Hazardous waste handler universe

TSDF = Treatment, storage and disposal facilities

LQG = large quantity generator

SQG = small quantity generator

CESQG = conditionally exempt small quantity generators

Transporters

**FACILITIES**

* ID\_NUMBER uniquely identifies RCRA sites/“handlers” (number of unique values = number of observations in dataset)
  + 4-12 characters, starts with valid state code
* FACILITY\_NAME: legal name of handler
* ACTIVITY\_LOCATION: state postal code for location of *agency regulating the handler*
* FULL\_ENFORCEMENT 🡪 includes all TSDFs that *could* potentially undergo a compliance evaluation inspection (CEI) to be evaluated for compliance w/applicable TSDF requirements
  + Six sub-universes:
    - L = Land Disposal
    - I = Incinerator
    - B = boiler/industrial furnace
    - S = storage
    - T = treatment
    - H = solid waste management
  + Only a TSDF will have this code
* Location: STREET\_ADDRESS, CITY\_NAME, STATE\_CODE, ZIP\_CODE, LATITUDE83, LONGITUDE83
* FED\_WASTE\_GENERATOR: regulatory status of the site w/r/t federal rules to which it is subject as determined by the quantity and/or toxicity of hazardous wastes generated/stored/accumulated over a specified period
  + 1 = Large quantity generator
  + 2 = Small quantity generator
  + 3 = Conditionally Exempt Small Quantity generator
  + N = not a generator
  + Null/blank = unverified
* TRANSPORTER: handler is engaged in transportation of hazardous waste (Y or N)
* ACTIVE\_SITE: indicator based on the information currently in the RCRAInfo database (not legally significant or enforceable)
  + H = Handler Activities (activities found on the federal SI form except TSDF)
  + P = Permitting Activities (site has at least one “active” unit according to RCRAInfo Lega/Operating Status)
  + A = Corrective Action Activities (in the Corrective Action Workload universe)
  + C = Converter (site has at least 1 unit that is a “converter” according to RCRAInfo Lega/Operating Status)
  + S = State-specific activities (site has state-defined activities in the RCRAInfo Permitting or Handler)
* OPERATING\_TSDF: captures every facility that currently has an operating TSDF unit
  + LIBSTH like FULL\_ENFORCEMENT: see how this is different

**ENFORCEMENTS**

* ID\_NUMBER
* ACTIVITY\_LOCATION
* ENFORCEMENT\_IDENTIFIER: 3-digit number used to uniquely identify multiple occurrences of an enforcement action
  + Is this a count? This field is messy as hell
  + Maybe discretize into a range and use as a feature
* ENFORCEMENT\_TYPE: code that identifies the type of action being taken against a handler
  + Don’t know what these codes mean: combine/aggregate to discretize?
  + Use ENFORCEMENT\_DESC to discretize (regular expressions to pull out keywords)
* ENFORCEMENT\_AGENCY: E (EPA) or S (State)
* ENFORCEMENT\_ACTION\_DATE: the date (MM/DD/YYYY) that the enforcement action (ENFORCEMENT\_TYPE) was issued
* PMP\_AMOUNT, FMP\_AMOUNT, FSC\_AMOUNT, SCR\_AMOUNT – dollar amounts for settlements/punishments etc., mostly null

**EVALUATIONS**

* ID\_NUMBER: unique identifier
* ACTIVITY\_LOCATION – state
* EVALUATION\_IDENTIFIER – three-digit number used to uniquely identify an evaluation action
* EVALUATION\_TYPE – three-character code used to report the type of evaluation of handler records or conducted at the handler site
* EVALUATION\_DESC – textual description of EVALUATION\_TYPE
* EVALUATION\_AGENCY – one character code indicating the agency responsible for conducting the evaluation
  + B: State/territory
  + C: EPA contractor or grantee
  + E: EPA
  + L: Local
  + N: Native American
  + S: State/Territory
  + T: State/territory-initiated oversight/training actions
  + X: EPA-initiated oversight/training actions
* EVALUATION\_START\_DATE: MM/DD/YYYY (first date regardless of duration)
* FOUND\_VIOLATION: Y (Yes), U (Undetermined), N (No)

**VIOLATIONS**

* ID\_NUMBER – corresponds to unique facility
* ACTIVITY\_LOCATION – postal code
* VIOLATION\_TYPE – code indicating specific area of the RCRA portions of the Federal Code of Regulations covered by the evaluation and/or found to be in violation with RCRA regulations or statutes
* VIOLATION\_TYPE\_DESC – text description of VIOLATION\_TYPE code: use these to discretize
* VIOL\_DETERMINED\_BY\_AGENCY – E (EPA) or S (State)
* DATE\_VIOLATION\_DETERMINED – The date MM/DD/YYYY that a determination that the violation exists – not necessarily the evaluation date (need to find another way to link violations to evaluations)
* ACTUAL\_RTC\_DATE – the date the agency determines the handler demonstrated full physical compliance
* SCHEDULED\_COMPLIANCE\_DATE – the date by which the handler is to submit to the agency its documentation the violation has been brought into compliance

**NAICS**

* ID\_NUMBER – facility ID number (assigned to RCRA site)
* ACTIVITY\_LOCATION – state code
* NAICS\_CODE – industrial activity of the facility as defined by a standard industrial coding system – see the documentation
* Use facility activity as a feature once you link

**VIOLATION/SNC HISTORY**

* ID\_NUMBER – unique RCRA
* ACTIVITY\_LOCATION – state
* YRMONTH – six-digit number representing the year and month for which significant noncomplier (SNC) and noncompliance status values are captured (YYYYMM)
* VIO\_FLAG – does handler have unresolved violations, as of YRMONTH? (Y/N)
* SNC\_FLAG – is handler designated an unresolved significant noncomplier, as of YRMONTH? (Y/N)

FEATURES:

FED\_WASTE\_GENERATOR, TRANSPORTER, ACTIVE\_SITE, discretize OPERATING\_TSDF, previous ENFORCEMENTS (number, discretized TYPE/DESC), previous VIOLATIONS, industrial activity (NAICS\_CODE), VIO\_FLAG/SNC\_FLAG (these flags might prioritize noncompliers?)