DEQ CAO Emissions Inventories Processing Methodology

We began with 364 DEQ CAO Emissions Inventories received by information request to the agency. These were the first accurate, complete emissions reporting for Oregon industry large enough to require a Title V, Standard, or Simple Air Contaminant Discharge Permit for DEQ. We wanted this Emissions Inventory data available online in an easier to read format. The original format were large, complicated Excel spreadsheets Also we wanted totals for total emissions by company as well as separate totals for controlled and uncontrolled emissions.

Nick Appelmans PhD processed all the Emissions Inventories Excel files and extracted data associated with what he called the aggregate data for the two sheets Material Balance and Emission Units. He also pulled out all of the CAS (chemical specific) data but I am close to getting that. This computer programming produced three files:

2016_deq_cas_data_descriptors.csv
2016_deq_co_details.csv
2016_deq_emission.csv

Nick chose to build this data set this way because there were numerous permutations of columns and headings that just wouldn't fit in one format. DEQ even felt it was necessary to include html code in one example (which I removed). Nick built three "tables" like a database with primary and foreign keys that let you cross reference.

Nick's procedures:
The first file (2016_co_details.tsv) is a summary of the company information and in addition, has two columns that indicate whether any data was entered for either the Material Balance sheet (with an entry like MAT_data-present_false) or the Emission Units sheet (with an entry like EMI_data-present_true). The column descriptions are Company Source Number, Excel file name, Company Name, Street Address, City, Zip, EMI data present, Mat data present).

The next file (2016_units.tsv) has a listing of all the permutations of descriptions and units (essentially the relevant column headers) for the columns up to but not including the CAS chemical specific portion. Rather than try to shoehorn the various versions of these headers, I put them in this table and gave them a reference number that rows in the final file (below) refer to. There is also a number in the last row that refers to the last "descriptor" column after which columns are what I consider aggregate data. I call this aggregate because the chemical specific data in the remaining columns of these sheets (CAS data) would presumably add up to what is in these aggregate columns. So this last descriptor column tells you where to break up the header columns into descriptors to the left (and including the column entered as an integer in this table) and aggregate data to the right. The headers for this file are a, units_key (to cross reference with the data rows file below), then each column of headers followed by the number representing the last descriptor column.

Finally, the data is found in 2016_aggregate_data.tsv. This table contains every row in the two sheets that contained data. To cross reference with the other two files (for example, to obtain the description headings and units for a particular row of data), each row has a unique row id (this will be needed when I finally pull out the CAS data), the Company Source Number, the sheet (EMI for Emission Units, MAT for Material Balance), the row number for that row in the sheet, the units key, all of the data descriptors columns and the aggregate data columns up to but not including the CAS data columns.

Brad's procedures:

Brad Baker, a web designer and computer programmer processed Nick's files into web pages for each industry. DEQ informed how to write an algorithm to determine whether a chemical emitted in a process that was controlled or uncontrolled. Below is the final procedure Brad used to determine that a control device is used in the industrial process for each chemical:

_____

1. If column "Control Device Type" is zero or nothing

and then also
2. "Overall Control Efficiency (% decimal)" is 0 or nothing

Then no control device is used for that industrial process and the entire row is removed.

_____

Note: "Control Device Type" must be the first sort. If done backward this will lead to errors. Some processes that use a control device will report an "Overall Control Efficiency" of 0 yet will list a control device under "Control Device Type." In these unusual instances the efficiency is not actually zero, it is likely in 90% or better - some industries was not required to list actual efficiency in the "Overall Control Efficiency" column so they reported zero or nothing instead.

For more information on our work with DEQ CAO Emissions Inventories see page one and two of PCA Guide to Industrial Air Pollution at:

http://www.portlandcleanair.org/files/PCA-Guide-to-Industrial-Air-Pollution.pdf