

# Documentation for RSEI Geographic Microdata (RSEI-GM)

---

*February 2021*

## Contents

---

Documentation for RSEI Geographic Microdata (RSEI-GM) .....	1
Documentation for RSEI Geographic Microdata (RSEI-GM) .....	4
RSEI Microdata are summarized differently than the RSEI facility-level data .....	5
RSEI facility-level results .....	5
RSEI Microdata results .....	5
Microdata users should familiarize themselves with RSEI methods and data .....	5
Outliers should be checked .....	6
Adjustments must be made for time-series analysis .....	6
Water Microdata .....	7
Microdata File Types .....	8
Grid Cell Microdata files .....	9
Block Group Microdata files .....	10
Census Tract Microdata files .....	10
ZIP Code Microdata files .....	11
Water Microdata files .....	12
Census Crosswalk Files .....	12
Public Data Release Files (data tables) .....	13
Grid Shape files .....	13
Linking Tables .....	13
Appendix 1. Field Descriptions for Microdata Tables .....	15
Disaggregated Grid Cell Microdata .....	15
Aggregated Grid Cell Microdata .....	16
State Aggregated Grid Cell Microdata Shapefiles .....	16
Disaggregated Block Group Microdata .....	17
Block Group Shapefiles (dbf file) .....	18
Block Group Aggregated Data (csv file) .....	18
Disaggregated Census Tract Microdata .....	19
Census Tract Shapefiles (dbf file) .....	20

Census Tract Aggregated Data (csv file) ..... 20

Disaggregated ZIP Code Microdata ..... 21

Disaggregated ZIP Code Microdata ..... 21

ZIP Code Shapefiles (dbf file) ..... 22

ZIP Code Aggregated Data (csv file)..... 22

Water Microdata ..... 23

Census Crosswalk Table ..... 24

# Documentation for RSEI Geographic Microdata (RSEI-GM)

---

The RSEI Geographic Microdata are a unique dataset that provides detailed air model results from the RSEI model at levels of geographic aggregation as small as 810 x 810 meter square grid cells. The Microdata can be very large, especially in their most disaggregated form, which is more than 100GB in size per year for the entire United States. Locations for downloading the RSEI data can be found on the RSEI website.

RSEI also produces Water Microdata, which provides toxicity-weighted concentrations for each stream segment downstream of a water discharge.

**Version Note: This documentation will continue to be updated, but will no longer be versioned. Beginning with reporting year 2018, the Microdata will be produced on a two-year cycle. In odd years, only grid-cell level Microdata for the most recent three-years will be produced.<sup>1</sup> In even years, an expanded set will be produced, including all years (1988-current) at different geographies (grid-cell, block group, census tract, and zip code), different sets (all chemicals, core chemicals/original industries, 2001 core chemicals). In some years, there may also be a spring dataset that supersedes the fall dataset.**

The 3-year, odd-year Microdata set will be available on the RSEI ftp site under the appropriate version number (<ftp://newftp.epa.gov/RSEI/>). The expanded data are available by request; use the [Contact Us form](#) on the RSEI website to send an email with your request.

---

<sup>1</sup> Some ad hoc data sets may also be produced for other reasons; they will also be posted as available.

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

## RSEI facility-level results

RSEI facility-level results are distributed in a compact form in tools like EasyRSEI and in applications like EPA's Envirofacts data warehouse. This dataset uses the same modeling results as the Microdata, but sums all of the grid-cell results for each release instead of presenting them separately. The release<sup>2</sup> is the basic unit of analysis; when you rank states or counties in EasyRSEI, you are actually ranking the sum of all modeled releases (and transfers) from all reporting facilities located in each state or county. Some of the releases may impact other states or counties, and some transfers to off-site incinerators may result in releases across the country from the reporting facility. The facility-level dataset is appropriate for ranking facilities, industries, chemicals, and screening for pollution prevention opportunities or further investigation. Ranking or looking at trends for geographic areas can be useful as long as users understand the facility-based nature of the data.

## RSEI Microdata results

With the Microdata, geographic analyses are more intuitive; when you rank states, you are ranking the impacts that occur within the geographic confines of each state, regardless of where the releases originate. Impacts from off-site transfers are attributed to the grid-cells that correspond to the area around the facility that receives the transfer and releases the chemical into the environment. The Microdata are provided for small geographic areas, and users can examine the impacts that releases from multiple facilities may have on the same area.

## Microdata users should familiarize themselves with RSEI methods and data

As with any model, RSEI is subject to the limitations of the underlying data sources and models that it incorporates. You should carefully consider the impact that the RSEI method may have on the results for any analysis. RSEI relies on TRI data, which only captures releases from certain industrial facilities; RSEI does not provide any information regarding many sources of environmental risk, including mobile sources, small industrial facilities, and hazardous waste (Superfund) sites.

The [RSEI website](#) contains important information about RSEI, and the RSEI methodology document and appendices contain full documentation of methods and data sources.

---

<sup>2</sup> More specifically, it is the release pathway; for instance, water releases can be split into different pathways for fish consumption and drinking water exposure.

## Outliers should be checked

EPA's TRI program conducts an [extensive data quality analysis](#) after TRI reporting forms are received. However, facilities often make revisions to their TRI reporting, especially to the most recent annual report. RSEI data uses the same dataset as the TRI National Analysis, which represents a point in time in the autumn following the July deadline for reports. [EPA's Envirofacts TRI search](#) may have more current data that should be checked for any revisions to anomalously high release amounts.

In addition, some high toxicity-weighted concentrations are driven by reporting of chemical groups, like glycol ethers or various groups of metal compounds like lead compounds or nickel compounds. Because specific data on the types of these compounds are not reported to TRI, RSEI models a worst-case scenario and assumes the most toxic form of the chemical is being released in most of these cases. Exceptions include PACs and chromium and chromium compounds, for which RSEI assumes a more likely release profile based on National Emissions Inventory (NEI) data<sup>3</sup>. In all of these cases, additional investigation needs to be done to determine the exact form of the chemicals being released before any conclusions about potential risk 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 period. The Chemical table (Chemical.csv, available in the "data tables" or "Public Release" datasets) has a series of fields (e.g., "Core01ChemicalFlag") to help in picking the appropriate chemical set for the time series; the year in the field name is the beginning of the period for which reporting requirements are constant. For instance, the '98 Core Chemical Flag' field is "1" (or "True") for any chemical whose reporting requirements have been constant over the period 1998-present. Users should limit their time series to those chemicals whose requirements have not changed over the period being considered; otherwise increases in reporting may erroneously imply increases in potential impacts.

Users should also be aware that seven high-scoring industry sectors were added for reporting year 1998, and so if a time series bridges that year, the new industries should be removed. There are various ways to define the new sectors, including by SIC code. The following SIC codes were added: 1041, 1044, 1061, 1099, 1221, 1231, 4911, 4931, 4939, 4953, 5169, 5171, and 7389. The RSEI Facility table (Facility.csv) also has a True/False field called NewIndustryFlag that is "1" (or "True") for facilities in industries added in 1998. [More information on the TRI industry expansion](#).

---

<sup>3</sup> The RSEI toxicity topic page has more information.

This document contains additional important information to consider when using TRI data: <https://www.epa.gov/toxics-release-inventory-tri-program/factors-consider-when-using-toxics-release-inventory-data>.

## Water Microdata

RSEI water microdata are produced at a different level than the air microdata. Instead of the grid cell, the water microdata use the stream segment as the unit of analysis. Before using the water microdata, users should familiarize themselves with the RSEI water modeling methodology and be aware that TRI does not include all chemical discharges to water (see the RSEI website for more information on the TRI universe and where to go for information on other releases).

The RSEI water modeling is indexed by stream segment Common ID (COMID), which is a unique identifier assigned by the [NHDPlus data set](#), Version 2. NHDPlus is based on the USGS's medium resolution National Hydrography Dataset (NHD), and adds important information like stream flow and velocity estimates.

The file "NHDMicroResults\_conc\_agg.csv" sums up the concentrations by release and COMID. In some cases, a stream path for a single release may split in two and then come together later downstream. In RSEI, those paths are calculated separately, but in this water Microdata file, the separate concentrations and associated metrics are summed (the disaggregated data are available upon request).

## 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:

- **Core-** limited to the set of chemicals whose reporting requirements have not changed over the entire reporting period (this is the smallest set of chemicals) and the set of original industries. Facilities in NAICS codes that correspond to the industries added in the 1998 expansion are not included.
- **Core01-** limited to the set of chemicals whose reporting requirements have not changed since 2001.
- **All chemicals-** includes all chemicals. Any file with no core specification in the name contains all chemicals.

The TRI program produces a list of all TRI chemical changes, which can be found on the TRI website here: <https://www.epa.gov/toxics-release-inventory-tri-program/changes-tri-list-toxic-chemicals>. Also, in the RSEI data tables, the table “Chemical” has a field “Added” which indicates the first TRI reporting year for that chemical (not the calendar year in which it was added), and fields like “Core01ChemicalFlag” which is “TRUE” if that chemical is on the list. Using these flags, users can extract any other core chemical subset from the set of all chemicals.

**Aggregations.** The Microdata can be presented at two levels of aggregation:

- **Disaggregated.** This is the most detailed level of data available, where each release is presented separately, so that users can see what chemicals and facilities are affecting each spatial unit (grid cell, ZIP code, etc.).
- **Aggregated.** In aggregated files, the toxicity-weighted concentration and score are summed over each spatial unit, so there is only one record per spatial unit. Because concentrations cannot be summed, concentration is not available at this level.

**Geography.** Microdata can be created for the following spatial units:

- **Grid cell.** The 810m by 810m basic unit of the RSEI model, defined by an x,y address. Section 5.1 of the [RSEI methodology document](#) has the technical specifications for the grid, as well as the equations for translating latitude and longitude into an (x,y) address. Grid cell shapefiles are available on the [RSEI ftp site](#):
- **Block group.** The 2010 Census boundaries are used.
- **Census tract.** The 2010 Census boundaries are used.
- **ZIP code.** 2017 ZIP code boundaries are used.

**Extent.**

- **National.** Most files are for the entire US, including island areas.
- **Conterminous U.S.** The aggregated grid-cell data contains only the conterminous United States, defined as the grid cells in grid 14.

- **State.** State shapefiles include all grid cells, regardless of modeled concentration, so that users can use the state shapefile as a set of x,y addresses to extract state-specific data from other grid-cell level files (e.g., for other years).

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

- **Comma-delimited text file.** See field descriptions at the end of this document.
- **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 a typical Arcview installation.

## Grid Cell Microdata files

### Disaggregated Microdata

**Description:** These are the raw Microdata files that contain the most disaggregated data possible. For each 810m grid cell, the file contains scores, concentrations, and toxicity-weighted concentrations for each chemical release. There may be multiple records for any one grid cell. Note that 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 grid cell, there will be separate records for each grid release.

**Naming:** These annual files are named as microXXXX\_YYYY, where XXXX is the reporting year for the data freeze, and YYYY is the year of the data contained in the file. So micro2014\_2010 is from the RY2014 RSEI update, and contains data for chemicals released in 2010. See the Disaggregated Grid Cell Microdata table in Appendix 1 for field names.

### Aggregated Microdata files

**Description:** Aggregated Microdata files use the same data as the disaggregated files, but sum the chemical releases over each grid cell. Because the values are summed, unweighted concentrations are not available (the sum of the concentrations of different chemicals would be meaningless).

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

**Naming:** These annual files are named aggmicroXXXX\_YYYY\_gcZZ, where XXXX is the reporting year for the data freeze, YYYY is the year of the data contained in the file, and ZZ is the grid code. So aggmicro2017\_2010\_GC14 is from the RY2017 RSEI update, and contains data for chemicals released in 2010 for grid 14 (the conterminous United States).

### State Microdata files

**Description:** These zip files contain shape files and associated files with aggregated grid cell level data for individual states. non

See the State Aggregated Grid Cell Microdata Shapefiles (.dbf files) table in Appendix 1 for field names.

**Naming:** These files are named ss\_aggmicroYYYY.zip, where ss is the 2-letter state abbreviation and YYYY is the 4-digit data year.

## Block Group Microdata files

### Disaggregated Block Group Microdata

**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 over Census block groups. For each Census defined block group, the file contains scores, concentrations, and tox-weighted concentrations for each chemical release. There may be multiple records for any one block group. Note that 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 reporting year for the data freeze, and YYYY is the year of the data contained in the file. So censusmicroblockgroup2017\_2010 is from the RY2017 RSEI update, and contains data for chemicals released in 2010.

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

### Block Group Shapefiles

**Description:** These files contain the same data as the disaggregated block group files, but the values are aggregated over all releases for each block group, so that there is only one record for each affected block group. 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.

See the Block Group Shapefiles table in Appendix 1 for field names.

**Naming:** These annual files are named as censusmicroblockgroupXXXX\_YYYY\_aggregated, where XXXX is the reporting year for the data freeze, and YYYY is the year of the data contained in the file. So CensusMicroBlockGroup2017\_2010\_aggregated is from the RY2017 RSEI update, and contains data for chemicals released in 2010.

### Aggregated Block Group Microdata

See block group shapefiles above; the aggregated block group data is available in the shapefile dbf, or as a separate csv. The fields differ slightly between the two files.

See the Aggregated Block Group Microdata table in Appendix 1 for field names.

## Census Tract Microdata files

### Disaggregated Census Tract Microdata

**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 over Census tracts. For each Census defined tract, the file contains scores, concentrations, and tox-weighted concentrations

for each chemical release. There may be multiple records for any one tract. Note that 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 tract, there will be separate records for each release.

**Naming:** These annual files are named as censusmicrotractsXXXX\_YYYY, where XXXX is the reporting year for the data freeze, and YYYY is the year of the data contained in the file. So CensusMicroTracts2017\_2010 is from the RY2017 RSEI update, and contains data for chemicals released in 2010.

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

### Census Tract Shapefiles

**Description:** These files contain the same data as the disaggregated Census tract files, but the values are aggregated over all releases for each Census tract, so that there is only one record for each affected Census tract. 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.

See the Census Tract Shapefiles table in Appendix 1 for field names.

**Naming:** These annual files are named as censusmicrotractsXXXX\_YYYY\_aggregated, where XXXX is the reporting year for the data freeze, and YYYY is the year of the data contained in the file. So CensusMicroTracts2017\_2010 is from the RY2017 RSEI update, and contains data for chemicals released in 2010.

### Aggregated Census Tract Microdata

See Census tract shapefiles above; the aggregated block Census tract is available in the shapefile dbf, or as a separate csv. The fields differ slightly between the two files.

See the Aggregated Census Tract Microdata table below for field names.

## ZIP Code Microdata files

### Disaggregated ZIP Code Microdata

**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 over U.S. ZIP codes. For each ZIP codes, the file contains scores, concentrations, and tox-weighted concentrations for each chemical release. There may be multiple records for any one ZIP codes. Note that 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 grid release.

**Naming:** These annual files are named as censusmicrozipcodeXXXX\_YYYY, where XXXX is the reporting year for the data freeze, and YYYY is the year of the data contained in the file. So CensusMicroZipcode2017\_2010 is from the RY2017 RSEI update, and contains data for chemicals released in 2010.

See the Disaggregated ZIP codes Microdata table below for field names.

### ZIP Code Shapefiles

**Description:** These files contain the same data as the disaggregated ZIP codes files, but the values are aggregated over all releases for each ZIP codes, so that there is only one record for each affected ZIP codes. These files do not contain chemical- or facility-specific information, except the number of facilities and chemicals affecting each ZIP codes. These files also contain percentile fields to help in mapping.

See the ZIP codes Shapefiles table in Appendix 1 for field names.

**Naming:** These annual files are named as censusmicrozipcodeXXXX\_YYYY\_aggregated where XXXX is the reporting year for the data freeze, and YYYY is the year of the data contained in the file. So CensusMicroZipCode2017\_2010 is from the RY2017 RSEI update, and contains data for chemicals released in 2010.

### Aggregated ZIP codes Microdata

See ZIP codes shapefiles above, the aggregated ZIP codes data is available in the shapefile dbf, or as a separate csv. The fields differ slightly between the two files.

See the Aggregated ZIP codes Microdata table below for field names.

## Water Microdata files

This file contains the toxicity-weighted concentrations downstream of TRI discharges by stream segment. All years of data are contained in the file, which is named NHDMicroResults\_conc\_agg. For some years, the files may be split into offsite and onsite facilities. Offsites include POTWs and onsites include direct water discharges from TRI reporting facilities.

## Census Crosswalk Files

Each set of crosswalk files links the RSEI grid cell geography to a different US decennial census year. There is one crosswalk for each area and decennial Census year (1990, 2000, 2010). Crosswalk files are named by area (Alaska, Con(terminous) US, etc.). The last three fields in each file contain percent values that can be used to adjust the block or cell contents when performing the crosswalk. PCT\_B\_C and PCT\_C\_B are area-weighted and can be used for metrics that do not involve population, such as concentration and toxicity-weighted concentration. PCT\_PC\_B is population weighted, and can be used to crosswalk fields that involve population, like score and pop. Note that the "PCT\_CP\_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, RSEI uses 2000

block boundaries and scales each cell's population by the overall ratio of 1990/2000 population for each area.

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

## Public Data Release Files (data tables)

These tables contain TRI reporting data and other data used in modeling, including facility locations and chemical toxicity. Figure 1 below shows the relationships between the main public data release tables. The [RSEI Data Dictionary](#) provides field descriptions.

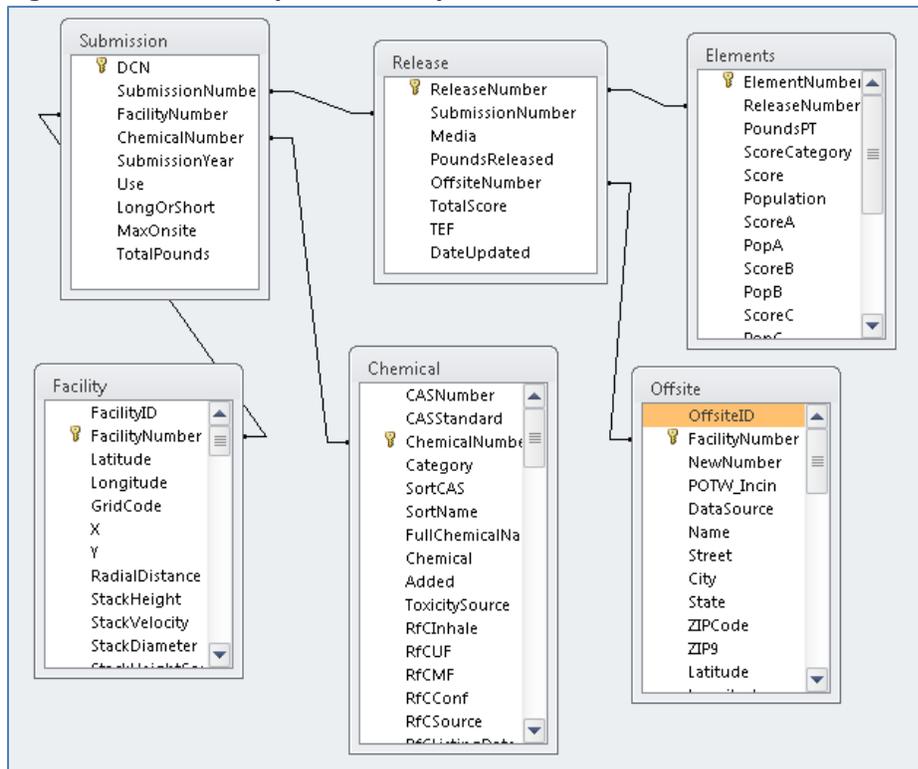
## Grid Shape files

Shape files of the RSEI grid geography are available by region, in two versions: polygon and center point. The [RSEI Data Dictionary](#) provides field descriptions.

## Linking Tables

**The Microdata must be used together with the other data tables provided with the RSEI public data release of the same vintage.** For instance, you must use the RY2018, V 2.3.8 data tables with the RY2018, V 2.3.8 Microdata. Each data table links into the Microdata using a key field (like ChemicalNumber, FacilityNumber, etc.), as indicated in Figure 1, below. Field descriptions can be found in the [RSEI Data Dictionary](#). The public data release files allow you to extract parameters like chemical name, CAS number, facility name, TRI facility ID (TRIFID), facility latitude/longitude, and any other associated data contained in the RSEI databases.

Figure 1. Relationships between public data release tables



## RSEI recent version numbers and TRI reporting years (RY)

- V2.3.1- RY 2010
- V2.3.2- RY 2011
- V2.3.3- RY 2012
- V2.3.4- RY 2014
- V2.3.5- RY2015
- V2.3.6- RY2016
- V2.3.7- RY2017
- V2.3.8-RY2018
- V2.3.9-RY2019

# Appendix 1. Field Descriptions for Microdata Tables

These 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 table "Release") <sup>1</sup> .
5	ChemicalNumber	Internal unique identifier of released chemical (lookup in table "Chemical") <sup>1</sup> .
6	FacilityNumber	Internal unique identifier of releasing facility (If media = 1 or 2, then lookup in table "Facility"; If media = 6 or 750 or 754, then lookup in table "Offsite") <sup>1</sup> .
7	Media	Code describing media into which chemical is released. (lookup in table "Media") <sup>1</sup> .
8	Conc	Concentration of chemical for release/media at grid cell ( $\mu\text{g}/\text{m}^3$ ).
9	ToxConc	Concentration multiplied by inhalation toxicity weight.
10	Score	Risk-related score (surrogate dose * toxicity weight * population).
11	ScoreCancer	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for cancer effects.
12	ScoreNonCancer	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for noncancer effects.

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
13	Pop	Number of people in grid cell (may be interpolated).

<sup>1</sup> 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.

## Aggregated Grid Cell Microdata

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 cell.
7	Score	Risk-related score (surrogate dose * toxicity weight * population), summed over all chemicals impacting cell.
8	Pop	Population of grid cell.
9	ScoreCancer	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for cancer effects.
10	ScoreNonCancer	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for noncancer effects.

## State Aggregated Grid Cell Microdata Shapefiles

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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.

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
9	TOXCONC	Concentration multiplied by inhalation toxicity weight, summed over all chemicals impacting cell.
10	SCORE	Risk-related score (surrogate dose * toxicity weight * population), summed over all chemicals impacting cell.
11	EXPOP	Population of grid cell, for cells with nonzero concentrations only.
12	CSCORE	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for cancer effects.
13	NCSCORE	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for noncancer effects.

## Disaggregated Block Group Microdata

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

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
<sup>1</sup> Other tables can be found in the public data set, on the RSEI ftp site ( <a href="ftp://newftp.epa.gov/RSEI/">ftp://newftp.epa.gov/RSEI/</a> ) or in the RSEI data dictionary ( <a href="https://www.epa.gov/rsei/rsei-data-dictionary">https://www.epa.gov/rsei/rsei-data-dictionary</a> ). Be sure to use the data tables from the same RSEI version as the Microdata.		

## Block Group Shapefiles (dbf file)

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 (surrogate dose * toxicity weight * 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 (surrogate dose * toxicity weight * 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 (surrogate dose * toxicity weight * 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)

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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.

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
6	SCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the block group.
7	POP	Sum of the population of the cells in the block group.
8	CSCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the block group. Score is calculated using only cancer toxicity weights.
9	NCSCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the block group. Score is calculated using only noncancer toxicity weights.

## Disaggregated Census Tract Microdata

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

<sup>1</sup> 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 Tract Shapefiles (dbf file)

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 (surrogate dose * toxicity weight * 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 (surrogate dose * toxicity weight * 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 (surrogate dose * toxicity weight * 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)

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 (surrogate dose * toxicity weight * population) of the cells in the tract.

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
7	POP	Sum of the population of the cells in the tract.
8	CSCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the tract. Score is calculated using only cancer toxicity weights.
9	NCSCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the tract. Score is calculated using only noncancer toxicity weights.

## Disaggregated ZIP Code Microdata

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

<sup>1</sup> 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.

## ZIP Code Shapefiles (dbf file)

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 (surrogate dose * toxicity weight * 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 (surrogate dose * toxicity weight * 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 (surrogate dose * toxicity weight * 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)

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 (surrogate dose * toxicity weight * population) of the cells in the ZIP code.

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
7	POP	Sum of the population of the cells in the ZIP code.
8	CSCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the ZIP code. Score is calculated using only cancer toxicity weights.
9	NCSCORE	Sum of the risk-related scores (surrogate dose * toxicity weight * population) of the cells in the ZIP code. Score is calculated using only noncancer toxicity weights.

## Water Microdata

<i>Field Number</i>	<i>Name</i>	<i>Description</i>
1	ReleaseNumber	Internal unique identifier for release  (links to Release table).
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
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.

## Census Crosswalk Table

Census Crosswalk Table		
<i>Field Number</i>	<i>Name</i>	<i>Description</i>
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 2/16/2021]