RSEI Data Dictionary

Contents

Facility-level Data Tables 2

Facility 2
Off-site 5
Standard Industrial Classification (SIC) 6
North American Industry Classification System (NAICS) 6
Chemical 7
Maximum Contaminant Level (MCL) 11
Media 12
Submission 12
Release 12
Elements 13
Category 14

RSEI Geographic Microdata 15

Disaggregated Microdata- 15
Aggregated Microdata 16
Averaged Block Group Microdata 17
Water Microdata 18

Other Available Data 18

Census Crosswalks 18
Population Data (US Decennial Census) 19
Shapefiles (Grid geography) 20

RSEI Data Dictionary

This document describes all of the data tables and fields used in the RSEI model and results data sets. Additional information can be found in the RSEI methodology document.

Facility-level Data Tables

This dataset presents RSEI results at the facility release level. These tables are distributed in the RSEI Queries database as well as in a set of flat csv files. These tables link to the Geographic Microdata using keys like FacilityNumber, ChemicalNumber, and ReleaseNumber. Note that the key values change with each version of RSEI, so you must use the same version of these tables as the Microdata.

Facility

The facility table contains data for reporting facilities, including location, stack parameters and discharge reach, and is also available in EasyRSEI. Note that, with Version 2.3.5, SIC codes and some facility identifiers (RCRA ID, NPDES ID) are no longer maintained. Users interested in these fields can find them in EPA's FRS (Facility Registry Service) system, identifying the facility with the FRS ID listed in the facility table,

Facility Data	
Variable Name	***Description***
FacilityID	Unique TRI identifier for facility (TRI Facility ID).
FacilityNumber	Internal identifier unique to each facility.
Latitude	Final latitude of the facility in decimal degrees used for modeling.
Longitude	Final longitude of the facility in decimal degrees used for modeling.
GridCode	Number that identifies the model grid within which the cell is located.
X	Assigned grid value based on latitude.
Y	grid value based on longitude.
RadialDistance	Distance from approximate center point of grid.
StackHeight	Height of facility stack that is emitting the pollutant (m).
StackVelocity	Rate at which the pollutant exits the stack (m/s).
StackDiameter	Diameter of facility stack that is emitting the pollutant (m).
StackHeightSource	Source of information on stack height.
StackVelocitySource	Source of information on stack velocity.
StackDiameterSource	Source of information on stack diameter.
NEIYear	National Emissions Inventory (NEI) version year, if NEI data were used for stack parameters.
FacilityName	TRI facility name.
Street	Street address of facility.
City	City where the TRI facility is located.
County	County where the TRI facility is located.
State	State in which the facility is located.
FIPS	FIPS (Federal Information Processing Standard) code which identifies the county associated with the facility.

Facility Data

Variable Name	*Description*
STFIPS	FIPS (Federal Information Processing Standard) code which identifies the state associated with the facility.
ZIPCODE	Five-digit facility ZIP code.
ZIP9	Nine digit facility ZIP code, if reported.
DUNS	The 9-digit number assigned by Dun & Bradstreet for the facility or establishment within the facility.
REGION	EPA region where facility is located.
FederalFacilityFlag	Code describing federal status for purposes of Executive Order 12856.
FederalAgencyName	Name of Federal Agency of which the federal facility is a part.
ParentName	Name of the corporation or other business entity located in the U.S. that directly owns at least 50 percent of the voting stock of the facility.
ParentDUNS	The 9-digit number assigned by Dun & Bradstreet for the US parent company.
PublicContactName	Name submitted by TRI facility as public contact.
PublicContactPhone	Phone number submitted by TRI facility for public contact.
PCT_CH6	Percent of chromium released that is assumed to be hexavalent (the remainder is assumed to be trivalent with negligible toxicity and not modeled.
ChromHexPercent	Percent of chromium released that is assumed to be hexavalent (the remainder is assumed to be trivalent with negligible toxicity and not modeled (same as PCT_CH6).
ChromSource	Source for PCT_CH6/ChromHexPercent.
ModChromReleases	True if facility has released or transferred chromium or chromium compounds to modeled media (fugitive/stack air releases, direct water, POTWs or off-site incineration).
NewIndustryFlag	True if the facility's primary SIC was added to TRI in the TRI industry expansion beginning in reporting year 1998.
NAICS1	Facility's 6-digit NAICS code designated as "primary" by facility on Form R. If no primary NAICS is designated, the field displays "NR."
NAICS2	Facility's most frequently reported non-primary 6-digit NAICS code.
NAICS3	Facility's second most frequently reported non
NAICS4	Facility's third most frequently reported non-primary 6-digit NAICS code.
NAICS5	Facility's fourth most frequently reported non
NAICS6	Facility's fifth most frequently reported non-primary 6-digit NAICS code.
NAICSCode3Digit	First 3 digits of facility's primary NAICS code.
NAICSCode4Digit	First 4 digits of facility's primary NAICS code.
NAICSCode5Digit	5
FRSID	EPA's Facility Registry System ID.
AssignedReach	NHDPlus reach identifier for final facility discharge reach.
AssignedCOMID	NHDPlus segment identifier for final facility discharge reach.
ReachSource	Source for final discharge assignment.
OutfallLatitude	Latitude for outfall.
OutfallLongitude	Longitude for outfall.
OutfallSource	Source for outfall coordinates.

Facility Data	
Variable Name	***Description***
NearReach	NHDPlus reach identifier for nearest discharge reach.
NearCOMID	segment identifier for nearest discharge reach.
NPDESReach	NHDPlus reach identifier for discharge reach as reported to ICIS-NPDES.
NPDESCOMID	segment identifier for discharge reach reported to ICIS
NPDESYear	Year of ICIS-NPDES data used.
DistanceToReach	The distance between an off site facility discharging to water and the reach of the receiving water body (m).
HEM3ID	The ID assigned to the nearest National Weather Service (NWS) observation station.
DistanceToHEM3	The distance between a facility and the nearest observation station.
LatLongSource	Source of final lat/long found in 'Latitude' and 'Longitude' fields.
LLYear	Year of lat/long data.
LLNotes	Notes for facility location.
DistanceToTribalLand	Distance to nearest Tribal Land within ten miles (miles).
TribalLandName	Name of nearest Tribal Landwithin ten miles, if any.
WaterReleases	True if facility reported direct water releases for any year 1988-2015.

Off-site

The Off-site table contains the condensed list of quasi-unique off-site facilities to which TRI reporters transfer waste. Only incinerators and POTWs are modeled by RSEI, so verification of addresses and locations are focused on those off-site facilities.

Off-site Data	
Variable	***Description***
OffsiteID	Unique internal identifier for each off-site facility.
FacilityNumber	Unique internal identifier for each off-site facility. [Note this is different from the FacilityNumber field in the Facility table]
NewNumber	This variable is not yet active.
POTW_Incin	Identifies off-site facilities for which releases are modeled: 1= POTW; 2=Incinerator; 3=POTW and Incinerator.
DataSource	This variable is not yet active.
TRIFID	TRI identifier, if applicable, for facility. Note this ID is not necessarily unique in this table, because there may be more than one record per facility (FacilityID is unique in the Facility table).
Name	Best submitted name for off-site facility.
Street	Best submitted street address for
City	Best submitted city for off-site facility.
State	Best submitted state for off
ZIPCode	Best submitted ZIP code for offsite facility.
ZIP9	This variable is not yet implemented.
Latitude	Geocoded latitude in decimal degrees for off-site facility.
Longitude	Geocoded longitude in decimal degrees for off
GridCode	Number that identifies the model grid within which the cell is located.
X	Assigned grid value based on latitude.
Y	Assigned grid value based on longitude.
Radial Distance	Distance from approximate center point of grid.
StackHeight	Default stack height used for off-site facilities.
StackVelocity	Default stack velocity used for off
StackDiameter	Default stack diameter used for off-site facilities.
TempField	This variable is not implemented.
FRSID	EPA's Federal Registry System ID for off-site facility.
NPDESPermit	Permit number issued by US EPA for facilities discharging to water.
RCRANumber	Number assigned by EPA to facilities handling hazardous waste under the Resource Conservation and Recovery Act.
HEM3ID	The ID assigned to the nearest National Weather Service (NWS) observation station.
DistanceToHEM3	The distance between a facility and the nearest observation station.
WaterReleases	Not currently updated.
OutfallLongitude	Latitude associated with end of the pipe used for off-site facility's discharge to water.
OutfallLatitude	Longitude associated with end of the pipe used for off-site facility's discharge to water.

Off-site Data	
Variable	***Description***
NearReach	14-digit NHDPlus reach identifier associated with the reach that is nearest to off site facility.
NearComID	ComID from NHDPlus dataset that uniquely identifies reach segment nearest facility.
DistanceToReach	The distance between an off-site facility discharging to water and the reach of the receiving water body (m).
AssignedReach	14-digit NHDPlus reach identifier associated with reach assigned by EPA or determined through QA.
AssignedComID	ComID from NHDPlus dataset that uniquely identifies reach segment for assigned reach.
StreamSource	Data source linking stream reach to facility.
LocationType	Type of geocoded match.
LatLongSource	Source used to determine lat/longs.
LatLongYear	Year lat/long was last updated.
Quality	Rank from 1 to 9 describing quality of geocoded match (1 is best).
Freq	Number of TRI transfers sent to this off-site facility.
NearStream	USGS Reach Identifier from RF1 stream network (not currently used).
DistanceToStream	The distance between the facility and its RF1 stream reach (m) (not currently used).
WBANID	The ID assigned to the Weather Bureau/Army/Navy WeatherStation nearest to the facility.
DistanceToWBAN	The distance between a facility and the nearest WBAN weather station (m).
Notes on Coordinates	Notes on how lat/long was derived.

Standard Industrial Classification (SIC)

This table is no longer maintained in RSEI. NAICS codes are now used to determine industry-level stack heights and chromium speciation rates.

North American Industry Classification System (NAICS)

NAICS codes are collected by TRI.

NAICS Data	
Variable	***Description***
NAICSCode	Six-digit NAICS code.
LongName	Text description of code.

Chemical

The chemical table contains data for chemicals reported to TRI, including toxicity, physico-chemical properties, and flag fields to facilitate user selections. The chemical table is also available in EasyRSEI.

Chemical Data	
Field Name	***Field Description***
CASNumber	Chemical Abstracts Service Registry Number, which identifies a unique chemical. For chemical categories, CAS Numbers begin with "N", followed by three digits.
CASStandard	The Chemical Abstracts Service Registry Number identifies a unique chemical. The standard format contains three sets of numbers divided by hyphens (00-00-0).
ChemicalNumber	Unique internal identifier.
Category	This identifier is not yet active.
SortCAS	Chemical Abstracts Service Registry Number, which identifies a unique chemical, formatted for sorting (no hyphens). For chemical categories, CAS Numbers begin with "N", followed by three digits.
SortName	Common name of chemical, with initial modifiers moved to end of name. Used for internal sorting purposes.
FullChemicalName	Full scientific name(s) of the chemical.
Chemical	Common name(s) of the chemical.
Added	The year the chemical was added to the Toxics Release Inventory
Toxicity Source	All sources used for toxicity data, and date of addition to database.
RfCInhale	The inhalation reference concentration (RfC) is defined as "an estimate (with uncertainty spanning perhaps an order of magnitude) of a continuous inhalation exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious noncancer health effects during a lifetime". Units are mg/m3.
RfCUF	The uncertainty factor (UF) is applied to the no-observed-adverse-effect level (NOAEL) upon which the RfC is based, thereby reducing the dose. The UF accounts for uncertainties in extrapolation from experimental data to an estimate appropriate to humans.
RfCMF	The modifying factor (MF) is a value applied to the NOAEL when scientific uncertainties in the study chosen for estimating the RfC are not explicitly addressed by the standard UFs.
RfCConf	Confidence levels are assigned to the study used to derive the RfC, the overall database, and to the RfC itself.
RfCSource	Source used for the RfC value.
RfCListingDate	Date that RfC was listed, if available.
RfCToxWeight	Toxicity weight based on the RfC (RfCToxWeight = 3.5/RfC). Noncancer/inhalation.
RfDOral	The oral reference dose (RfD) is "an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure [by ingestion] to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime". (mg/kg-day)

Chemical Data	
Field Name	***Field Description***
RfDUF	The uncertainty factor (UF) is applied to the no-observed-adverse-effect level (NOAEL) upon which the RfD is based, thereby reducing the dose. The UF accounts for uncertainties in extrapolation from experimental data to an estimate appropriate to humans.
RfDMF	The modifying factor (MF) is a value applied to the NOAEL when scientific uncertainties in the study chosen for estimating the RfD are not explicitly addressed by the standard UFs.
RfDConf	Confidence levels are assigned to the study used to derive the RfD, the overall database, and to the RfD itself.
RfDListingDate	Date that RfD was listed, if available.
RfDSource	Source used for the RfD value.
RfDToxWeight	Toxicity weight based on the RfD (RfDToxWeight = 1/RfD). Noncancer/oral.
UnitRiskInhale	The unit inhalation risk is the excess lifetime risk due to a "continuous constant lifetime exposure of one unit of carcinogen concentration"(51 FR 33998). (1/mg/m3)
QSTAROral	The oral cancer slope factor (q1*) or oral slope factor (OSF): a measure of the incremental lifetime risk of cancer by oral intake of a chemical, expressed as risk per mg/kg-day. (1/mg/kg-day)
WOE	Weight of evidence (WOE) categories indicate how likely a chemical is to be a human carcinogen, based on considerations of the quality and adequacy of data and the type of responses induced by the suspected carcinogen. EPA WOE classifications include the following categories and associated definitions (51 FR 33996): A Carcinogenic to humans B Probable carcinogen based on: •B1 Limited human evidence •B2 Sufficient evidence in animals and inadequate or no evidence in humans: C Possible carcinogen D Not classifiable E Evidence of non-carcinogenicity
UnitRiskListingDate	Date that Unit Risk was listed, if available.
UnitRiskSource	Source used for the Unit Risk value.
IURToxWeight	Toxicity weight based on the IUR (IURToxWeight = IUR/2.8e-7). Cancer/inhalation.
QStarListingDate	Date that QStar was listed, if available.
QStarSource	Source used for the QStar value.

Chemical Data	
Field Name	***Field Description***
OSFToxWeight	Toxicity weight based on the QStar or OSF (OSFToxWeight = QSTAROral/1e-6). Cancer/oral.
WOEListingDate	Date that WOE was listed, if available.
WOESource	Source used for the WOE classification.
ITW	Inhalation Toxicity Weight: the RSEI toxicity weight for a chemical for the inhalation pathway.
OTW	
ToxicityClassOral	This indicates whether the toxicity weight for the oral pathway is based on cancer or noncancer health effects.
ToxicityClassInhale	This indicates whether the toxicity weight for the inhalation pathway is based on cancer or noncancer health effects.
ToxicityCategory	This indicates whether the oral and inhalation toxicity weights are based on cancer health effects, non-cancer health effects, or both.
AirDecay	The rate at which a chemical degrades in air, due primarily to photooxidation by radicals (hr-1).
Koc	The organic carbon-water partition coefficient, used in estimates of chemical sorption to soil (mL/g).
H2ODecay	The rate at which a chemical degrades in water, due to abiotic hydrolysis, biodegradation, or photolysis (hr-1).
LOGKow	The logarithm of the octanol-water partition coefficient. Kow is the ratio of a chemical's concentration in the octanol phase to its concentration in the aqueous phase at equilibrium in a two-phase octanol/water system.
Kd	The soil water partition, or distribution, coefficient. For organics, the value is often estimated as the product of Koc and foc (the fraction of organic carbon in the soil) (L/kg).
WaterSolubility	The amount of chemical that dissolves in water at a particular temperature (mg/L).
POTWPartitionRemoval	Percent of chemical removed from the wastewater by the POTW (Publicly Owned Treatment Works).
POTW PartitionSludge	Percent of total POTW removal efficiency attributable to sorption of the chemical to sewage sludge.
POTW PartitionVolat	Percent of total POTW removal efficiency attributable to volatilization of the chemical.
POTW PartitionBiod	Percent of total POTW removal efficiency attributable to biodegradation of the chemical.
IncineratorDRE	Destruction/removal efficiencies, expressed as the percent of chemical fed to the incinerator that is not released to the air.
BCF	Bioconcentration factor: the ratio of a chemical's concentration in fish to its concentration in water at equilibrium (L/kg).
Henrys	Henry's law constant: the ratio of a chemical's concentration in the air to its concentration in the water at equilibrium (atm·m3/mol).

Chemical Data	
Field Name	***Field Description***
MCL	Maximum Contaminant Level, which is EPA's national primary drinking water standard for the chemical. This is the current value; historical data are contained in the table, 'MCL.'
Molecular Weight	The mass in grams of one mole of molecules of the chemical.
T33/50Flag	This flag is a marker which indicates that the chemical is included in EPA's 33/50 program, a program in which facilities voluntarily reduce their chemical releases by 33 percent and 50 percent by certain dates.
HAPFlag	This flag marks the chemicals that are hazardous air pollutants, as defined by the Clean Air Act.
CAAFlag	This flag marks the chemicals that are Clean Air Act pollutants.
PriorityPollutantFlag	This flag marks the chemicals that are priority pollutants, as defined by the Clean Water Act.
SDWAFlag	This flag marks the chemicals that have national primary or secondary drinking water standards under the Safe Drinking Water Act.
CERCLAFlag	This flag marks the chemicals that are regulated under Superfund (CERCLA—the Comprehensive Environmental Response, Compensation, and Liability Act).
OSHACarcinogens	This flag indicates whether the chemical is a known or suspect human carcinogen based on OSHA criteria. Known human carcinogens are defined as those that have been shown to cause cancer in humans. Suspect human carcinogens have been shown to cause cancer in animals. The list of chemicals flagged as OSHA carcinogens is based on the list of carcinogens provided in the 1997 TRI Public Data Release.*
ExpansionFlag	This flag marks the chemicals that were added to the Section 313 toxic chemical list for the 1995 Reporting Year.
Core88ChemicalFlag	This flag marks the chemicals that are common to all reporting years of TRI and that have had no modifications of reporting requirements, as determined by the 1988 Core Chemical List found on the TRI Explorer website.
Core95ChemicalFlag	This flag marks the chemicals that are common to TRI reporting years 1995 through the current year and that have had no modifications of reporting requirements in that time period, as determined by the 1995 Core Chemical List found on the TRI Explorer website.
Core98ChemicalFlag	This flag marks the chemicals that are common to TRI reporting years 1998 through the current year and that have had no modifications of reporting requirements in that time period, as determined by the 1998 Core Chemical List found on the TRI Explorer website.
Core00ChemicalFlag	This flag marks the chemicals that are common to TRI reporting years 2000 through the current year and that have had no modifications of reporting requirements in that time period.
Core01ChemicalFlag	This flag marks the chemicals that are common to TRI reporting years 2001 through the current year and that have had no modifications of reporting requirements in that time period. The only difference between this flag and the Core00ChemicalFlag is the inclusion of lead and lead compounds.

Chemical Data	
Field Name	***Field Description***
HPVFlag	Indicates whether the chemical is designated as a High Production Chemical.
HPVChallengeValue	Describes the value or combination of values assigned to the chemical by EPA's HPV Challenge program to describe the chemical's status under the program.
PBTFlag	Indicates whether EPA has designated this chemical as a priority chemical under the Persistent Bioaccumulative and Toxic (PBT) Chemical Program.
Metal	This flag indicates whether the chemicals are metals and also whether they are core chemicals. (Core chemicals are those that are common to all reporting years of TRI and which have had no modifications of reporting requirements.)
User Tags 1 through 5	Using these tags, you can select a set of chemicals based on your own selection criteria.
HasTox	Indicates that the chemical has a toxicity weight (either oral or inhalation) in the data set.
MaxTW	Shows the greater of the two possible toxicity weights (oral or inhalation).

Maximum Contaminant Level (MCL)

MCLs are used to cap maximum concentrations in drinking water systems.

MCL Data	
Variable	***Description***
CASNumber	Chemical Abstracts Service Registry Number, which identifies a unique chemical. For chemical categories, CAS Numbers begin with "N", followed by three digits.
CASStandard	The Chemical Abstracts Service Registry Number identifies a unique chemical. The standard format contains three sets of numbers divided by hyphens (00-00-0).
ChemicalNumber	Unique internal identifier.
	Common name(s) of the chemical.
MCL1988...MCL2016	MCL for each year an MCL was in effect.

Media

The media table provides descriptions for the media codes used in the Release table.

Media Data	
Variable	***Description***
Media	Code associated with the media and/or method of release, as reported by facility in TRI Reporting Form R. See Media Information for codes.
MediaText	Descriptions of receiving media associated with Media Codes
Sum	This variable is not yet implemented.
Itw	Internal dummy variable used for modeling.
Otw	Internal dummy variable used for modeling.
Mtw	Internal dummy variable used for modeling.
MediaCode	Internal variable used for modeling.

Submission

The submission table contains Form R information submitted to TRI.

Submission Data	
Variable	***Description***
DCN	Unique identifier assigned by TRI to each facility submission (document control number).
SubmissionNumber	Internal identifier assigned to each submission.
FacilityNumber	identifier unique to each facility (links to Facility table).
ChemicalNumber	Internal identifier unique to each chemical (links to Chemical table).
SubmissionYear	Year of facility release.
Use	Code describing how chemical is used in reporting facility, as reported on TRI Reporting Form R. See On-site Chemical Information for an explanation of the codes.
LongOrShort	Code describing whether the submission came from a short or long form.
MaxOnsite	Code describing the maximum amount of the chemical on-site at reporting facility, as reported in TRI Reporting Form R. See On-site Chemical Information for an explanation of the codes.
TotalPounds	Total pounds released. (not currently used, see PoundsPT in Elements table.

Release

This table contains data for each chemical release. There can be multiple release records per submission record.

Release Data	
Variable	***Description***
ReleaseNumber	Unique internal identifier.

Release Data	
Variable	***Description***
SubmissionNumber	Unique internal identifier (links to Submission table).
Media	Code associated with the media and/or method of release, as reported by facility in TRI Reporting Form R. See Media table for explanation of codes.
PoundsReleased	Total pounds released, without accounting for treatment.
OffsiteNumber	Unique identifier for off-site facility receiving this release, if any. Links to Facility Number in the Off site table.
TotalScore	Total score (risk-related result) for release. (not currently populated- use Score in Elements table.)
TEF	Toxicity Equivalency Factor used to adjust toxicity for dioxins.
DateUpdated	Null

Elements

The Elements table contains the calculated results for each release. There can be multiple elements records for each release. Note that all values in the elements table are rounded to six significant figures.

Elements Data	
Variable	***Description***
ElementNumber	Unique internal identifier.
ReleaseNumber	Unique internal identifier (links to Release table).
PoundsPT (TRI Pounds)	Total pounds after any treatment by POTWs or other offsite facilities.
ScoreCategory	Codes corresponding to the medium into which the chemical is released. Examples of the information include: direct air releases from the stack using a "rural" air dispersion model, fugitive air releases, releases to an onsite landfill. [See Score Category Information for descriptions]
Score	Total Indicator Element score- modeled surrogate dose multiplied by toxicity weight and by population, using the higher cancer/noncancer toxicity weight for each air/water pathway.
Population	Total population exposed.
ScoreA	Score for children 0 through 9 years of age (inclusive).
PopA	Number of exposed children
ScoreB	Score for children 10 through 17 years of age (inclusive).
PopB	Number of exposed children
ScoreC	Score for adults 18 through 44 years of age (inclusive).
PopC	Number of exposed
ScoreD	Score for adults 45 through 64 years of age (inclusive).
PopD	Number of exposed
ScoreE	Score for adults 65 years old and greater.
PopE	Number of exposed .
NCScore (NonCancer Score)	Indicator Element score, limited to chemicals with non-cancer endpoints.
CScore (Cancer Score)	Indicator Element score, limited to chemicals with cancer endpoints.

Elements Data	
Variable	***Description***
Hazard	Toxicity weight times TRI pounds, using the higher cancer/noncancer toxicity weight for each air/water pathway.
HazardC (Cancer Hazard)	Toxicity weight times TRI pounds, limited to chemicals with cancer endpoints.
HazardNC (Non-Cancer Hazard)	Toxicity weight times TRI pounds, limited to chemicals with non-cancer endpoints.

Category

The Category table describes the codes used in the Elements table to indicate the release pathway.

Category Data	
Variable	***Description***
ScoreCategory	Codes corresponding to the medium into which the chemical is released. Examples of the information include: volatilization from a transfer to a POTW, fugitive air releases, releases to an onsite landfill.
Category	Descriptions of release media and other descriptors corresponding with the score category codes.
Model	A dummy variable that is '1' when that category can be modeled and '0' when it cannot.
InhaleTox	A dummy variable that is '1' when the model requires an inhalation toxicity score to model this kind of release and '0' when it does not.

RSEI Geographic Microdata

A separate guidance is available for use with the Microdata.

Disaggregated Microdata-

These are the raw Microdata files that contain the most disaggregated data possible. For each 810m grid cell, the file contains scores, concentrations, and tox-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 have historically been named 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 Micro 2014_2010 is from the RY2014 RSEI update, and contains data for chemicals released in 2010. The new naming convention substitutes the version number for the version year, as in vXXX_micro_YYYY, where XXX is the version number and YYYY is the year of the data contained in the file; for example v234_micro_2014.csv.There is one annual file for the entire country, which is over 100 GB in size.

Disaggregated Microdata Table

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”)*
5	ChemicalNumber	Internal unique identifier of released chemical (lookup in table “Chemical”)*
6	FacilityNumber	Internal unique identifier of releasing facility (lookup in table “Facility” if media = 1 or 2; if media = 6 or 750 or 754, then lookup in table “Offsite”)*
7	Media	Code describing media into which chemical is released. (lookup in table “Media”)*
8	Conc	Concentration of chemical for release/media at grid cell.
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

Disaggregated Microdata Table		
Field Number	***Name***	***Description***
12	ScoreNonCancer	Risk-related score (surrogate dose * toxicity weight * population) using only toxicity values for noncancer effects
13	Pop	Number of people in grid cell (may be interpolated)

Aggregated Microdata

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). **Naming:** These annual files have historically been named 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 Micro 2014_2010 is from the RY2014 RSEI update, and contains data for chemicals released in 2010. The **new naming convention** substitutes the version number for the version year, as in vXXX_micro_YYYY, where XXX is the version number and YYYY is the data year; for example v234_micro_2014.csv.These files have historically been named in the format 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 2-digit grid code (see Field 1 in the Table 1 below for grid codes). The **new naming convention** substitutes the version number for the version year, as in vXXX_aggregated_micro_gcZZ_YYYY; for example, v234_aggregated_micro_gc14_2014.csv.

Aggregated Microdata Table		
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 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 cancer effects

Averaged Block Group Microdata

These files are the same as the aggregated Microdata files, but instead of being presented at the grid cell level, the values are averaged over Census block groups. The file BG_RSEI_XXXX_3yr is a csv file with the block group-level data averaged over 2012 through 2014. There are also shape files (tl_2010_bg_US_RSEI) with the same data; that is, the .dbf file and the .csv have the same fields.

Averaged Block Group Microdata		
Field Number	***Name***	***Description***
1	GEOID10	US Census Block Group ID
2	ALAND10	Land area of the block group (m^2)
3	AWATER10	Water area of the block
4	TOXCONC	Average toxicity-weighted concentration of the cells in the block group, averaged over three years.
5	PTOXCONC	Percentile associated with field TOXCONC.
6	SCORE	Average risk-related score (surrogate dose * toxicity weight * population) of the cells in the block group, averaged over three years.
7	PSCORE	Percentile associated with field SCORE.
8	NCSCORE	Average risk-related score (surrogate dose * toxicity weight * population) of the cells in the block group, averaged over three years. Score is calculated using only noncancer toxicity weights.
9	PNCSCORE	Percentile associated with field NCSCORE.
10		Average risk-related score (surrogate dose * toxicity weight * population) of the cells in the block group, averaged over three years. Score is calculated using only cancer toxicity weights.
11	PCSCORE	Percentile associated with field CSCORE.
12	POP	Average population of the cells in the block group, averaged over three years.
13	PPOP	Percentile associated with field POP.
14	COVERED	Internal field.
15	FOUND	
16	GC	Grid code.

Water Microdata

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_XXXX, where XXXX is the reporting year of the data freeze.

Water Microdata

Field Number	Name	Description
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

Other Available Data

Census Crosswalks

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.

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_PC_B	Percent of the cell's population that is within the Census block (Population-Cell to Block)

Population Data (US Decennial Census)

RSEI Census data are contained in three tables, Census 90 (data from the 1990 Census), Census 00 (data from the 2000 Census) and Census 10 (data from the 2010 Census). These three tables contain the Census data that has been transposed onto the RSEI model grid. Each Census table is over 600 MB in size. 1990 Census data have been provided by Geolytics, Inc.

Census data were last updated in 2012.

Census 90 Data	
Variable	***Description***
Grid Code	Number that identifies the model grid within which the cell is located.
X	Assigned grid cell value based on latitude.
Y	Assigned grid cell value based on longitude.
Male0to9 through Female65andUp	The number of people in the grid cell in each Census subpopulation group in the year 1990.
PrimaryFIPS	The FIPS code for the county within which most or all of the grid cell is contained.

Census 00 Data	
Variable	***Description***
Grid Code	Number that identifies the model grid within which the cell is located.
X	Assigned grid cell value based on latitude.
Y	Assigned grid cell value based on longitude.
Male0to9 through Female65andUp	The number of people in the grid cell in each Census subpopulation group in the year 2000.
PrimaryFIPS	The FIPS code for the county within which most or all of the grid cell is contained.

Census 10 Data	
Variable	***Description***
Grid Code	Number that identifies the model grid within which the cell is located.
X	Assigned grid cell value based on latitude.
Y	Assigned grid cell value based on longitude.
Male0to9 through Female65andUp	The number of people in the grid cell in each Census subpopulation group in the year 2010.
PrimaryFIPS	The FIPS code for the county within which most or all of the grid cell is contained.

Shapefiles- Current Version (Grid geography)

RSEI shapefiles define the grid and can be used for mapping. They do not contain any RSEI results. New shapefiles were posted on the RSEI ftp site in early 2017. The shapes are the same; however, the fields and format are different, and now additional files for grid cell sizes other than 810m are available. More information on the RSEI grid can be found in the RSEI methodology document.

Attribute Table for Grid Shapefiles	
Variable	***Description***
CELLX	Assigned grid cell value based on latitude.
Y	Assigned grid cell value based on longitude.
LAT	Latitude for center point of grid cell.
LONG	Longitude for center
CX	Vertical distance from the grid center point to grid cell (m). Equivalent to CELLX*grid size (m) (for standard RSEI grid, CELLX*810).
CY	Horizontal distance from the grid center point to grid cell (m). Equivalent to CELLY*grid size (m) (for standard RSEI grid, CELLY*810).

Shapefiles- Older Version (Grid geography)

RSEI shapefiles define the grid and can be used for mapping. They do not contain any RSEI results. There are two sets: polygon (con_us_810m_poly) and center point (con_us_810m). The grid is split into 4 files for each type, numbered 1-4. The attribute table is the same for all shapefiles. More information on the RSEI grid can be found in the RSEI methodology document.

Attribute Table for Grid Shapefiles	
Variable	***Description***
X	Assigned grid cell value based on latitude.
Y	Assigned grid cell value based on longitude.
LONGX	Easting coordinate for Albers projection.
LATY	Northing coordinate for Albers projection.
LONGITUDE	Longitude for center point of grid cell.
LATITUDE	Latitude for center point of grid cell.
RADIALDIST	Radial distance from center point of grid (m).
AREA	Area of grid (m) (note that grid cells vary slightly in size).
NORTHADJ	Internal.

[revised 3/27/2017]