model version numbers and TRI reporting year (RY) data freeze used for that RSEI model update:

- V2.3.1- RY 2010
- V2.3.2- RY 2011
- V2.3.3- RY 2012
- V2.3.4- RY 2014
- V2.3.5- RY 2015
- V2.3.6- RY 2016
- V2.3.7- RY 2017
- V2.3.8- RY 2018
- V2.3.9- RY 2019
- V2.3.10- RY 2020
- V2.3.11-RY 2021

Appendix 1. Field Descriptions for Microdata Tables

These Microdata tables, along with the tables for the RSEI facility-level data, can also be found in the RSEI Data Dictionary.

Disaggregated Grid Cell Microdata

Field Number	Name	Description
1	GridCode	Identifies grid. 14=Conterminous US 24=Alaska 34=Hawaii 44=Puerto Rico/Virgin Islands 54=Guam/Marianas 64=American Samoa
2	X	X-coordinate of grid cell.
3	Y	Y-coordinate of grid cell.
4	ReleaseNumber	Internal unique identifier for release (lookup in Releases table)[1].
5	ChemicalNumber	Internal unique identifier of released chemical (lookup in Chemical table)[1].
6	FacilityNumber	Internal unique identifier of releasing facility (If media = 1 or 2, then lookup in Facility table; If media = 6 or 750 or 754, then lookup in Off-site table)[1].
7	Media	Code describing media into which chemical is released. (lookup in Media table)[1].
8	Conc	Concentration of chemical for release/media at grid cell (µg/m^3).
9	ToxConc	Concentration multiplied by inhalation toxicity weight.
10	Score	Risk-related score (estimated dose * toxicity weight * potentially exposed population).
11	ScoreCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for cancer effects.
12	ScoreNonCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for noncancer effects.
13	Pop	Number of people in grid cell (may be interpolated).

[1] Other tables can be found in the Public Release Data tables, on the RSEI FTP site (ftp://newftp.epa.gov/RSEI/) or https://gaftp.epa.gov/RSEI, or in the RSEI data dictionary (https://www.epa.gov/rsei/rsei-data-dictionary). Be sure to use the data tables from the same RSEI model version as the Microdata.

Aggregated Grid Cell Microdata

Field Number	Name	Description
1	X	X-coordinate of grid cell.
2	Y	Y-coordinate of grid cell.
3	NumberOfFacilities	Number of facilities with releases affecting grid cell.
4	NumberOfReleases	Number of individual releases affecting grid cell.
5	NumberOfChemicals	Number of chemicals with nonzero concentrations for grid cell.
6	ToxConc	Concentration multiplied by inhalation toxicity weight, summed over all chemicals impacting grid cell.
7	Score	Risk-related score (estimated dose * toxicity weight * potentially exposed population), summed over all chemicals impacting grid cell.
8	Pop	Population of grid cell.
9	ScoreCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for cancer effects.
10	ScoreNonCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for noncancer effects.

State Aggregated Grid Cell Microdata Shapefiles

Field Number	Name	Description
1	CELLX	X-coordinate of grid cell.
2	CELLY	Y-coordinate of grid cell.
3	CLAT	Latitude in decimal degrees of center of grid cell.
4	CLONG	Longitude in decimal degrees of center of grid cell.
5	PRIFIPS	5-digit FIPS code associated with the county in which the largest percentage of the grid cell is found.
6	NUMFACS	Number of facilities with releases affecting grid cell.
7	NUMRELEASES	Number of individual releases affecting grid cell.
8	NUMCHEMS	Number of chemicals with nonzero concentrations for grid cell.
9	TOXCONC	Concentration multiplied by inhalation toxicity weight, summed over all chemicals impacting grid cell.

Field Number	Name	Description
10	SCORE	Risk-related score (estimated dose * toxicity weight * potentially exposed population), summed over all chemicals impacting grid cell.
11	EXPOP	Population of grid cell, for cells with nonzero concentrations only.
12	CSCORE	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for cancer effects.
13	NCSCORE	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for noncancer effects.

Disaggregated Block Group Microdata

Field Number	Name	Description
1	GEOID	US Census Block Group ID
2	ReleaseNumber	Internal unique identifier for release (lookup in Releases table)[1].
3	ChemicalNumber	Internal unique identifier of released chemical (lookup in Chemical table)[1].
4	FacilityNumber	Internal unique identifier of releasing facility (If media = 1 or 2, then lookup in Facility table; If media = 6 or 750 or 754, then lookup in Off-site table)[1].
5	Media	Code describing media into which chemical is released. (lookup in Media table)[1].
6	Conc	Concentration of chemical for release/media at block group ($\mu g/m^3$).
7	ToxConc	Concentration multiplied by inhalation toxicity weight.
8	Score	Risk-related score (estimated dose * toxicity weight * potentially exposed population).
9	ScoreCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for cancer effects.
10	ScoreNonCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for noncancer effects.
11	Pop	Number of people in block group (may be interpolated).

[1] Other tables can be found in the Public Release Data tables, on the RSEI FTP site (ftp://newftp.epa.gov/RSEI/) or https://gaftp.epa.gov/RSEI, or in the RSEI data dictionary (https://www.epa.gov/rsei/rsei-data-dictionary). Be sure to use the data tables from the same RSEI model version as the Microdata.

Block Group Shapefiles (dbf file)

Field Number	Name	Description
1	GEOID	US Census Block Group ID.
2	NUMFACS	Number of facilities affecting the block group.
3	P_NUMFACS	Percentile associated with field NUMFACS.
4	NUMRELEASES	Number of releases affecting the block group.
5	P_NUMRELEASES	Percentile associated with field NUMRELEASES.
6	NUMCHEMS	Number of chemicals affecting the block group.
7	P_NUMCHEMS	Percentile associated with field NUMCHEM.
8	TOXCONC	Average toxicity-weighted concentration of the cells in the block group.
9	P_TOXCONC	Percentile associated with field TOXCONC.
10	SCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the block group.
11	P_SCORE	Percentile associated with field SCORE.
16	POP	Sum of the population of the cells in the block group.
17	P_POP	Percentile associated with field POP.
14	CSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the block group. Score is calculated using only cancer toxicity weights.
15	P_CSCORE	Percentile associated with field CSCORE.
12	NCSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the block group. Score is calculated using only noncancer toxicity weights.
13	P_NCSCORE	Percentile associated with field NCSCORE.

Block Group Aggregated Data (csv file)

Field Number	Name	Description
1	GEOID10	US Census Block Group ID.
2	NUMFACS	Number of facilities affecting the block group.
3	NUMRELEASES	Number of releases affecting the block group.
4	NUMCHEMS	Number of chemicals affecting the block group.
5	TOXCONC	Average toxicity-weighted concentration of the cells in the block group.
6	SCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the block group.
7	POP	Sum of the population of the cells in the block group.

Field Number	Name	Description
8	CSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the block group. Score is calculated using only cancer toxicity weights.
9	NCSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the block group. Score is calculated using only noncancer toxicity weights.

Disaggregated Census Tract Microdata

Field Number	Name	Description
1	GEOID	US Census Tract ID.
2	ReleaseNumber	Internal unique identifier for release (lookup in Releases table) [1].
3	ChemicalNumber	Internal unique identifier of released chemical (lookup in Chemical table) [1].
4	FacilityNumber	Internal unique identifier of releasing facility (If media = 1 or 2, then lookup in Facility table; If media = 6 or 750 or 754, then lookup in Off-site table) [1].
5	Media	Code describing media into which chemical is released. (lookup in Media table) [1].
6	Conc	Concentration of chemical for release/media at census tract ($\mu g/m^3$).
7	ToxConc	Concentration multiplied by inhalation toxicity weight.
8	Score	Risk-related score (estimated dose * toxicity weight * potentially exposed population).
9	ScoreCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for cancer effects.
10	ScoreNonCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for noncancer effects.
11	Pop	Number of people in tract (may be interpolated).

[1] Other tables can be found in the Public Release Data tables, on the RSEI FTP site (ftp://newftp.epa.gov/RSEI/) or https://gaftp.epa.gov/RSEI, or in the RSEI data dictionary (https://www.epa.gov/rsei/rsei-data-dictionary). Be sure to use the data tables from the same RSEI model version as the Microdata.

Census Tract Shapefiles (dbf file)

Field Number	Name	Description
1	GEOID	US Census Tract ID.
2	NUMFACS	Number of facilities affecting the tract.
3	P_NUMFACS	Percentile associated with field NUMFACS.
4	NUMRELEASES	Number of releases affecting the tract.
5	P_NUMRELEASES	Percentile associated with field NUMRELEASES.
6	NUMCHEMS	Number of chemicals affecting the tract.
7	P_NUMCHEMS	Percentile associated with field NUMCHEM.
8	TOXCONC	Average toxicity-weighted concentration of the cells in the tract.
9	P_TOXCONC	Percentile associated with field TOXCONC.
10	SCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the tract.
11	P_SCORE	Percentile associated with field SCORE.
16	POP	Sum of the population of the cells in the tract.
17	P_POP	Percentile associated with field POP.
14	CSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the tract. Score is calculated using only cancer toxicity weights.
15	P_CSCORE	Percentile associated with field CSCORE.
12	NCSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the tract. Score is calculated using only noncancer toxicity weights.
13	P_NCSCORE	Percentile associated with field NCSCORE.

Census Tract Aggregated Data (csv file)

Field Number	Name	Description
1	GEOID10	US Census Tract ID.
2	NUMFACS	Number of facilities affecting the tract.
3	NUMRELEASES	Number of releases affecting the tract.
4	NUMCHEMS	Number of chemicals affecting the tract.
5	TOXCONC	Average toxicity-weighted concentration of the cells in the tract.
6	SCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the tract.
7	POP	Sum of the population of the cells in the tract.

Field Number	Name	Description
8	CSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the tract. Score is calculated using only cancer toxicity weights.
9	NCSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the tract. Score is calculated using only noncancer toxicity weights.

Disaggregated ZIP Code Microdata

Field Number	Name	Description
1	GEOID	US ZIP code
2	ReleaseNumber	Internal unique identifier for release (lookup in Releases table) [1].
3	ChemicalNumber	Internal unique identifier of released chemical (lookup in Chemical table) [1].
4	FacilityNumber	Internal unique identifier of releasing facility (If media = 1 or 2, then lookup in Facility table; If media = 6 or 750 or 754, then lookup in Off-site table) [1].
5	Media	Code describing media into which chemical is released. (lookup in Media table) [1].
6	Conc	Concentration of chemical for release/media at ZIP code ($\mu g/m^3$).
7	ToxConc	Concentration multiplied by inhalation toxicity weight.
8	Score	Risk-related score (estimated dose * toxicity weight * potentially exposed population).
9	ScoreCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for cancer effects.
10	ScoreNonCancer	Risk-related score (estimated dose * toxicity weight * potentially exposed population) using only toxicity values for noncancer effects.
11	Pop	Number of people in ZIP code (may be interpolated).

[1] Other tables can be found in the Public Release Data tables, on the RSEI FTP site (ftp://newftp.epa.gov/RSEI/) or https://gaftp.epa.gov/RSEI, or in the RSEI data dictionary (https://www.epa.gov/rsei/rsei-data-dictionary). Be sure to use the data tables from the same RSEI model version as the Microdata.

ZIP Code Shapefiles (dbf file)

Field Number	Name	Description
1	GEOID	US ZIP code.
2	NUMFACS	Number of facilities affecting the ZIP code.
3	P_NUMFACS	Percentile associated with field NUMFACS.
4	NUMRELEASES	Number of releases affecting the ZIP code.
5	P_NUMRELEASES	Percentile associated with field NUMRELEASES.
6	NUMCHEMS	Number of chemicals affecting the ZIP code.
7	P_NUMCHEMS	Percentile associated with field NUMCHEM.
8	TOXCONC	Average toxicity-weighted concentration of the cells in the ZIP code.
9	P_TOXCONC	Percentile associated with field TOXCONC.
10	SCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the ZIP code.
11	P_SCORE	Percentile associated with field SCORE.
16	POP	Sum of the population of the cells in the ZIP code.
17	P_POP	Percentile associated with field POP.
12	NCSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the ZIP code. Score is calculated using only noncancer toxicity weights.
13	P_NCSCORE	Percentile associated with field NCSCORE.
14	CSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the ZIP code. Score is calculated using only cancer toxicity weights.
15	P_CSCORE	Percentile associated with field CSCORE.

ZIP Code Aggregated Data (csv file)

Field Number	Name	Description
1	GEOID10	US ZIP code.
2	NUMFACS	Number of facilities affecting the ZIP code.
3	NUMRELEASES	Number of releases affecting the ZIP code.
4	NUMCHEMS	Number of chemicals affecting the ZIP code.
5	TOXCONC	Average toxicity-weighted concentration of the cells in the ZIP code.
6	SCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the ZIP code.
7	POP	Sum of the population of the cells in the ZIP code.

Field Number	Name	Description
8	CSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the ZIP code. Score is calculated using only cancer toxicity weights.
9	NCSCORE	Sum of the risk-related scores (estimated dose * toxicity weight * potentially exposed population) of the cells in the ZIP code. Score is calculated using only noncancer toxicity weights.

Water Microdata: Annual Aggregated Shapefiles

Field Number	Name	Description
1	ComID	"Common Identifier" of a flowline (sub-segment of a reach)- atomic unit of reach data that matches one-to-one to NHDPlus.
2	ReachCode	Code for reach
3	ToxConc	Sum of toxicity-weighted concentrations in the flowline. Concentration is in mg/L.
4	Lat1	Latitude of the 25th point of the 100 that constitute a flowline in NHDPlus.*
5	Long1	Longitude of the 25th point of the 100 that constitute a flowline in NHDPlus.*
6	Lat2	Latitude of the 50th point of the 100 that constitute a flowline in NHDPlus.*
7	Long2	Longitude of the 50th point of the 100 that constitute a flowline in NHDPlus.*
8	Lat3	Latitude of the 75th point of the 100 that constitute a flowline in NHDPlus.*
9	Long3	Longitude of the 75th point of the 100 that constitute a flowline in NHDPlus.*

* Note that points may not be evenly spaced.

Water Microdata: All Years Discharges

Field Number	Name	Description
1	ReleaseNumber	Internal unique identifier for release (links to Releases table).[1]
2	Counter	Auto-increment count of COMIDs
3	ComID	"Common Identifier" of a flowline (sub-segment of a reach)- atomic unit of reach data that matches one-to-one to NHDPlus.
4	ReachCode	Code for reach

Field Number	Name	Description
5	Conc	Concentration of chemical in flowline (mg/L)
6	Sequence	Number defining pathway of release (used to indicate branching).
7	TravelTime	Time(s) for release to go from top of flowline to bottom.
8	TravelLength	Distance (m) for release to go from top of flowline to bottom
9	Paths	Number of branches in stream path
10	FCode	Descriptor from NHDPlus for type of flowline (e.g., pipeline, stream)
11	ResCode	Internal code
12	OTW	Oral toxicity weight for the chemical.
13	ToxConc	Toxicity-weighted concentration in the flowline for the discharge. Concentration is in mg/L.
14	NCTW	Noncancer toxicity weight for the chemical.
15	CTW	Cancer toxicity weight for the chemical.
16	Year	Year the discharge occurred.

[1] Other tables can be found in the public data set, on the RSEI ftp site (ftp://newftp.epa.gov/RSEI/) or in the RSEI data dictionary (https://www.epa.gov/rsei/rsei-data-dictionary). Be sure to use the data tables from the same RSEI version as the Microdata.

Census Crosswalk Table

Field Number	Name	Description
1	GridID	Identifies grid. 14=Conterminous US 24=Alaska 34=Hawaii 44=Puerto Rico/Virgin Islands 54=Guam/Marianas 64=American Samoa
2	X	X coordinate of the cell address.
3	Y	Y coordinate of the cell address.
4	Block_ID00	US Census Block ID.
5	UR	Internal.
6	PCT_B_C	Percent of the Census block that is within the cell (Block to Cell).
7	PCT_C_B	Percent of the cell that is within the Census block (Cell to Block).
8	PCT_CP_B	Percent of the cell's population that is within the Census block (Population-Cell to Block).

[updated 2/2/2023]