

Managing EPA's Emission Inventory Databases

Douglas A. Solomon

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
MD-14
Research Triangle Park, NC 27711

Anne A. Pope

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
MD-14
Research Triangle Park, NC 27711

Rebecca L. Tooly

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
MD-14
Research Triangle Park, NC 27711

ABSTRACT

This paper describes EPA's two large databases for storing national emission inventory information. The National Emission Trends (NET) database and the National Toxics Inventory (NTI) database store information for emissions of criteria and toxic pollutants, respectively. The paper will cover design and structure of these databases, their implementation in Oracle, and the plans to merge them into one large air emission database. The paper will also discuss how data from State and local air agencies gets into these databases and EPA's use of the Web and other tools to makes these databases available to the public.

INTRODUCTION

The Clean Air Act requires that the U.S. Environmental Protection Agency (EPA) publish a list of pollutants that have adverse effects on public health and welfare, and are emitted from numerous and diverse stationary and mobile sources. For each pollutant, EPA must compile and publish a "criteria" document containing studies documenting adverse effects of specific pollutants at various concentrations in the ambient air. For each pollutant, National Ambient Air Quality Standards (NAAQS) are set at levels that, based on the criteria, protect the public health and the public welfare from any known or anticipated adverse effects. These regulated pollutants are called "criteria pollutants." To support many of its activities related to the development and implementation of the NAAQS, EPA annually estimates emissions of criteria pollutants. These emission estimates of criteria pollutants are stored in the National Emission Trends (NET) database.¹

While EPA has routinely collected emissions inventory data for criteria pollutants for about 20 years. The 1990 Amendments to the CAA provided a new focus on hazardous air pollutants (HAPs) that resulted in a need for HAP emissions inventories. The CAA presents a list of 189 HAPs (one has been delisted to bring the current number of HAPs to 188), which are also known as air toxics. The EPA is required to identify the sources of these HAPs, quantify their emissions by source category, develop regulations for each source category, and assess the public health and environmental impacts after the regulations are implemented. As EPA began to implement the requirements of the 1990 CAA, a strong need became clear for a central repository of air toxic emissions inventory data from which to conduct the analyses required by the CAA. While development and maintenance of a national database for air toxic emission inventories is not explicitly required by the CAA, availability of air toxic emissions data assists numerous program offices and other stakeholders in meeting CAA program requirements. EPA's estimates of air toxics emissions are stored in the National Toxics Inventory (NTI) database.

These two emission inventory databases are developed and maintained by EPA's Office of Air Quality Planning and Standards (OAQPS). Both of these databases are large due to the number of sources, number of pollutants, and years of historical data. The balance of this paper will describe EPA's approach to managing these data, including data input, database design and structure, access to these databases, and future plans for these databases.

BACKGROUND

The National Emission Trends Inventory

The NET, developed annually by EPA, is a national inventory of stationary and mobile sources that emit criteria pollutants and their precursors. The NET includes emission estimates for all 50 States. The specific pollutants included in the NET are VOCs, NO_x, CO, SO₂, PM-10, PM-25, and NH₃. The NET database currently contains emissions for the years 1985 through 1999. Documentation on the data contents and development methodologies for the NET can be found on EPA's Emission Trends site.

(www.epa.gov/ttn/chief/trends.html) EPA's Air Quality site presents detailed information on criteria pollutants, their effects, and EPA programs to reduce emissions.

(www.epa.gov/oar/oaqps/cleanair.html)

EPA compiles the NET using various sources of data. Table 1 lists the main sources of data for each major source category in the NET. For the 1996 NET (the most recent year compiled with State data), over 40 States submitted data for incorporation in the NET. The NET contains emission estimates for approximately 55,000 point sources, over 100 stationary area sources, over 100 different types of non-road engines, and approximately 100 different vehicle type/road functional class combinations for on-road mobile sources. Point source data in the NET is collected and stored at the emission process level. Emissions are collected and stored at the county level for all other source categories in the NET. The latest version of the NET was released in March 2001.

Table 1. NET Data Sources

Major Source Category	NET Data Source
Electric Generating Units	EPA's Emission Tracking System/Continuous Emissions Monitoring Data (ETS/CEM) and DOE Fuel Use Data
Other Large Stationary Sources	State Data, Older Inventories Grown Where No State Data Submitted
On-Road Mobile Sources	FHWA's Estimate of Vehicle Miles Traveled and Emission Factors from EPA's MOBILE Model
Non-Road Mobile Sources	EPA's NONROAD Model
Stationary Area Sources	State Data, EPA Developed Estimates for Some Sources, Older Inventories Grown Where No State or EPA Data Submitted

The National Toxics Inventory

Developed every third year (1993, 1996, 1999, etc.) by EPA, the NTI is a national inventory of stationary and mobile sources that emit HAPs in the 50 states, District of Columbia, and territories. HAPs are generally defined as those pollutants that are known or suspected to cause serious health problems. Section 112(b) of the Clean Air Act currently identifies a list of 188 pollutants as HAPs.

(www.epa.gov/ttn/uatw/pollsour.html) EPA's Unified Air Toxics Web site (UATW) presents more information on HAPs, their effects, and EPA's programs to reduce HAPs. (www.epa.gov/ttn/uatw/basicfac.html)

The 1996 NTI is the most recent for which data development is complete. It includes estimates of emissions from stationary point and non-point and mobile source categories. Point source categories include major and area sources as defined in section 112 of the CAA. Non-point source categories include area sources and other stationary sources that may be more appropriately addressed by other programs rather than through regulations developed under certain air toxics provisions (sections 112 or 129) in the CAA. Mobile sources include on-road and non-road categories. The 1996 NTI contains approximately 58,000 point sources and 500 non-point stationary source categories. The 1996 NTI includes data on the emissions of the 188 HAPs from the 50 states, District of Columbia, Puerto Rico and Virgin Islands. Point source data are available at the individual stack level within a facility. Non-point stationary and mobile source data are reported at the county level. The EPA publicly released the final 1996 NTI in August 2000.²

The EPA compiled the 1996 NTI using various sources of data. The five primary sources of 1996 NTI data were: (1) state and local HAP inventories developed by state and local air pollution control agencies, (2) existing databases related to EPA's Maximum Achievable Control Technology programs to reduce HAP emissions (www.epa.gov/ttn/uatw/epaprogs.html), (3) Toxic Release Inventory data (www.epa.gov/tri/), (4) emissions estimated by using mobile source methodology

developed by experts in EPA's Office of Transportation and Air Quality, and (5) stationary non-point source emission estimates generated using emission factors and activity data. Extensive documentation is available for the NTI.

(<http://www.epa.gov/ttn/chief/nti/index.html>) Forty-seven states, Puerto Rico, Virgin Islands and the District of Columbia participated in the development of the 1996 NTI.

CURRENT STATUS

Both the NET and the NTI recently finished a development cycle that focused on collecting information on emissions generated in the calendar year 1996. The current focus for both databases is collecting and developing data for the calendar year 1999 and merging the two databases into one umbrella database, the National Emission Inventory (NEI). The balance of this section describes the current status of the databases in terms of how data will flow into the databases for 1999, how the physical databases are managed, and how the databases can be accessed.

Data Input

During the 1996 inventory development cycle, EPA received data from State and local air agencies through a variety of ad hoc methods (e.g., FTP, E-mail attachments, floppy diskettes, etc.). For the 1999 inventory cycle, OAQPS has teamed with the EPA Office of Environmental Information to provide for submission of emissions data through EPA's Central Data Exchange (CDX). CDX is the central point of entry supporting EPA reporting systems. It provides new and existing functions for exchanging data in diverse formats including consolidated/integrated data.³

For the NEI, CDX provides the following functionality:

- receives and virus scans submitted air emissions data in two formats, flat file (in the NEI Input Format (NIF)) and extensible markup language (XML);
- logs the data submittal transaction;
- provides automated acknowledgement of receipt of the data;
- provides security by authenticating the data and preserving the contents of the original transmission;
- archives the inbound data;
- converts XML files into the NIF;
- detects errors in XML files; and
- notifies OAQPS that files have been received in the CDX Facility.

Data submitters can use CDX to submit criteria and air toxic data separately or as a consolidated criteria/air toxics inventory submittal. Figure 1 illustrates one of the CDX input screens data submitters will use to submit their inventory data.

Figure 1. CDX Input Screen

Home - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites History

Address <https://epacdx1.ilo.org/epacdx/> Go Links

EPA United States
Environmental Protection Agency

CDX

1999 National Emission Inventory Submittal Form

I. Contact Information (Please provide information on the person submitting data.)

Name:

Organization:

Telephone No: Fax No:

E-Mail:

B. Submittal Status

Note: All data sent to EPA should be considered publicly available. All data received by EPA will be processed as publically available.

Date of Submission: Submission Number (numeric):

III. Geographic Coverage (Please provide information on the geographic coverage of the submittal and check all that apply to the submittal.)

For State: Statewide: Non-attainment Area(s): Other:

For Local:

IV. Pollutant Information (Please check all pollutants included in submittal.)

VOC: NO_x: CO: SO₂: PM₁₀: PM_{2.5}: NH₃: HAPs:

V. Source Category Information (Please check all source categories included in the submittal.)

a. Criteria

Point: Size Cutoff: Stationary Area: Nonroad Mobile: Biogenic:

Onroad Mobile: Emissions VMT: MOBILE Input Files:

b. Toxics

Done Internet

During the 1996 inventory development cycle, data went through a quality assurance (QA) process after it was received from a data submitter. While the QA was rigorous, it was not standardized across criteria and air toxics data. In addition many of the QA

checks and reports were generated using ad hoc queries and reporting tools. These factors resulted in a QA process that was not very efficient or duplicable.

For the 1999 reporting cycle, one of EPA's goals is to develop a more efficient QA process. To accomplish more robust and efficient data processing, EPA has developed and published a set of standard QA checks that it will perform on both criteria and air toxics data received during the 1999 reporting cycle. A listing of the standard QA checks are documented in EPA's 1999 National Emission Inventory preparation plan, which can be downloaded from the National Emission Inventory Data site.⁴

(www.epa.gov/ttn/chief/net/index.html)

In addition, EPA has developed and made available an automated QA tool. The automated QA tool will check a set of emissions data for format and content. The tool is a stand-alone application written in MS Access. It accepts data in the NIF, checks for format errors against the NIF, checks the data content against the published QA checks and generates reports based on its findings. The automated QA tool was completed and published in April 2001.

Database Management

Both the NET and NTI are Oracle databases. Due to their different histories, however, their path to becoming an Oracle database was very different. The NET database started out in the early 1990s as a stand alone MS FoxPro database. For the NTI, the 1996 reporting cycle was the first effort to put the air toxics data into a comprehensive database. Therefore, the 1996 NTI was developed as an Oracle database.

Conversion of the NET data from flat files in MS FoxPro into a relational Oracle database proved to be a major effort. Normalizing the NET data resulted in some tables growing to more than one million records for each of the fifteen calendar years in the NET database. Final conversion of the NET data into the Oracle platform was completed in March 2001.

After completion of the conversion of the NET database from MS FoxPro to Oracle, the major data management task EPA is currently working on is the effort to merge the NET and NTI into a single database. Merging these two databases is a large challenge due to differences in the structure and content of these databases. EPA is taking a phased approach to the merger of these databases. The first step is a "virtual database merge." The goal of the "virtual database merge" is bring consistency to the data input, quality assurance methods, and data output tools developed for external users.

This has largely been accomplished for the 1999 reporting cycle. The NIF Version 2 provides a consistent format for reporting both criteria and air toxic emissions and, for the first time, provides a mechanism for data submissions to EPA of consolidated criteria and air toxics emissions data. The QA tool discussed above was also developed to bring consistency to the QA checks for both the criteria and air toxics data. EPA is also developing a consistent set of tables and reports for use in the review of the 1999 data.

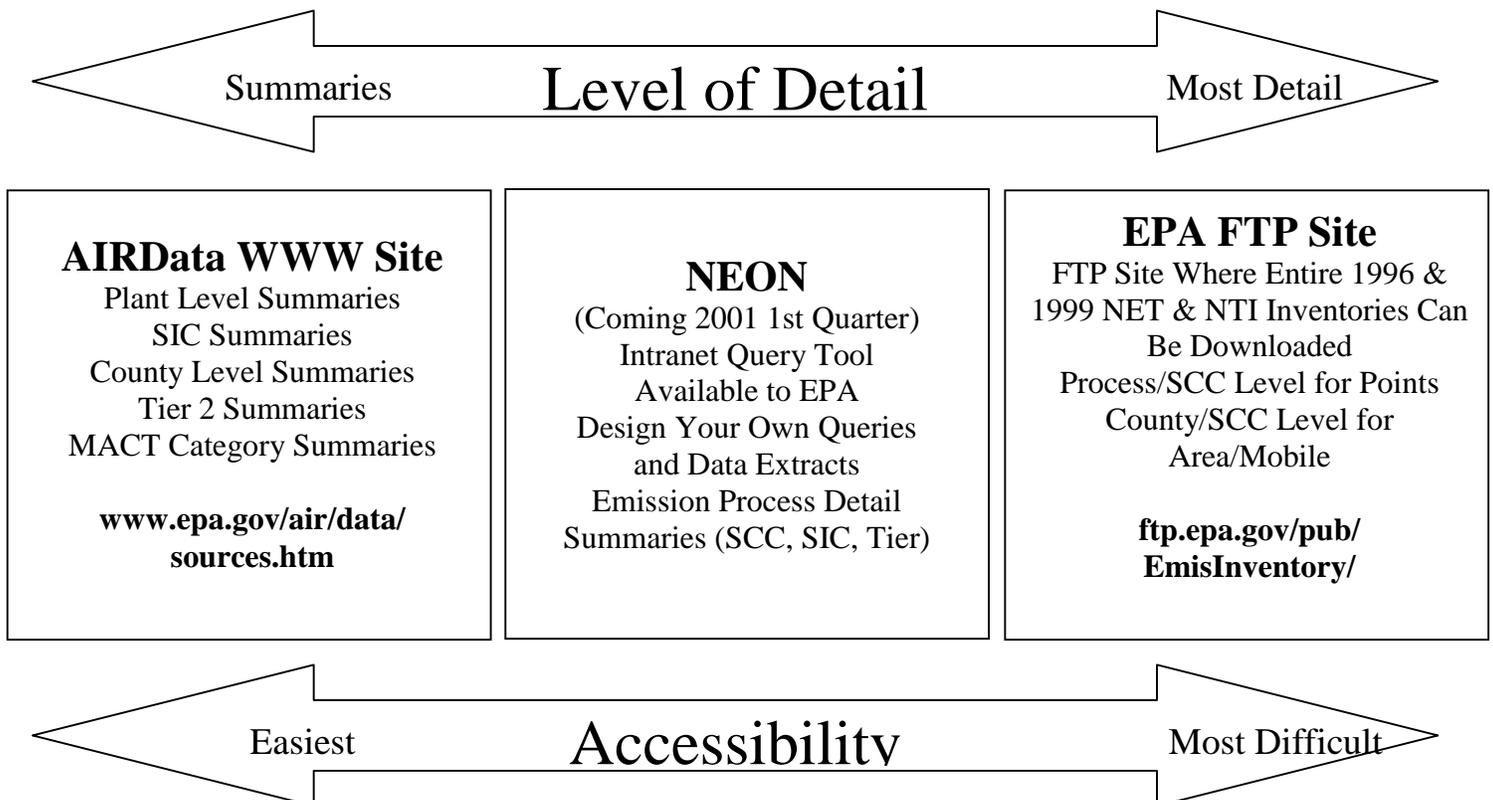
Finally, the data access options discussed below allow for users to obtain criteria and air toxics data using a single set of tools. The more difficult task of physically merging the databases is just beginning and is discussed in the Future Plans section of this paper.

The merger of these two databases is important for several reasons. First, it will allow users to get comprehensive and consistent emissions data for both criteria and air toxics from a single source. Second, it will remove the inefficiencies involved in managing two large databases for which large portions of the data overlap. Third, it will reduce the burden on stakeholders submitting and reviewing the data.

Data Access

Access to the NET and NTI data is one of the most important functions of the databases. Direct access to the databases for those outside of EPA is not currently feasible due to security issues (e.g., the databases are housed on a server behind EPA's firewall). Taking this security issue into consideration, EPA has developed a data access plan that provides 3 options for internal EPA and external access to the emissions data. Figure 2 provides an illustration of the NET/NTI data access options. Each of these data access options will be discussed below.

Figure 2. NET/NTI Data Access Options



The AIRData site (www.epa.gov/air/data) provides public access to summary data from both the NET and NTI. For both the NET and NTI there are four different reports available. Table 2 describes the reports available on the AIRData site. For each report, the user defines the data summary they would like to see by selecting the geographic location, source categories, and level of aggregation. Figure 3 is an example of an AIRData selection screen. Figure 4 is an example of an AIRData output screen. Users can export and download the results from their AIRData query. The summarized data presented on the AIRData site is detailed enough to fulfill the data needs of the majority of external users

Table 2. List of AIRData NET and NTI Reports

Report Name	NET/NTI	Report Description
Emissions	NET & NTI	Facility Level Emissions for Point Sources in Selected Geographic Area
Count	NET & NTI	Number of Point Sources and Emissions in Selected Geographic Area
SIC	NET	Number of Point Sources and Emissions in Each SIC in Selected Geographic Area
Tier	NET	Point, Area, and Mobile Emissions by EPA Tier Category in Selected Geographic Area
MACT	NTI	Facility Level Emissions With MACT Code for Selected Geographic Area
Summary	NTI	County Level Emissions Broken into Major Source, Area, On-road Mobile, and Non-road Mobile Components

Figure 3. AIRData Selection Screen



Figure 4. AIRData Output Screen



The second part of the NET/NTI data access plan is the NEI Emissions On the NET (NEON). NEON is a graphical user interface (GUI) that directly queries the NET Oracle database. Because NEON queries the Oracle database directly, there is much more flexibility in designing detailed queries. For example, with NEON process level data is available for point sources, where on AIRData the lowest level of detail available is facility level data. NEON was released for internal Beta testing in March 2001. The current version of NEON accesses the criteria data only. The second phase of NEON, which is currently under development, will include access to the air toxics data. NEON as it is currently implemented is only available to users inside the EPA firewall. EPA is exploring options for making NEON available to a wider audience. Figures 5 and 6 are examples of NEON selection and Output screen, respectively.

Figure 5. NEON Selection Screen

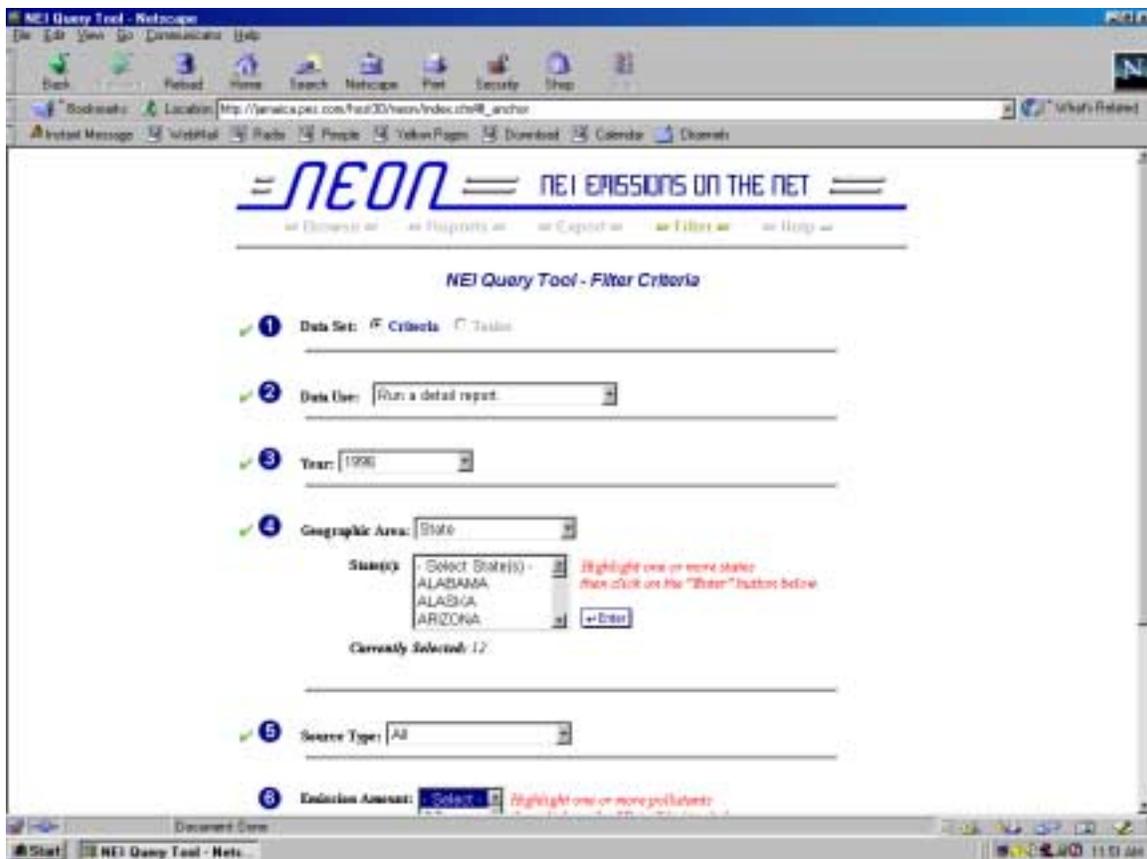
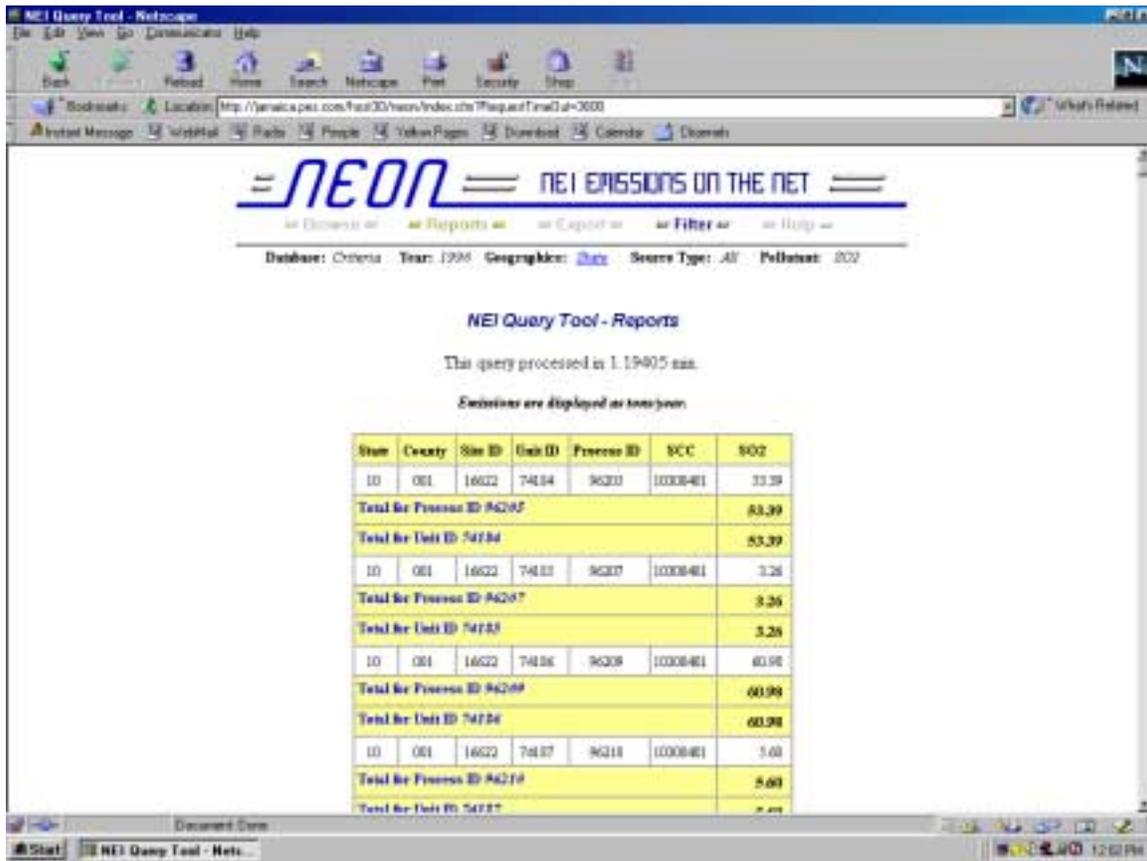


Figure 6. NEON Output Screen



The third and final part of the current NET/NTI data access plan is the provision for full file downloads via EPA's anonymous FTP site (<ftp.epa.gov/EmisInventory/>). On this site users can download the entire 1996 and 1999 NET and NTI data. These data extracts are the complete inventories in the NIF. The data files are large and are appropriate for users wishing to use these inventories in air quality modeling or other detailed analyses requiring emissions data.

FUTURE PLANS

The above section described the current status of the NET and NTI databases as of June 2001. This section describes the major changes planned for these databases over the next 12 to 18 months.

Complete Merger of NET and NTI

EPA plans to go beyond the "virtual database merge" described above and begin work on the physical database merge. The first step in this process will be to bring the NTI onto the same server that currently houses the NET. This will allow for NEON to access the NTI data. Beyond that EPA will develop bridges between the NET and NTI data so consolidated queries can be performed by the various data access tools. Finally, EPA will examine the structural database and business process differences and will develop an action plan for consolidating the databases.

Expanded Role of Central Data Exchange

EPA is exploring ways to further reduce burden on data submitters. One potential option under examination is Active Data Retrieval (ADR). ADR is an extension of CDX, where CDX with permission of the data submitter would pull data from the submitters site rather than having the submitter push the data to EPA.

Enhanced Data Access

AIRData will be enhanced to include new and enhanced reports. A NET Summary Report similar to the NTI summary report will be added to provide county level emissions broken into point, area, on-road mobile, and non-road mobile components. The ability to select data by Tribal Identifier will be added to the Emissions Reports. Drill down capability will be added to the Count and SIC reports. The ability to select multiple years will be added to the Tier Report.

New data access options will be available to the public. These include adding the NET data to Envirofacts (EPA's data warehouse that provides the public with direct access to the wealth of information contained in its databases) (www.epa.gov/enviro) and adding both the NET and NTI data to the AIRS Graphics site (www.epa.gov/agweb). Adding the NET and NTI data to the AIR Graphics site will allow users to generate a variety of user defined maps and graphs of EPA's emissions data.

CONCLUSIONS

Managing EPA's emission databases is a large effort due to the volume and complexity of the data. EPA has taken steps to put these databases on a stable, reliable, and secure platform. EPA will continue to make improvements to the databases to increase their usefulness. One of the most important improvements will be the merger of the two databases so that users will be able to get comprehensive emissions of both criteria and air toxic emissions for any given source. Other improvements such as adding Tribal Identifiers to the database will make the data useful for new groups of stakeholders.

EPA is also working on developing tools to automate tasks like electronic data submission, data quality assurance, and data access. EPA is committed to continue the development of these tools to aid both EPA data users and external stakeholders. Efforts are underway to expand the role of CDX so that the burden on data submitters continues to decrease. EPA is also placing a special emphasis on making the data available through a variety of data access options. EPA is continuing to enhance the current data access options such as AIRData and is exploring new options for data access.

REFERENCES

1. *National Air Pollutant Emission Trends, 1900-1998*, U.S. Environmental Protection Agency, Research Triangle Park, NC, 2000; EPA-454/R-00-002.
2. Pope, A.; Dombrowski, S.; Wilson, D. *The 1996 National Toxics Inventory - A Key Component in the 1996 National Air Toxics Assessment*. Presented before the Air and Waste Management Association, Orlando FL, 2001.
3. Chaudet, R; Nobles, M.; Kutscher, K. *Summary of Air Emission Inventory Central Data Exchange Pilot Test of NIF Flat File Submissions, Draft Report*, 2001.
4. *National Emission Inventory Preparation Plan*, U.S. Environmental Protection Agency, Research Triangle Park, NC, 2001.

KEY WORDS

Emissions
Inventory
Databases
Criteria pollutants
Air Toxics
Hazardous Air Pollutants (HAPS)