

EPA Data Extractor

A utility for parsing emissions data files from the EPA Acid Rain Program

Version 1.2

Getting Started

Checklist

Windows 98, NT, or 2000

Microsoft Access 2000 (should work with Access 97, has not been tested)

Suitable Access database for storing emissions data (CO2Sources.mdb)

Database can be called anything, but must include the table structure of CO2Sources.mdb.

EPA emissions data files (files may be downloaded from:

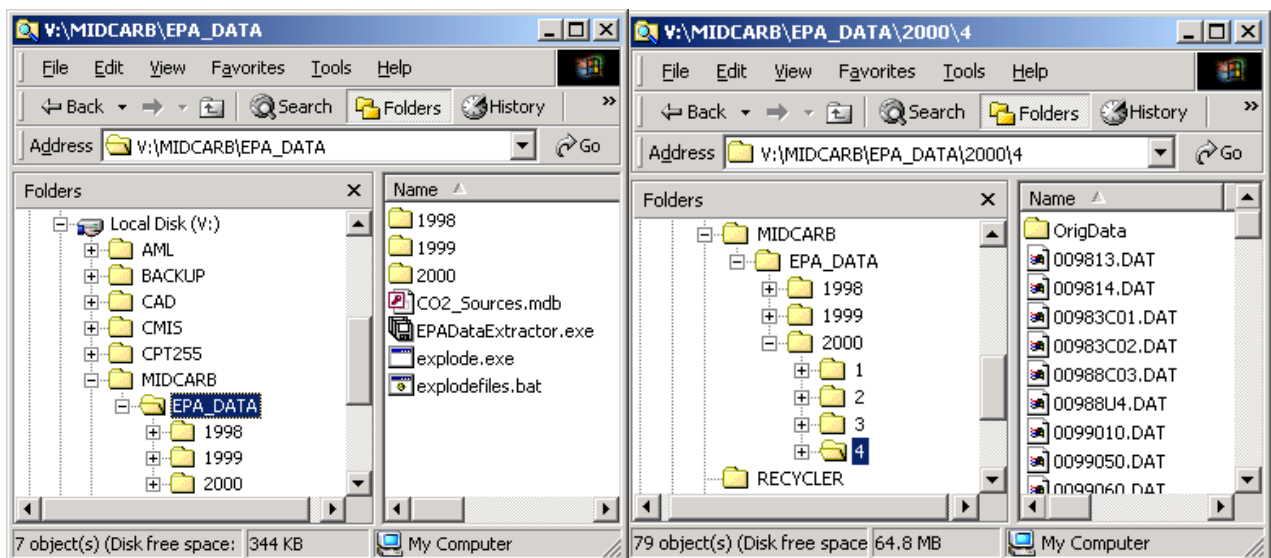
<http://www.epa.gov/airmarkets/emissions/raw/index.html>. Files must be uncompressed using the explode.exe file provided at the above Web site. For detailed instructions for obtaining EPA emissions data refer to: <http://www.midcarb.org/DataInfo.shtml>.

EPA Data Extractor application (EPADDataExtractor.exe)

Installation

To install the EPA Data Extractor utility, copy *EPADDataExtractor.exe* and *CO2Sources.mdb* to a working directory on your local drive. Figure 1 shows a suggested directory structure for managing EPA emissions data.

Figure 1. Directory structure for managing EPA emissions data and related files.



Using the EPA Data Extractor

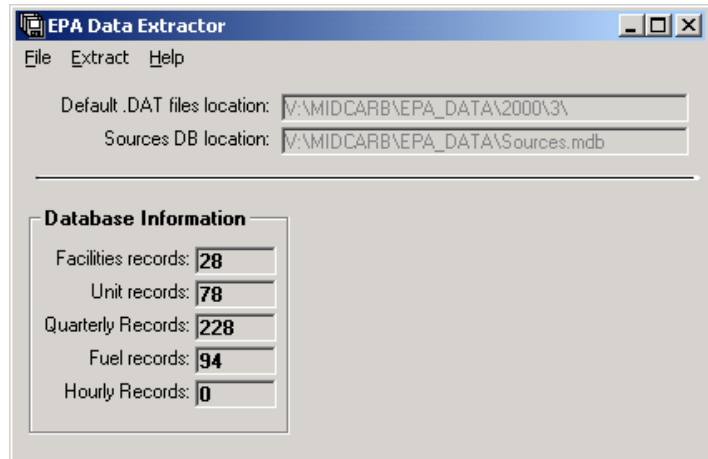
Connecting to the CO2Sources database

Run the **EPADataExtractor.exe** program (by double-clicking the file from your file manager)

Select the **File | Open Sources DB** menu and navigate to CO2Sources.mdb.

Once you have selected the sources database, the **Database Information** on the EPA Data Extractor window will show the number of records in the database. Figure 2 shows updated database information after opening the sources database.

Figure 2. The Database Information is updated after the sources database is opened.



Setting the path to .DAT files

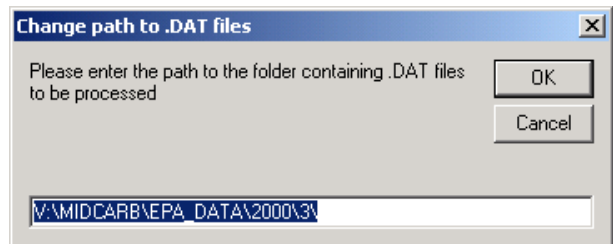
IMPORTANT NOTE:

If you plan to extract emissions data from .DAT files for quarters prior to quarter 2, 2000, you must start by extracting data from .DAT files from the most recent quarter and work backwards. This will alleviate problems associated with the older .DAT file format used prior to quarter 2, 2000.

Select the **File | Change path to .DAT files** menu and enter the full path to the folder containing .DAT files to be parsed. Figure 3 shows the window used to select the path to the .DAT files from which data is to be extracted.

Note: You must enter the full path to the folder, including a trailing “\”.

Figure 3. Dialog box used to set the path to the .DAT files to be extracted.



Using the EPA Data Extractor

Extracting Hourly and Quarterly Records

The EPA Data Extractor was designed to extract both hourly and quarterly data from the EPA .DAT files. Hourly data includes hourly emissions records for NO_x, SO_x, CO₂ mass, CO₂ concentration, effluent volume and several other parameters. Quarterly records include information for facilities, units, fuels, and emissions.

To extract **Quarterly Records** from all the .DAT files in a directory:

- 1) verify that the **Default .DAT files location** points to the .DAT files that you wish to parse (if the path is incorrect use the **File | Change Path to .DAT Files** menu to set the path)
- 2) select the **Extract | Extract Quarterly Records** menu.

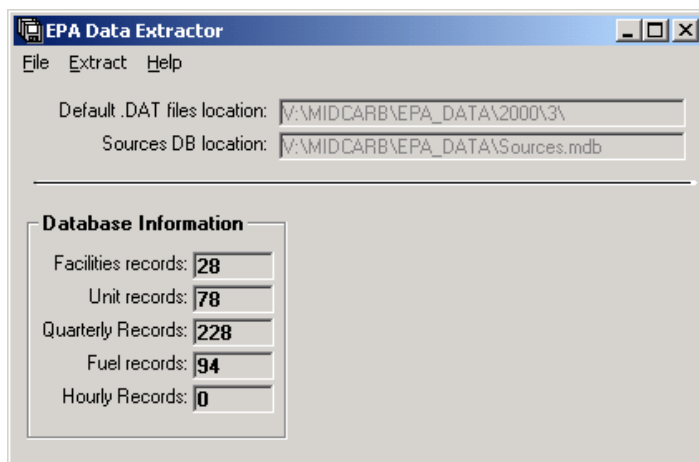
Progress indicators will show the status of data extraction. The program will execute until records have been extracted from all .DAT files in the specified directory. This process should take less than one minute if you are processing one quarter's .DAT files, with the .DAT files and the sources database on a local machine. Processing files over a network may increase the processing time several fold.

To extract **Hourly Records** from all the .DAT files in a directory, follow **Step One** above, and select the **Extract | Extract Hourly Records** menu.

To continue processing data for other calendar quarters, modify the path to the .DAT files and select the appropriate **Extract** menu. Each extraction will append records to the CO2Sources.mdb Access database.

Note: Each .DAT file contains over 2000 hourly records. It takes more than two minutes to process each .DAT file and insert its records into the Access database. If you are attempting to process all the .DAT files for a quarter, this process could easily last more than an hour. During processing, EPA Data Extractor utilizes most of the available resources on your computer. Thus, you should plan to extract hourly records only when you don't need to use your computer for other tasks.

Figure 4. Main EPA Data Extractor window showing the number of records contained in each table of the CO2Sources.mdb database.



Once record extraction has been completed, the database information is updated in the Database Information frame on the EPA Data Extractor window (Figure 4).

Using the EPA Data Extractor

CO2Sources.mdb Access Database

The CO2Sources.mdb Access database is provided with the EPA Data Extractor and acts as a simple, yet effective mechanism for temporarily storing and managing data extracted from the EPA emissions .DAT files. The table structure for CO2Sources.mdb was developed by the Kansas Geological Survey and includes placeholders for the most useful hourly and quarterly air quality and quantity statistics contained in the EPA .DAT files. The following list describes all of the tables in CO2Sources.mdb.

Emissions table – contains hourly emissions data including SO₂, NO₂, and CO₂ concentrations, SO₂ and CO₂ mass emissions, and effluent volumetric flow. This table contains only hourly records so if you don't extract hourly records this table will be empty.

Facilities table – contains one record for each facility including facility name, EPA ID numbers and Lat-Lon coordinates. This table will be populated with records from the first .DAT file that you process. In order for this table to be correctly populated you must process .DAT files submitted for **calendar quarter 2, 2000 and later** before processing older .DAT files. EPA .DAT files submitted before calendar quarter 2, 2000 use an older format and contain much less detailed information about facilities.

Quarters table – contains quarterly summaries of data included in the Emissions table as well as heat input and hours of operation. The EPA Data Extractor was designed to extract/calculate minimum, maximum, and mean values for volumetric effluent flow and CO₂ concentration from hourly emissions records and include them in the quarters table along with the standard EPA quarterly records. This table should include one record for each .DAT file processed per calendar quarter. The reporting date is typically the month following the end of the quarter so you will see records for January 2001, for year 2000, quarter 4 data and this is correct.

Units table – contains generating-unit specific data including unit type, date of first operation, stack height, stack area, and load range values. This table should include one record for each generating unit operating in your state. After the first calendar quarter of .DAT files have been processed this table will only increase as new generating units become active.

Fuels table – contains fuel type, fuel start date, fuel end date, and startup fuel. New records will be added only when a facility reports a change in fuel type.

About EPA Data Extractor

The EPA Data Extractor is a simple utility that allows users to extract hourly emissions data from raw data files obtainable from the EPA Acid Rain Program. This application was created by Nathan K. Eaton, Indiana Geological Survey, with technical assistance from Kurt Look and Melissa Moore, Kansas Geological Survey, and testing by Brandon Nuttall, Kentucky Geological Survey and Joseph Wells, Ohio Geological Survey.

This utility was developed to provide a means of easily extracting data from the EPA's raw data format for use in databases created for the Mid-Continent Interactive Digital Carbon Atlas and Relational Database (MIDCARB) project.

Provide comments and suggestions about the EPA Data Extractor to:

Nathan K. Eaton
Indiana Geological Survey
611 N. Walnut Grove
Bloomington, IN 47405
neaton@indiana.edu