

2002 National Emission Inventory (NEI) Preparation Plan

- Final -

USEPA

Emission Factor and Inventory Group
Office of Air Quality Planning and Standards
August 10, 2004

Table of Contents

Chapter 1: Plan Overview	5
What is the purpose of this document?	5
What are the intended uses of the NEI?	5
What is the schedule for the NEI preparation?	5
Will we prepare more than one version of the NEI?	6
Chapter 2: General Issues	7
What will we call the different versions of the NEI?	7
How will we organize source categories?	7
What source categories will we include?	8
What pollutants will we include?	8
What geographic areas will we include?	8
What time periods will we cover?	8
Chapter 3: How We Will Collect Data	11
What data will we collect from whom, when, and in what form?	11
What does the Consolidated Emissions Reporting Rule (CERR) require?	14
What are the acceptable File Formats for S/L/T's Transferring Data to EFIG?	14
What are the File Formats for Other Data Providers Transferring Data to EFIG?	14
What are we expecting re XML submissions and how will we process them?	15
How will Data Providers transfer data to EPA? Who in EFIG will handle the data receipt?	15
How should S/L/Ts indicate in their data submittals that they believe emissions are zero?	16
How will we transfer the data to the data development contractors?	16
What are our criteria for acceptance/nonacceptance of the data?	17
Where will we store and/or copy the data for archival purposes?	17
What impact will the EPA Data Standards and Information Quality Guidelines have?	18
Chapter 4: How We Will Process Data and Compile the Inventory	19
How will we Blend/Merge the data?	19
How are MACT codes assigned in the NEI?	20
What steps will we take to meet the EPA's Information Quality Guidelines?	20
Chapter 5: How We Will Distribute Data	29
What are the User Needs for NEI data distribution?	29
What will be the procedure to document the codes used in the final NEI?	30
What will be the steps to generate summary files?	30
What CAPs Summary Files will we generate?	30
What HAPs Summary Files will we generate?	31
What other distribution files will we create?	31
What README and Documentation files will we prepare?	32
What are the final steps for distribution of the NEI files?	32
Chapter 6: How We Will Check Data Quality	33
What tools will we and our contractors use to check the quality of the data?	33
What QC steps should State/Local/Tribal agencies take prior to submittal of data to EPA?	33
What procedures will we follow prior to, or as part of, sending data submittals to our contractors?	33
What procedures will our contractors follow to check the quality of the data?	34
What QC procedures will we follow after the contractors complete their work but prior to publication of the draft NEI?	35
What additional QC procedures will we conduct during the draft comment period, if time permits?	36

What steps will we follow to process the draft inventory revisions and comments?	37
Chapter 7: How We Will Peer Review the NEI	39
Why is the NEI undergoing peer review?	39
At what milestones in the NEI preparation process will peer review be conducted?	39
What is the purpose of the NEI peer review?	40
How will the NEI peer review be conducted?	40
Chapter 8: NEI Data Management System Flow	43
What are the operational components, and sequence of data flow for the Data Collection component?	43
What are the operational components, and sequence of data flow for the Data Processing and Storage component?	44
What are the operational components, and sequence of data flow for the Data Distribution component?	46
Chapter 9: Special Issues	55
Appendix A: Reporting of HAP Groups	59
Appendix B: Source Categories in the NEI	63
Appendix C: Data Elements in the NEI	69
Appendix D: Emissions Data Needs for Air Quality Modeling	75
Appendix E: Example Chain of Custody Form for Data Submittals	83
Appendix F: Example Communications with Point Source Data Providers	85
Appendix G: Example Communications with NonPoint Source Data Providers	87

Chapter 1: Plan Overview

What is the purpose of this document?

We have prepared this document (herein referred to as “the Plan”) to serve as an internal planning tool for EPA’s Emission Factor and Inventory Group (EFIG), which manages the NEI. However, it also serves to inform our State, Local, and Tribal agency partners (S/L/Ts) of our plans for the 2002 NEI preparation and the opportunities for providing input to, and for reviewing, the NEI. The EPA NEI planners will record in this document the steps that we will use to prepare, store, and distribute the 2002 NEI. We will post it on the CHIEF web site (www.epa.gov/ttn/chief). You can contact David Misenheimer, EPA/EFIG (misenheimer.david@epa.gov) if you have any comments or questions concerning this plan.

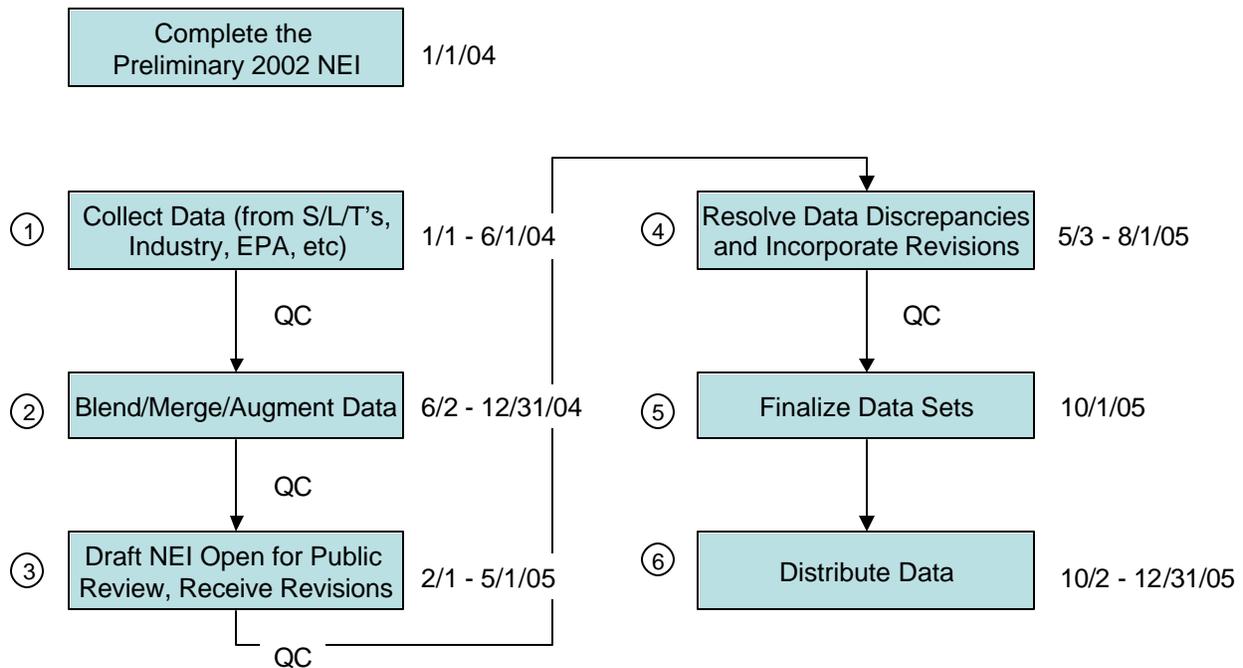
What are the intended uses of the NEI?

We are preparing the 2002 NEI is to meet several specific needs, including:

- Key input to regional/national modeling by EPA, RPO’s, S/L/T’s, etc.
- Basis for National Air Toxics Assessment (NATA) type risk analyses
- Starting point for rule development (past and ongoing examples include Nonroad Rule, Transport Rule, Clear Skies)
- Trends and GPRA tracking
- Public information (esp. important with expanded use of web)

What is the schedule for the NEI preparation?

The schedule for the main NEI preparation steps is as follows:



Each of the above steps is discussed in greater detail later in this Plan.

Will we prepare more than one version of the NEI?

We will complete a preliminary version of the 2002 NEI on January 1, 2004. In this version we will include criteria pollutant emissions for point sources, and we will include HAPs and criteria pollutant emissions for all other sources. We will prepare a subsequent version of the 2002 NEI, including criteria and HAPs for all sectors, and including data provided by S/L/Ts and industry. We will complete a draft of this subsequent version by February 1, 2005, and we will finalize it by October 1, 2005. We will distribute the final version by December 31, 2005. We will include in this final version S/L/T and industry data provided between January 1 and June 1, 2004, and revisions provided during the draft review period.

Both of the preliminary and final versions will be suitable for anyone wanting to use the best data available at the time, as both will have undergone substantial quality review. We do not recommend that anyone use the draft version except to review and identify needed corrections.

Chapter 2: General Issues

What will we call the different versions of the NEI?

We will use the following format for naming the preliminary, draft, and final versions of the 2002 NEI (e.g., when transferring files to/from the Data Development Contractors, when transferring files to other EPA groups, when posting NEI files on our web site or FTP site). [Note: We do not intend for S/L/Ts to use this format when submitting data to EPA. They should follow the naming conventions in the NIF 3.0 guidance.]

- "2002NEI", followed by
- Sector (PT, NP, OR, NR), followed by
- Pollutants (HAP, CAP), followed by
- Underscore, followed by
- Month and Year of release (2 digit-4 digit, example: 012003)

An example of this is: 2002NEIPTCAP_102004

We will insert a two letter State abbreviation at the beginning of the name for a specific state inventory, and the term "Prelim" (for preliminary), "Draft", or "Final" at the end, when appropriate. Also, we will note the file type (.zip, .txt, .mdb, etc.) at the end when naming a specific electronic file (e.g., when posting files on the ftp site).

An example of this is: CA2002NEIORHAP_032004DRAFT.ZIP

We will use the following version names for the major 2002 NEI products according to the dates in the proposed preparation schedule:

- Preliminary NEI: 2002NEIPTCAP_012004PRELIM,
2002NEINPHAPCAP_012004PRELIM,
2002NEIORHAPCAP_012004PRELIM, &
2002NEINRHAPCAP_012004PRELIM
- Draft NEI: 2002NEIHAPCAP_022005DRAFT
- Final NEI: 2002NEIHAPCAP_102005FINAL

How will we organize source categories?

The NEI is organized into 4 main groupings of categories which we call "source sectors":

- Point sources (stationary)
- NonPoint sources (stationary)
- Onroad sources (mobile)
- Nonroad sources (mobile)

The Point source sector is further divided into two main "subsectors", Electric Generating Units (EGU's) and NonEGU's. All significant sources of emissions for these sectors are included in the NEI. Biogenic source emissions, while not included in the NEI, are accounted for and generated during processing for air quality modeling. We discuss source category definitions in more detail in Appendix B.

What source categories will we include?

We will include a large number of source categories in the 2002 NEI. In the preliminary version of the 2002 NEI, we will carry forward emission estimates for many source categories from the 1999 NEI, version 3, and will calculate updated emissions for the remainder of categories. In the draft and final versions, we plan to calculate 2002 emissions for a large number of source categories using updated activity data, emission factors, and/or other methods. There are also a small number of source categories for which we plan to use the 1999 NEI emission estimates as is. In Appendix B, we list the groups of source categories that we will calculate or carry forward emissions in the 2002 NEI and note the source of the estimates in the preliminary and draft/final versions. S/L/Ts may also submit emission estimates for additional source categories. We will include these in the NEI where possible.

What pollutants will we include?

In order to make the 2002 NEI a reliable starting point for anticipated regional scale modeling and human exposure modeling, we will include emissions information for the following pollutants:

- all criteria pollutants, including precursors of ozone and PM_{2.5}, and
- the 188 HAPs, including individual HAPs reported for compound groups listed in the Clean Air Act (CAA).

More specifically, for criteria pollutants and precursors, we will include the following:

- Nitrogen Oxides
- Sulfur Oxide
- Volatile Organic Compounds
- Carbon Monoxide
- Primary PM-10
- Filterable PM-10
- Primary PM-2.5
- Filterable PM-2.5
- Ammonia

Note: Because the final 2002 NEI will include both criteria pollutants and HAPs, and because lead (Pb) is both a criteria pollutant and a HAP, we found it necessary to identify it either as a criteria pollutant or as a HAP. The list of 188 HAPs includes the compound group "lead compounds." Where possible, we maintain speciated HAP compound data. The Chemical Abstract System (CAS) pollutant identifiers available in the list of HAP compounds in the NEI assist us in maintaining speciated emissions for lead. Including lead also as a criteria pollutant would be redundant so we will track it only as a HAP.

Whether we or the S/L/Ts include pollutant emissions for specific source categories depends on the availability of source test data, emission factors or estimation methodologies, and activity data. For information on HAPs that will be included in the NEI and on reporting HAP compound groups, refer to Appendix A.

What geographic areas will we include?

We will include emissions information for all 50 states and their counties, D.C., the U.S. territories of Puerto Rico and Virgin Islands, and the territories of federally-recognized American Indian nations. Note: Although we will work on an approach to include Tribal emissions data into the NEI, currently we do not have a mechanism to include Nonpoint and Mobile data not submitted at the county/state level.

What time periods will we cover?

We will include in the NEI annual emissions for criteria and HAPs for calendar year 2002. Data preparers (S/L/Ts and EPA) will identify annual emissions for many sources in the NEI as occurring January 1 through December 31. Because emissions for a smaller number of sources are seasonal in nature (e.g., woodstoves), data preparers will identify them as occurring for only a portion of the year (e.g., November 15 through March 15).

We anticipate that S/L/T's will also submit seasonal daily emissions for incorporation into the NEI. We have encountered a significant number of problems in the past with completeness and accuracy of reported seasonal emissions values. Because of this, we will evaluate seasonal data submitted for the 2002 NEI to determine if the completeness and accuracy is at a sufficient level to warrant the incorporation into the NEI.

For O3 nonattainment areas, S/L/T's prepare O3 season daily emission values for O3 SIP inventory purposes, because they must base Rate of Progress (ROP) plans on O3 season daily emissions. They should submit these data to us (via CDX) because: 1) it is a convenient means for the EPA Regional Offices to review the draft SIP inventories, and 2) for certain sources and source categories it would not be possible for us to accurately calculate seasonal emissions from annual emissions. Also, for many agencies, it is easier to submit all of their emissions data (annual and seasonal) rather than report a subset of their data (annual only).

For the source categories and geographic areas where S/L/Ts do not report seasonal emissions, we will not calculate seasonal daily emissions in the NEI. Instead, we will calculate these seasonal daily emissions when preparing files for air dispersion modeling, either during preparation of inputs to SMOKE, or during processing of emissions in SMOKE.

Mobile sources are somewhat unique concerning time periods. We will calculate mobile source emissions on a monthly basis. For S/L/T that provide us with mobile source data, we are requesting those data on a monthly basis.

Chapter 3: How We Will Collect Data

In this chapter, we describe how we will collect data for compiling the 2002 NEI, specifically the version of the 2002 NEI which includes data submitted by S/L/Ts, as well as from other data sources. We describe:

- the data we will collect
- from whom and when we will collect it
- how and in what form(s) we will collect it
- quality control checks that we will apply during data collection
- how we will meet EPA Data Standards and Information Quality Guidelines

What data will we collect from whom, when, and in what form?

S/L/Ts will provide much of the information for the NEI. However, because data from these agencies may be incomplete and because we consider data from certain other information sources for a few source categories to be the most representative available, we will use additional sources of data to prepare the 2002 NEI. In the tables below, we discuss the source and form of the data by source sector. This will help identify differences in the various data sets we collect or receive for specific sectors, including schedule differences and data format differences.

Point Sources - <i>Electric Generating Utilities (EGUs)</i>				
<u>Pollutant(s)</u>	<u>Data Source</u>	<u>Data Source Format</u>	<u>EFIG Data Organizer</u>	<u>Delivery Date to Contractor</u>
SO ₂ , NO _x ¹	EPA Clean Air Markets Division (CAMD) 2002 Continuous Emission Monitoring (CEM) Database - final	CAMD CEM Database report	Pt source contractor (ERG) and subcontractor (Pechan)	9/30/03
Hg ²	ESD Hg Model	NIF V3.0	Pt source contractor	6/1/04
Non SO ₂ /NO _x criteria and HAPs (including Hg) ³	EIA 767 Data	767 database output	Pt source contractor	9/30/03
Point Sources - <i>non-EGUs</i>				
Criteria / HAPS	S/L/T data submissions	NIF V3.0 or XML V3.0	Pt source contractor	1/1-6/1/04
HAPs	EPA ESD MACT, Residual Risk, and Section 112k area source data	multiple formats, including NIF, residual risk modeling files (from Ted Palma of EPA/ESD), etc.	Pt source contractor	?
HAPs	Industry	multiple formats	Pt source contractor	1/1-6/1/04
HAPs and NH3	EPA TRI Database	TRI Air Emission Report	Pt source contractor	?

<u>Pollutant(s)</u>	<u>Data Source</u>	<u>Data Source Format</u>	<u>EFIG Data Organizer</u>	<u>Delivery Date to Contractor</u>
HAPs and criteria	1999 NEI for HAPs and Prelim. 2002 NEI for facilities not included in the data sources listed above	NIF V3.0	Pt source contractor	??

Notes:

1) EPA/CAMD will provide Pechan with preliminary CEM data (i.e., data partially QC'd) 6 months after close of the calendar year. They will format it for NIF, transfer it to Point Source Subcontractor, Pechan. Pechan will prepare final 2002 and projected 2003 data set during Spring 2004.

2) EPA/ESD Contact is Bill Maxwell. The Hg Model is for calendar year 1999. We will use a ratio of heat input data for 2002 and 1999 to project emissions to 2002.

3) We will use fuel usage, heat input, control device efficiencies, location, and stack parameters from DOE/EIA 767 reports, combined with AP-42 emission factors to estimate emissions for non-CEM EGUs and for estimating emissions from all EGUs for pollutants other than SO₂, NO_x and Hg. We will provide details about step in the 2002 NEI documentation for point sources.

Nonpoint Stationary Sources				
<u>Data Source</u>	<u>Pollutant(s)</u>	<u>Data Source Format</u>	<u>EFIG Data Organizer</u>	<u>Delivery Date to Contractor</u>
S/L/T data submissions	HAPs and criteria	NIF V3.0	Nonpoint contractor	6/2004
ESD MACT	HAPs and criteria	various	Nonpoint contractor	
Nonpoint contractor	HAPs and criteria	Contractors' working formats	Nonpoint contractor	

Onroad Mobile Sources				
<u>Data Source</u>	<u>Data (HAPs & criteria)</u>	<u>Data Source Format</u>	<u>EFIG Data Organizer</u>	<u>Delivery Date to Contractor</u>
OTAQ ¹	NMIM county-level inputs	NMIM input format	Mobile source contractor	
S/L/T data submissions ¹	VMT, and/or MOBILE6-generated emissions w/ input/output files, and/or NMIM county-level inputs	VMT in NIF 3.0 MOBILE6 emissions in NIF 3.0 and MOBILE6 input/output in MOBILE6 file format NMIM inputs in NMIM input format	Mobile source contractor	6/2004
FHWA ²	VMT	FHWA database	Mobile source contractor	6/2004

Notes:

1) OTAQ will provide us with an NMIM county-level input database populated with default MOBILE6 information (input files, temperatures, RVP, fuel parameters, etc.). S/L/Ts can submit replacement county-level inputs. Another option is for S/L/Ts to instead provide emissions generated with the current version of MOBILE6, but are asked to also provide the associated model input and output files. S/L/Ts can also provide VMT. OTAQ will provide a guidance document on how to revise the NMIM inputs with local data (see <http://www.epa.gov/ttn/chief/net/2002inventory.html#nmim>).

2) FHWA sent us a preliminary 2002 VMT dataset in Spring 2003 and will send a final data set Spring 2004. We will use the preliminary VMT data for the Preliminary version of the 2002 NEI and the final VMT data for the Draft and Final versions of the 2002 NEI. For States that submitted 1999 VMT, we will project 2002 VMT based on their 1999 values.

Nonroad Mobile Sources				
<u>Data Source</u>	<u>Data (HAPs & criteria)</u>	<u>Data Source Format</u>	<u>EFIG Data Organizer</u>	<u>Delivery Date to Contractor</u>
OTAQ ¹	NMIM county-level inputs	NMIM input format	Mobile source contractor	
S/L/T ¹	NMIM county-level inputs, or NONROAD model emissions w/ input/output files	NMIM input format NONROAD emissions in NIF V3.0 and NONROAD model input/output files in NONROAD file format	Mobile source contractor	6/2004
S/L/T	Emissions for Locomotives, Commercial Marine Vessels, Aircraft	NIF V3.0	Mobile source contractor	6/2004

Notes:

1) OTAQ generated preliminary 2002 NONROAD emission estimates. They will provide an NMIM county-level data base populated with default NONROAD information (input files, temperatures, RVP, fuel parameters, etc.) (see <http://www.epa.gov/ttn/chief/net/2002inventory.html#nmim>). S/L/Ts can submit replacement county-level inputs. Alternatively, S/L/Ts can provide emissions generated with the current

version of NONROAD, but are asked to also provide the associated model input and output files.

What does the Consolidated Emissions Reporting Rule (CERR) require?

The CERR requires State and Local air agencies to report criteria pollutant air emissions data to EPA each year, including 2002. Details of the final rule, as well as a summary of its major provisions, are available at <http://www.epa.gov/ttn/chief/cerr/index.html>. The rule requires State and Local air agencies to report their emission inventory data to EPA in the latest electronic format designated by the Emission Factor and Inventory Group.

We make use of CERR reports in developing the NEI, but there are also other uses for the CERR submissions. We will post the CERR submissions after we have completed our QC of the data files.

What are the acceptable File Formats for S/L/T's Transferring Data to EFIG?

We have specified the formats for S/L/Ts reporting of data as the NEI Input Format (NIF) Version 3.0 and the NEI extensible markup language (XML) Schema. Other agencies or groups (e.g., Tribes, Industry Trade Groups) submitting emissions data to us for incorporation into the NEI must also use these formats. The NIF version 3.0 is documented at <http://www.epa.gov/ttn/chief/nif/index.html#ver3>. S/L/Ts can obtain a copy of the current XML schema for NEI by contacting our staff at <http://cdx.epa.gov/contactus.asp> or at (888) 890-1995. The NIF 3.0 and NEI XML formats apply to NEI reporting of criteria and HAP data.

The NIF Version 3.0 is available now and is the appropriate NIF version for transferring 2002 emissions data to the NEI. We plan to retire the previous version, Version 2.0, prior to January 1, 2004. At the time of this writing, we are updating the XML format to reflect all of the content of NIF Version 3.0. We expect to complete the NEI XML Version 3.0 by March 31, 2004. We will include in the NIF and XML formats the same data element content, though packaged differently. S/L/Ts may use either file format to provide the data of interest to us for incorporation into the NEI and to meet the reporting requirements of the CERR.

We will accept the NIF and XML submitted files with HAPs and CAPs together or in separate files. However, we prefer that S/L/Ts include them together in the same file. This will improve our ability to QC their data (e.g., compare consistency of VOC and organic HAP, and PM and particulate HAP methodologies and emission estimates for the same categories) and will allow us to respond more quickly concerning the results of our review of their data. Note that for onroad mobile, file size might make submittal of HAPs and CAPs together impossible.

While S/L/T's and Industry must use these file formats, our contractors and other data providers such as EPA/ESD that are developing and delivering 2002 data for incorporation into the NEI will also use the NIF.

Note: The Regional Planning Organizations (RPOs) are also expecting to exchange 2002 emissions inventory data with S/L/Ts in a form that allows direct input to emissions processing for air quality modeling. The RPO data exchange protocols are arranged by source type, and for specific sector data, include some use of the NIF Version 3.0 as is or a modified version of the NIF V3.0. The RPO protocols also include many formats for exchanging detailed temporal, spatial, and speciation data that is desired and used as direct input to emissions processing for air quality modeling. While RPOs may use some of the 2002 NEI data we compile as a starting point for their efforts, we do not expect the RPOs to submit us the RPO data exchange protocols that are non-NIF for incorporation into the NEI.

What are the File Formats for Other Data Providers Transferring Data to EFIG?

The table above lists file formats for data providers, including formats other than NIF 3.0 and XML 3.0.

What are we expecting re XML submissions and how will we process them?

For the upcoming 2002 submittals, we anticipate that some States will submit XML files. These will be predominantly point source files; it is unlikely that we will receive non-point or mobile in XML format. We anticipate that 3 to 10 States will submit point files in the NEI XML format.

Presently, we are updating the NEI XML schema to reflect the NIF V3.0 requirements, and are also adjusting the XML to NIF convertor to read the new XML files and translate them into NIF V3.0 files in ASCII for purposes of our data processing, including blending with other data sources. The conversion software will be a 'stand-alone' product that we can use within EFIG if necessary, and which each EFIG WAM must provide and instruct their sector contractor to use for translation of XML submittals. We will complete the convertor software by March 31, 2004.

We are establishing a process that will result in the XML files submitted to CDX automatically passing through validation software maintained at CDX, to confirm presence of mandatory fields, correct data type and length, valid code values, and in some cases appropriate data ranges. If the file 'fails' those validation checks then the file will be automatically returned to the sender with problem statements and the status will be noted as "not received". After successful iteration and validation at CDX, the XML file, like others submitted through CDX, will pass into EFIG's (Sally's) box at CDX with notification to her that the file is received, upon which she will do a chain-of-custody procedure, and pass it like any other file to the appropriate data development contractor. For XML files, the contractor will have to operate the XML file translator provided by EFIG to produce the NIF 3.0 file (which should by now be assured as 'clean') for their further processing.

How will Data Providers transfer data to EPA? Who in EFIG will handle the data receipt?

As noted above, many of the data providers will transfer data via EPA's Central Data Exchange (CDX). Whether a S/L/T uses the NIF or XML format noted above, they will all move their files to us through the CDX node on the National Environmental Information Exchange Network (NEIEN). The NEIEN is administered in partnership with the Environmental Council of States (ECOS) and is maintained by the EPA's Office of Environmental Information (OEI) for all data flows in the Agency. Some S/L/T agencies also operate a node on the NEIEN, and more are expecting to do so in the future. Node-to-node, or automated machine-to-machine transmissions over the NEIEN are planned in the future for the NEI data transfers and for other data flows within the Agency. Node-to-node exchanges will likely use XML. The OEI and ECOS are designing templates for trading partner agreements (TPAs) that will be made between entities (agencies or other groups) exchanging data over the network. The TPAs may reflect agreements on security plans, data format(s), the data to be provided, and the exchange schedule. We will participate in TPAs with OEI and S/L/T partners if, and as, they become a necessity for our collective participation on the NEIEN - including any TPAs that would need to be put in place for exchange of 2002 NEI data.

With the support of OEI, we will test node-to-node NEI data exchanges with a few States during FY2004; however, at this time, we do not expect to do real-time node-to-node data exchanges with States by June 1, 2004 for the transfer of the 2002 data to NEI. Should node-to-node exchanges occur, we may modify this Plan at a later time to describe how the node-to-node originated file transfer will interact with the NEI data management system flow and processing.

CDX registration and submission procedures for S/L/Ts are located at <http://www.epa.gov/ttn/chief/nif/cdx.html>. The sequence of events involved in the electronic transfer of files through CDX are summarized below, and our areas of responsibility and action are noted:

- Data submitter registers on-line at CDX.
- CDX on-line - the data submitter completes NEI submittal form to generally describe the data, zips and names the file as prompted by CDX instruction, and sends the file.
- Submitters receive an automated message from CDX upon receipt of files at CDX and a second automated message from CDX upon download by us.
- EFIG personnel (Sally Dombrowski with Lee Tooty serving as backup) downloads the files received at CDX from a secure password-protected CDX website to a secure local network drive on one of our servers accessible by us.

How should S/L/Ts indicate in their data submittals that they believe emissions are zero?

It is possible to note this information on the CDX Transmittal Form. However, a better place to note it is on the Transmittal Record which is more closely connected to the data. The comments need to clearly state that emissions values of zero in the transmittal are actually zero and should not be replaced by EPA with nonzero values.

We suggest that S/L/Ts include in the comment field of the Transmittal Record, or in an email to Sally Dombrowski sent just before they submit their data to CDX, the following statement (or something to this effect):

"Emissions values of zero in this transmittal should not be replaced by EPA with nonzero values."

Note that this statement is 79 characters which fits into the Transmittal Record comment field which allows a maximum of 80 characters.

How will we transfer the data to the data development contractors?

In summary, when a file is received through CDX, Sally will:

- 1) Download the original file into the directory designated as original and mark the file read only;
- 2) Make a copy of the original file, place it into the working directory, complete the Chain of Custody (COC) form and rename the file according to the file naming convention described in Chapter 2 (General Issues);
- 3) Run the QC tool (nonpoint files only and only as time permits), and place the resulting error file in the working directory. Any changes to field or table names, to make the file run through the tool, will be designated on the COC;
- 4) Download submittal form information to the DataSubmitted file. These forms generally contain additional information from the State with regard to their data. We will place a copy of the submittal form in the original and working directory; and
- 4) Send an E-mail message to the data developer, RPO, regional contact, and data development contractor that the data has been received and is available on the EPA ftp site.

Turn around time to the data developers should be the same day that we download the data.

Sally will manage the processing of files from CDX to our data development contractors. The NEI submission made for each individual CDX / NEI account consists of two files. One is a

text file containing the information entered into the NEI Submittal Form by the account holder during the on-line submittal process. The text file includes the submitter contact information, a description of the data being submitted, and any comments provided. Each text file is automatically loaded into a spreadsheet which contains a macro program to format the meta data in a reader-friendly format. The result is a running summary of meta data for all the files received in EFIG through CDX. The spreadsheet <datasubmittals2002.xls> is posted on a designated local drive.

Note: Our experience indicates that the meta information provided on-line by the data submitters about the file data contents (and subsequently copied into the spreadsheet mentioned above) does not always accurately reflect the actual contents of the data file. Thus, we should confirm the actual contents before we begin working with the data file. Errors in the meta information should be noted when identified.

The second file is a zipped (.zip) file which contains the emission inventory data. We will archive a copy of the data files to a designated S/L/T directory on a local area network drive accessible by EFIG staff. We will lock the original 'archived' file(s) against general shared LAN access to prevent changes. We will place a copy of the file(s) in a designated shared working directory for processing.

We expect the file(s) in the zip file to be in NIF V3.0 as a .mdb or .txt file type., or in XML as a .xsd file type. We will further process only files in NIF or XML. If a data file is confirmed to be in some other format, we will send an email to the data submittal contact person describing why we are unable to process the data file(s).

We will complete a Chain of Custody (COC) form for each source category received. On the COC, we will include contact information for:

- submitter
- original file name
- file size
- number of records in each table
- new file name
- name of the data development contractor

Once the data development contractors have been notified that a data submittal is available, they will download a copy of the zip file which contains the data, COC, and any additional information files to their offices for further processing (QC, blend/merge, augmentation). Each data development contractor will be responsible for converting the XML files into NIF Version 3.0 and determining if all the NIF files are correctly formatted, including contacting the data file submitters to resolve problems. If necessary, and if time allows, we may request that certain S/L/T agencies send us a corrected data submittal through CDX.]

What are our criteria for acceptance/nonacceptance of the data?

We describe details of the acceptance criteria during initial data submittal in the QC chapter.

Where will we store and/or copy the data for archival purposes?

We will archive the following types of files on the EFIG servers by S/L/T agency:

- original data submittals and resubmittals
- internal EPA draft and revised data sets
- documentation of procedures and methodologies
- documentation of communications (phone calls, emails, etc)

We will set up directories in EFIG servers that the data development contractors can access to upload data. We will structure the directories according to source sector (point, nonpoint, onroad, nonroad), type of archive files (e.g., original submittals), and geographic location (tribe, state, county, etc).

What impact will the EPA Data Standards and Information Quality Guidelines have?

In a continued effort to foster consistently defined and formatted data elements and sets of data values, we have designed the NIF V3.0, and subsequently will complete the update of the NEI XML to Version 3.0, to comply with EPA data standards. The EPA data standards relevant to NEI include: Facility Identification, Latitude/Longitude, Chemical Identification, SIC/NAICS, Contact Information, Tribal Code, and Date. The desired endpoint for adopting these standards across EPA is to enable Agency systems to share related or same data in a consistent manner in order to provide public access to more meaningful data. Strategies for complying with the standards relevant to the NEI begins with how we collect the data. Many of the standards are adopted directly into the NIF Version 3.0 and NEI XML Version 3.0. Some data for a relevant standard will be collected in a non-standard manner, but one that is historically familiar and successful to the NEI data collection process. In those instances, we will transform the data element terminology received to the standard terminology required for data system sharing and distribution.

The data collection phase is the earliest stage in the NEI data compilation process to invoke some measures to ensure the IQ Guideline performance goal of disseminating information that has a basic standard of quality, objectivity, utility, and integrity. The data collection process sets into motion our documentation of information in a standard form, where the data originated, how we handled it to preserve meaning and respectful use, and if necessary, what steps we took to ensure its utility.

Chapter 4: How We Will Process Data and Compile the Inventory

In the previous chapter (How We Will Collect Data), we described our plans for collecting the data that will serve as the basis for the NEI. In this chapter, we begin with the point at which our contractors have the necessary data in hand and we describe the steps they and we will follow to prepare the draft and final versions of the NEI.

We will accomplish the following major steps to compile the NEI:

- Blending/Merging of data from multiple data sources;
- Augmentation of blended data for missing data elements;
- QC/QA of data (discussed in the QC Chapter);
- Preparation of draft 2002 NEI for external and internal review;
- Incorporation of external and internal review comments on the draft 2002 NEI and incorporation of new inventory data submitted during review period; and
- Preparation of final 2002 NEI.

While these steps help to distinguish the various activities involved, there will actually be much overlap between these steps (i.e., one step may still be ongoing when we begin the next step). Certain activities are performed simultaneously during the development of the 2002 NEI. The description in this chapter provides a description of key steps, but does not describe all steps within each step chronologically.

How will we Blend/Merge the data?

We describe below the procedures we will follow to blend and merge the data. The discussion is divided by source sector.

NonEGU Point Sources

The procedures we will follow to compile the point source portion of the NEI are described in the document titled “NEI Quality Assurance and Data Augmentation for Point Sources”, available at www.epa.gov/ttn/chief/net/2002inventory.html#point

EGU Point Sources

The steps for preparing the EGU portion of the point source inventory are described in the document titled “Documentation for the 2002 Electric Generating Unit (EGU) National Emissions Inventory (NEI)”, available at ftp://ftp.epa.gov/EmisInventory/prelim2002nei/point/documentation/egu2002nei_final.pdf

Nonpoint Stationary Sources

Procedures we will follow to prepare the nonpoint source inventory are described in “Documentation for the 2002 Nonpoint Source National Emission Inventory for Criteria and Hazardous Air Pollutants (January 2004 Version)”, and associated documents (re Commercial Cooking, Wildland Fires, and Ammonia Emissions), available at www.epa.gov/ttn/chief/net/2002inventory.html#nonpoint

Onroad Mobile Sources

Procedures for the compilation of the Onroad Mobile Source sector of the NEI are

described in "Documentation for the On Road NEI for Base Years 1970-2002", available at www.epa.gov/ttn/chief/net/2002inventory.html#onroad

Nonroad Mobile Sources

Procedures for the compilation of the Nonroad Mobile Source sector of the NEI are described in "Draft Documentation for the NONROAD Model Criteria and Hazardous Air Pollutant Components of the National Emissions Inventory (NEI) for 2002 Version: January 2004" and in several documents related to Air, Rail, Locomotive, Commercial Marine Vessels, available at www.epa.gov/ttn/chief/net/2002inventory.html#nonroad

How are MACT codes assigned in the NEI?

To evaluate EPA's progress in reducing air toxic emissions through the Maximum Achievable Control Technology (MACT) standards and to identify sources that may be modeled as part of residual risk assessments, operations within facilities that are subject to MACT standards will be identified in the NEI by MACT codes. The tagging of data with MACT codes allows EPA to determine reductions attributable to the MACT program. The NEI associates MACT codes with major and area source data for HAPs. MACT codes are assigned at the process level. MACT codes are also assigned to source categories in the nonpoint source file. In the 2002 NEI, we will add additional codes for 70 area source categories that will be evaluated for the development of area source standards. We request that S/L/Ts include MACT codes as part of their submittal of 2002 HAP emission inventory data. If data submitters do not include MACT codes in their inventories, then we will assign MACT codes.

What steps will we take to meet the EPA's Information Quality Guidelines?

We also intend to store in the 2002 NEI additional information that we generate (when not reported by the S/L/T's) such as activity, emission factors, and control information that we will use to calculate and add missing data. We will include codes noting the origin of the data (e.g., reported by S/L/T agency or generated by EPA) and information on the quality of emission estimates. Due to implementation of EPA's new Information Quality Guidelines, we are increasing emphasis on our documentation of the source and basis of data in the NEI (e.g., methodology used to generate NEI emissions). To help us track the quality of the data, we will prepare emission data quality ratings for the point source emissions data.

What are the procedures for preparing the draft and final NEI files?

Preparation of Draft 2002 NEI Files for Review

To improve the quality of the first draft of the NEI for HAPs, we will make the draft 2002 NEI available for review from February 1, 2005 - April 30, 2005. Reviewers will be asked to identify duplicate facilities, sites, and emission records; identify facilities not operating in 2002; and to provide emissions data for missing facilities and source categories.

The draft 2002 NEI for HAPs will be posted on the following ftp site:
<ftp://ftp.epa.gov/EmisInventory/draftnei2002/>. The ftp site will contain the following files:

- README file
- Documentation
- State data files in NIF Version 3.0. and
- Data Summary Reports.

README file

The README file contains an overview of the files posted on the ftp site, organization of files, description of files, helpful hints on how to review data, instructions on submitting revisions, schedule, and software needed to review the data. Reviewers are strongly encouraged to first read the README before downloading any files or reviewing the draft inventory.

Documentation

Separate reports will be available that provide documentation for the draft 2002 NEI point, nonpoint, onroad mobile, and nonroad mobile source inventories. Reviewers are encouraged to read the documentation to better understand how the draft inventory is compiled and the sources of data. The documentation will specifically address how the NEI meets IQ Guidelines and EPA Data Standards.

State Data Files in NIF Version 3.0

The ftp site will have separate point, nonpoint, and mobile zipped files for each state including Washington DC, Puerto Rico, and Virgin Islands. We will provide ACCESS 97 data files to post in NIF-like format with additional data fields to indicate defaulted fields. The naming convention for these files will be: XX2002NEIPTCAP01022004DRAFT.zip, XX2002NEINPHAP01022004DRAFT.zip, etc. where XX is the state abbreviation.

Data Summary Reports

A series of reports will summarize the data in the draft 2002 NEI to facilitate its review. We will post the following summary data files. Refer to Appendix B for data field descriptions for each of the summary reports listed below.

- 99-2002 Site List - provides a list of sites found in the 1999 base year NEI, the 2002 draft, and sites common to both the 1999 and 2002 inventories. By sorting this list by state, county, and facility name, reviewers could evaluate the sites listed in each county and detect potential duplicates. If a site is found in the 1999 version but is not in the 2002 draft, reviewers could verify that the facility closed.
- 2002 Facility Data Source Summary provides summary emissions data for each facility/pollutant/source category combination where more than one data source (state or local agency, TRI, ESD, etc.) is available. The report indicates which source was selected for inclusion in the draft and allows a quick comparison of data from the different sources.
- 2002 County Emissions Summary - presents NEI pollutant emissions by state and county for major, area, onroad and nonroad sources. This report helps reviewers identify states and counties to investigate pollutant emissions by source sector
- 2002 Source Category Summary - presents NEI pollutant emissions by state, county, and tribe for source categories. Tier reports are used for criteria pollutants and the NEON specs are used for HAPs.
- 2002 Point Source Facility Summary - presents NEI pollutant emissions for each facility with facility data and information about the emissions (major vs area, MACT codes, etc.). This report helps reviewers to target facilities to investigate further by using the State data Files.
- 2002 Point Source Stack Summary - presents NEI pollutant emissions for each emission release point in the NEI. Included will be unit ID, process ID, emission release point ID, SCC, MACT code, SIC/NAICS code, emission release point type, latitude/longitude, data source, etc.

- 2002 Nonpoint Source Summary presents a summary of nonpoint emissions by pollutant for each county and include information on the source category by SCC and MACT code. The source of data is also contained in this report.
- 2002 Onroad Summary presents a summary of onroad mobile data by the 8 vehicle classes for each pollutant emitted within a county.
- 2002 Nonroad Summary presents nonroad mobile source data summarized by the TIER 3 categories for each pollutant and county.

Incorporation of Draft Review Comments

For incorporation into the final 1999 NEI, reviewers will need to submit their revisions in the format specified by EPA.

Data Revisions

Data revisions will be submitted electronically to EFIG using the NIF Version 3.0. We will request that revisions be coded in terms of how reviewers want us to change data as it appears in the public review version of the draft NEI. The data changes(s) must be coded in the NIF Version 3.0 in the form of an addition, deletion, or revision to what exists in the draft NEI. All data changes must be done for **whole**, not partial records.

NIF Data Records and Elements Involved

Submitting data revisions in the NIF V 3.0 will include the following records and data elements to describe the type of revision being requested:

Record(s) Data Elements. Transmittal (TR) TRANSACTION TYPE, INCREMENTAL SUBMISSION NUMBER All except TR SUBMITTAL FLAG

Use Conventions. At the end of each NIF record, except for the Transmittal record, the last data element field is the **SUBMITTAL FLAG**. The SUBMITTAL FLAG data element is a coded field that must be reported in each record to specify the type of data correction being requested. The code values and their meaning are:

- A** Add - the record data reported (e.g., included therein) does not exist in the NEI public review version and reviewer wants us to add it. If the reviewer requests that we add (A) a record that has subordinate related records, the reviewer must include those specific records in the Add request.
- D** Delete - the record data reported does exist in the NEI public review version and reviewer wants us to delete it, without replacement. When a reviewers request we delete a specific record (e.g., D), we will also delete, if present, any related, subordinate records.
- RD** Revise/Delete - the record data reported exists in the draft NEI and a reviewer wants us to revise it with the respective corresponding record with SUBMITTAL FLAG code = RA.
- RA** Revise/ Add - the record data reported should supercede (*replace*) that in respective record indicating the RD (above).

If a reviewer wants us to *replace* some of the data that exists in the draft NEI, there must always be two records provided in that case - one indicating SUBMITTAL FLAG =Revise/Delete (RD), and a corresponding record type indicating the Revise/Add (RA). The RD acknowledges the *whole* record in its entirety that exists in the draft, and the RA indicates any corrections (replacements) requested for those same data elements that exist in the draft record, plus may report additional data elements if relevant (that did not exist in the NEI record).

Important: For RD / RA record pairs, the reviewer does not need to re-submit the subordinate records of a changed record if the subordinate records do not contain changes. *The RD / RA record type pairs must report the same data values for the primary key fields to allow the records to be sorted together for processing.* If a review is correcting a data value for one of the primary key fields in a record, that actually creates a *new* record (that does not exist in the NEI), then the data should be submitted as an Add, and should include any subordinate records. The specific record in the NEI that contains the incorrect primary key data value should then be coded with a D, to Delete that record.

A **Transmittal record (TR)** must accompany each source type file (i.e., point, nonpoint, nonroad mobile, onroad mobile) of records the review submits to EFIG. The information in the Transmittal record is used to describe the entire source file and its origin. Submit one TR record per unique county. This, along with ORGANIZATION NAME will help distinguish and track local (county) agency submittals separately from the state agency submittals. Please refer to the NIF for a full explanation of how to implement the TR record. The following TR record data elements are especially relevant to the data correction / replacement scenario.

TRANSACTION TYPE = 'replacement' (05) for submittal of a source file type previously submitted, or data correction.

INCREMENTAL SUBMISSION NUMBER = a unique number that differentiates incremental submissions of data. Use '1' for first submittal of data correction / replacement set.

Electronic Submittal of the Data Corrections – Central Data Exchange (CDX)

Once the data changes are implemented in the NIF, the electronic file will be submitted to EFIG through the EPA's Central Data Exchange facility. CDX will be used in the same manner as for submission of the original data sets from agencies. Those agencies that have previously submitted data sets through CDX will already have an account and password. Be aware that when transferring data electronically through CDX, users are asked to zip their files and name the zip file according to a specified CDX file naming convention. When we unzip the files received from CDX, the initial file names described above remain apparent.

Process of Incorporating Revisions

Revisions to the 2002 NEI will be subjected to a rigorous review process in order to ensure internal consistency of the NEI. Specifically, the following steps will be performed:

- QC format of revised and new records;
- Review of the documentation provided by the commenter to ensure that it is consistent with the actual changes to the inventory submitted;
- Verification that the add/revise/delete designations of the revisions are accurate (e.g., that a record designated for addition is not in fact a revision to an existing record);
- Verification that there is no source category overlap between the existing draft and revised data (non-point stationary source categories);
- Review of the revised emission estimates, by source category and pollutant, to identify outliers and determine the validity of such estimates;
- Verification that added pollutants are HAPs or CAPs; and
- Verification that added pollutants have correct or valid Chemical Abstracts Service numbers.

Revisions will be grouped into the following three categories.

- Changes to draft emission records (e.g., draft emissions for a HAP were 10 tons and

- revision changed the emissions to 5 tons);
- Additions of new facilities not in the draft NEI or additions of pollutants to a facility/nonpoint or mobile source category in the draft NEI; and
- Deletions of draft NEI facilities/source categories or deletions of pollutants within a facility/source category.

For the most part, revisions provided by state and local agencies and tribes will be incorporated to produce the final inventory. When questions arise over specific revisions, the reviewers will be contacted by the EPA. Memos describing how specific revisions are addressed into the final NEI will be sent to the reviewers.

Point source revisions will be incorporated into the final NEI using the following methodology.

1. All revisions and additions in correct format, with valid codes, and within valid data ranges will be made for most non-emissions fields including:
 - Process throughput information,
 - Zip codes,
 - Control/capture efficiencies,
 - Control device information,
 - SCC/process descriptions,
 - Stack parameters,
 - Unit descriptions,
 - Lat/long or UTM coordinates,
 - FIPS codes, and
 - SIC codes.
2. Additions of new facilities not in the draft NEI will be automatically processed.
3. Emission record additions and revisions to existing emission records will be automatically processed if the original data are provided by the state or local agency or tribe.
4. Facilities and pollutants marked for deletion will be evaluated to determine whether the deletions should be processed.
 - If the draft NEI data are provided by state or local agencies, then we will process the deletion.
 - If the draft NEI data are provided by TRI or in MACT databases, then we will further evaluate the proposed deletions. We will use trade association journals to determine if facilities were in operation in 2002. If a facility was closed in 2002, the EPA will delete the facility.
 - If a TRI or MACT facility is recommended for deletion, but was open in 2002, and no duplicate facility is present in the draft inventory, then we will not process the deletion.
 - If a TRI or MACT facility is recommended for deletion, was open in 2002, and a duplicate facility is present in the draft inventory, then we will evaluate the HAPs emitted from both facilities. If the HAP emission records are duplicated, then we will process the deletions. The HAP emission records that are not duplicated will be retained. We will then merge the HAPs for the duplicate sites using a NEI ID.

For nonpoint and mobile source sectors, data submitters will be required to submit data for all pollutants. Thus a reviewer should not submit partial pollutant revisions, i.e, revise VOC mobile source emissions and not revise HAP mobile source revisions or submit PM- CON revisions but not PM25-PRI revisions. We will contact reviewers to request all pollutants if partial submissions are received. If the reviewer does not provide data for all pollutants, we will

not process the revisions.

Industry is encouraged to review the draft NEI and provide proposed revisions directly to state and local agencies. Industry revisions submitted directly to the EPA will be sent to state and local agencies to determine if they should be incorporated into the final NEI.

Maintaining an Emissions History Table

We will create and maintain an emissions history table which tracks the changes in total emissions for each facility/site/pollutant combination in iterative versions of the 2002 NEI. This table will enable users to see past and current total emissions for each pollutant from a facility/site at a glance. This table will contain the following elements:

- Status - Indicates if a facility has closed or was added after the initial version;
- State FIPS Code - Two digit state code;
- County FIPS Code - Three digit county code;
- State or Local Site ID - Plant ID assigned by state, local agency or EPA;
- NEI ID;
- Facility Name - Plant Name;
- Pollutant Code - Chemical Abstract Service (CAS) number;
- Emissions Value 1 - Emissions in TPY associated with 01012004 draft 2002 NEI;
- Data Source 1- Source of emissions value in 01012004 draft 2002 NEI;
- Date of release of 01012004 draft 2002 NEI;
- Proposed Emissions Value 2 - Emissions in TPY of proposed revision
- Proposed Data Source 2 - Source of emissions value in proposed revisions;
- Proposed Date 2 - Date of proposed revision;
- Resolution of Revision (process don't process)
- Emission Value 2;
- Data Source 2;
- Date of Release 2

This table will be aggregated to the site/pollutant code level and will provide a total HAP and the 6 criteria pollutant emission estimates for all processes associated with each site. As new versions of the NEI are created, a new column will be added to this table providing the total pollutant estimate and the data source (i.e. state or local agency, industry, EPA or TRI) for the newest version.

If a site closed (and is deleted from the inventory), the history record will be retained, and the status column will be updated to "D." If a site is added that did not exist in previous versions, the status will be updated to "A" and the columns related to the most recent version (and every version, thereafter) will be filled. Maintaining a derived table like this can quickly become unwieldy if too many data elements are added or the level of disaggregation is too detailed. This table is not meant to provide all the information that the database contains, but to provide a quick overview of the changes in emissions values over the course of the revision process. It is meant to help the user target facilities for further investigation, not provide all of the supporting information.

This emissions history table will be supplemented by the "holding" tables which contain the specific and individual changes submitted during each review period. These holding tables mirror the NIF tables and contain complete NIF records. These records have a "submittal flag" which indicates the change type ("A", "RA" or "D") and contain a source field which describes the submitters of the change. We will improve this table by adding a comment field. This field would be populated in those cases in which a user has provided a specific reason for a change.

In addition to the "holding" tables, when we use the "merge" algorithm to choose one pollutant value from among multiple data sources for the same pollutant/facility, We will save the "rejected values" to subsidiary tables along with the list of data source choices. We will experiment with merging this information with the emissions history table. If supplementing the

history table with the “rejected values” is not efficient, we will retain this as a stand-alone table.

Preparation of Final 2002 NEI

After the revisions to the draft 2002 NEI are incorporated, the EPA will conduct final format QC and produce the 2002 NEI for HAPs. The final 2002 NEI for HAPs will be posted on the following ftp site, <ftp://ftp.epa.gov/EmisInventory/nei2002/>. The ftp site contains the following files:

- README file.
- Documentation files
- State data files in NIF Version 3.0.
- NEI lookup database; and
- Data Summary Reports.

The files are similar to the draft files described earlier.

README File

The README file contains an overview of the files posted on the ftp site, organization of files, description of files, and description of how the 2002 NEI meets OMB Information Quality Guidelines. Users of the 2002 NEI are strongly encouraged to first read the README before downloading any files.

Documentation

Separate reports will be available that provide documentation for the 2002 NEI point, nonpoint, onroad mobile, and nonroad mobile source inventories. We encourage people who access the data to read the documentation to better understand how the inventory is compiled and the sources of data. The documentation will specifically address how the NEI meets IQ Guidelines and EPA Data Standards.

State Data Files in NIF Version 3.0

The ftp site will have separate point, nonpoint, and mobile zipped files for each state including Washington DC, Puerto Rico, and Virgin Islands. We will provide ACCESS XP data files in posted in NIF-like format with additional data fields to indicate defaulted fields. The naming convention for these files will be: XX2002NEIPTCAP10022004.zip, XX2002NEINPHAP10022004DRAFT.zip, etc. where XX is the state abbreviation.

NEI Lookup Database

The NEI lookup database will contain all of the codes and flags used in the data files. The pollutant HAP dictionary will comply with the U.S. Environmental Protection Agency’s (EPA’s) Office of Environmental Information (OEI) Data Standards and contains several important fields which map the NEI pollutant codes to the Chemical Identification Data Standard.

Data Summary Reports

In addition to the NEI documentation and data files, additional files are provided to facilitate the evaluation of the NEI, and to help put the emission estimates presented into perspective by state, county, source category, and facility. The summary files and documentation reports also allow one to clearly identify the source of emissions data selected for each point source facility and each nonpoint source category. In each summary file, emissions are presented for each 188 HAP category, as the sum of the 188 HAPs, and as the sum of the 33 urban HAPs used by EPA in many air toxics programs. Each 33 urban HAP is flagged as such. Emissions are also presented for each individual HAP species in all files except for the county emission summary, the source category summary and the point source facility

summary files. Each county is flagged with the urban/rural designation developed under EPA/s Integrated Urban Air Toxics Strategy.

- *County Emission Summary* presents the NEI pollutant emissions by state, and county for major, area, onroad, and nonroad sources. Major and area sources are also summarized as Maximum Achievable Control Technology (MACT) vs. non-MACT source categories.
- *Source Category Summary* presents the NEI emissions by state, and county for major, area, onroad, and nonroad sources. The area sources are delineated as point or nonpoint. Each stationary source category is presented by MACT code, Standard Industrial Classification (SIC) code, or just source category name if there is no applicable MACT or SIC code.
- *Point Source Facility Summary* presents the pollutant emissions by NEI Facility and individual site for major and area point sources. Included with each facility record is the address, site latitude/longitude, emission type (actual, allowable, potential, etc.), MACT and/or SIC code. The source of the emission estimate, whether original data or recently revised, is also noted as state/local/tribal, MACT, Toxics Release Inventory (TRI), industry, or 1996 NTI.
- *Point Source Stack Summary* presents the pollutant emissions by NEI ID and individual site for major and area point sources. Included with each record is the emission type (actual, allowable, potential, etc.), emission unit ID, process ID, emission release point ID, source classification code (SCC), MACT and/or SIC code, emission release point type (stack/vent or fugitive), and latitude/longitude of the emission release point. The source of the emission estimate, whether original data or recently revised, is also noted as state/local/tribal, MACT, TRI, industry, or 1999 NEI.
- *Nonpoint Source Summary Report* presents the NEI pollutant emissions by state, county, and area source category. Included with each record is the emission type (actual, allowable, potential, etc.), SCC/AMS code, MACT and/or SIC code.
- *2002 Nonroad Summary* presents nonroad mobile source emissions data summarized by the TIER 3 categories for each pollutant by state, county, and source category.
- *2002 Onroad Summary* presents a summary of onroad mobile emissions by the 8 vehicle classes for each pollutant by state, county, and source category. Included with each record is the mobile model used to estimate emissions.

Preparation of the Point Source Summary Files

Prior to the creation of the summary files, the point source files will undergo a "standardization" process to eliminate possible double counting and annualize all estimates. Where there are multiple estimates for a HAP at a given emission release point (i.e., defined by state and county FIPs, site ID, unit ID, process ID, and emission release point ID), one record is chosen according to the following logic:

- Data for the most recent year gets preference over older data (e.g., 2002 data are preferred over 1999 data);
- When information is provided for two different periods for the same emission release point, the more complete period gets preference over incomplete periods (365 days over 79);
- An emission type hierarchy is established, and higher types get preference over lower ones (entire period > average > potential > maximum annual > maximum > maximum allowable > average daily > actual hourly > maximum hourly > unknown);
- Where there are multiple metallic HAPs associated with the same emission release point and one of the compounds is a specific compound and the other is not, the specific HAP is retained over the non-specific HAP grouping (e.g., "Chromium (VI)" is retained over "Chromium and Compounds"); and
- Finally, all emissions are converted to tons/year.

The standardized emissions are then used to determine if the facility is major or area based on the CAA definition of major vs area source. The facility category field in the sites table is updated using this assignment.

Note: Records eliminated from the standardized file are retained in the inventory and can still be found in the state output files.

In addition, public summary data will be available via NEON and Air DATA. Modelers files will also be prepared.

Chapter 5: How We Will Distribute Data

Data distribution will occur after the data development contractors produce the final 2002 NEI. All files will be generated from the final NEI database residing at EPA and we will insure that consistency is maintained between the different files.

We will post the files on the ftp site: <ftp://ftp.epa.gov/EmisInventory/nei2002>

We will include the following types of files:

- README
- Documentation
- State data files in NIF Version 3.0
- NEI lookup database - dictionary
- Data summary reports

In addition, we post modeling files at: <ftp://ftp.epa.gov/pub/EmisInventory/neimodel>

We will include the following types of modeling related files:

- README
- Files by source sector (point, nonpoint, onroad, and nonroad)
- NEI lookup database - dictionary

We will make NEI summary data accessible from a variety of locations including the NEI ftp site, NEON, and Air DATA.

What are the User Needs for NEI data distribution?

Our User Needs for NEI data distribution include:

- 4 Sets of NEI Input Format Files (Point, Nonpoint, OnRoad, NonRoad) with additional fields added by data developer contractor, as agreed upon by us
- Blank field in EM Table will identify whether emissions are Annual (A), Seasonal (S), or Daily (D)

The Data Development Contractors will provide us, in addition to the standard NIF fields, additional fields necessary for managing the NEI. One of the fields will identify the source of the data. This is necessary because our contractors send us data that came from many sources and we need to be able to easily identify the source of the data. The data source field is also utilized when the contractors make changes to State data.

The emission ton (pounds for HAPs) field will be included because the data submitters and contractors send us emissions values using various units. Including emissions standardized in terms of tons enables us to easily create summary reports summing all of the ton values.

The start and end date and emission type field do not provide sufficient information to determine if an emission number is Annual, Seasonal, or Daily. For example, a number could appear to be seasonal (start date > 1/1/0X and end date < 12/31/0X) yet actually be annual because the plant was not in operation for part of the year. This is especially important for our data developer contractors who will attempt to contact the source of the data to confirm the period type and place it in the period type column so that we do not have to guess the basis for the number. This will allow us to easily create reports of annual emissions summing all of the annual values.

The Data Development Contractors will provide a list of ALL deviations from NIFv 3.0. This will include any and all table and field name changes, field data type changes, field length changes, and any codes that were not in the NIF v3.0 set of acceptable code tables.

The contractors will transfer the zipped files to us via the contractors' public ftp sites. They will also include Chain of Custody (COC) forms that identify the size of the unzipped files, row numbers in each, and emission ton totals by pollutant. (*Refer to the QC Chapter for more information on the COC forms.*) Data developer Contractors will send the documentation files when they send the NIF files. A Row count for the documentation files will not be necessary on the COC.

The EFIG Data Base Administrator (DBA), will then load the new year (2002 NEI) into the NEI (1990-2001). After the 2002 load is complete, the NEI will include data for all years 1990 - 2002.

What will be the procedure to document the codes used in the final NEI?

As the developer contractors process data they may need to add codes. There is one set of code tables in the NEI that is accessed by all of the yearly sets of tables. There is a field in every table that identifies each acceptable code as active or inactive for input. This is especially useful for codes that get split into multiple codes like County FIPS and SCCs. For 1999 Version 3 Criteria, this set of codes is simply updated with additions. For the 1996 and 1999 HAPs there is a set of code tables for each year. Because 2002 is the first attempt at HAP and criteria pollutant integration, we are requiring along with the NEI data sent by the contractors a codes database and COC form. The pollutant HAP dictionary will comply with the EPA Office of Environmental Information (OEI) Data Standards and will contain several important fields which map the NEI pollutant codes to the Chemical Identification Data Standard.

What will be the steps to generate summary files?

As we are still in the early stages of the process of combining the Hazardous Air Pollutants (HAPs) and criteria pollutants, for the 2002 NEI we have yet to automate the summary report generation for HAPs in the NEI database. For 2002, the HAP summary files will be generated by the contractor while the raw data for all sources is being loaded into the NEI and the criteria pollutant summary files are being created. When the HAP summary files are complete, they will be sent to EFIG via the same contractor ftp site and accompanied by the same COC form. Our slightly longer term goal is to make the summary files a consistent format for both HAP and criteria pollutants and to distribute the same set of files to all receivers mentioned below.

After loading the raw data, our DBA will create the NEON summary files (5), AIRData Summary files (2), and Long Range Transport of Air Pollution (LRTAP – international organization) files from the Annual CAP data only. The National Association of Records Administration (NARA), Envirofacts, and OAQPS Data Warehouse databases require all 7 of these files. Once the summary files are created, TRs can be retrieved via NEON, and we can enable NEON for year 2002 retrievals.

What CAPs Summary Files will we generate?

The five NEON files are SCCANNUAL, SICANNUAL, TIER1ANNUAL, TIER2ANNUAL, and TIER3ANNUAL. These files have county level totals by SCC, SIC, or TIER categories. They each contain the Region, State and County FIPS, either the SIC or SCC or TIER code, the SCC_TYPE (PT, NP, ON, NR for 2002 forward but only kept P and A where A is all not P for 1999 and back), and the annual emission number. The 2 AIRData files are AIRDATA and COUNTY_TIER1_TIER2. The AIRDATA file has State and County FIPS, TIER1 code, TIER2 code, YEAR, State Abbrev., Region, and 4 columns for each of the 12 pollutants (CO, NOX, VOC, SO2, PMCON, PMFIL, PMPRI, PM10FIL, PM10PRI, PM25FIL, PM25PRI, NH3) annual emissions – (PT, NP, ON, NR for 2002 on but only kept P and A where A is all not P for 1999 and back). The LRTAP format is still to be determined.

What HAPs Summary Files will we generate?

In addition to the NEI documentation and data files, additional files are provided to facilitate the evaluation of the NEI, and to help put the emission estimates presented into perspective by state, county, source category, and facility. The summary files and documentation reports also allow one to clearly identify the source of emissions data selected for each point source facility and each non-point source category. In each summary file, emissions are presented for each 188 HAP category, as the sum of the 188 HAPs, and as the sum of the 33 urban HAPs used by EPA in many air toxics programs. Each 33 urban HAP is flagged as such. Emissions are also presented for each individual HAP species in all files except for the county emission summary, the source category summary and the point source facility summary files. Each county is flagged with the urban/rural designation developed under EPA's Integrated Urban Air Toxics Strategy.

- *County Emission Summary* presents the NEI pollutant emissions by state, and county for major, area, on-road, and non-road sources. Major and area sources are also summarized as Maximum Achievable Control Technology (MACT) vs. non-MACT source categories.
- *Source Category Summary* presents the NEI emissions by state, and county for major, area, on-road, and non-road sources. The area sources are delineated as point or non-point. Each stationary source category is presented by MACT code, Standard Industrial Classification (SIC) code, or just source category name if there is no applicable MACT or SIC code.
- *Point Source Facility Summary* presents the pollutant emissions by NEI Facility and individual site for major and area point sources. Included with each facility record is the address, site latitude/longitude, emission type (actual, allowable, potential, etc.), MACT and/or SIC code. The source of the emission estimate, whether original data or recently revised, is also noted as state/local/tribal, MACT, Toxics Release Inventory (TRI), industry, or 1996 NTI.
- *Point Source Stack Summary* presents the pollutant emissions by NEI ID and individual site for major and area point sources. Included with each record is the emission type (actual, allowable, potential, etc.), emission unit ID, process ID, emission release point ID, source classification code (SCC), MACT and/or SIC code, emission release point type (stack/vent or fugitive), and latitude/longitude of the emission release point. The source of the emission estimate, whether original data or recently revised, is also noted as state/local/tribal, MACT, TRI, industry, or 1999 NEI.
- *Non-point Source Summary Report* presents the NEI pollutant emissions by state, county, and area source category. Included with each record is the emission type (actual, allowable, potential, etc.), SCC/AMS code, MACT and/or SIC code.
- *2002 Non-road Summary* presents non-road mobile source emissions data summarized by the TIER 3 categories for each pollutant by state, county, and source category.
- *2002 On-road Summary* presents a summary of on-road mobile emissions by the 8 vehicle classes for each pollutant by state, county, and source category. Included with each record is the mobile model used to estimate emissions.

What other distribution files will we create?

In addition, we will prepare public summary data for access via NEON and AirDATA. We will also prepare Modeling files. While we are creating summary files, the Data Administrator – Rhonda, will create the modeling files for criteria pollutants. The modeling files will be the 2002 NEI in ascii but only the annual. Rhonda will also create, for BOTH HAP and criteria pollutants, the State NEI access databases for annual only. These will be ACCESS XP data files posted in NIF-like format with additional data fields to indicate defaulted fields. The naming convention for these files will be: XX2002NEIPTCAP10022004.zip, XX2002NEINPHAP10022004DRAFT.zip, etc. where XX is the state abbreviation.

What README and Documentation files will we prepare?

The README file will contain an overview of the files posted on the ftp site, organization of files, description of files and format, and description of how the 2002 NEI meets OMB Information Quality Guidelines. We will create this file after delivery of the 2002 NEI. We will strongly encourage users of the 2002 NEI to first read the README before downloading any files.

Separate reports will be available that provide documentation for the 2002 NEI Point, NonPoint, OnRoad, and NonRoad source inventories. We will encourage people who access the data to read the documentation to better understand how the inventory is compiled and the sources of data. The documentation will specifically address how the NEI meets IQ Guidelines and EPA Data Standards.

What are the final steps for distribution of the NEI files?

Rhonda will then put the 7 summary files, LRTAP file, the modeling files, the State files, the README, and the Documentation files on the EPA ftp site at:
<ftp://ftp.epa.gov/EmisInventory/nei2002>

Rhonda will also create an excel version of the TR retrieval from NEON. Ann Ingram will post the TR on CHIEF and create links from CHIEF to the appropriate ftp site for the modeling files, State files, README, and documentation. The ftp site will have separate point, non-point, and mobile zipped files for each state including Washington DC, Puerto Rico, and Virgin Islands. Rhonda will notify AIRData of the location of their 2 files and NARA and Envirofacts contacts of the location of the 7 summary files. The goal is to have the LRTAP report available from NEON but it will at least be available via the ftp.

Chapter 6: How We Will Check Data Quality

What tools will we and our contractors use to check the quality of the data?

We developed a personal computer software tool, called the Basic Format and Content Checker, and an additional tool, called the Extended QC Tool.

The Basic Format and Content Checker is intended to facilitate the incorporation of emissions data into the NEI. It is to be used by us, as well as S/L/T agencies, to check the initial quality of the inventory before additional processing takes place. This tool works only with files formatted in NIF version 3.0. The Basic Format and Content Checker can QC data for the point source, nonpoint stationary source, nonroad mobile source, and onroad mobile source sectors of the inventory. It does not QC the biogenic and geogenic portions of the inventory. An option allows the tool to process NIF 3.0 files in either Access database or ASCII text file format.

The Basic Format and Content Checker checks the format (structure) of the data files, including:

- Correct naming of tables and fields, within the tables;
- Field type;
- Mandatory field entries;
- Duplicate records (based on primary keys); and
- Referential integrity between tables (for widows and orphans).

The Extended QC Tool can be used when format/content problems have been identified and corrected. This tool performs additional QC, including:

- Compare to previous emission inventories; and
- Create "Top 10" lists of emitting facilities and source categories within an area

What QC steps should State/Local/Tribal agencies take prior to submittal of data to EPA?

State/local/tribal agencies should complete a full set of QC checks and any necessary corrections prior to submission of their data. The most basic of these checks are the ones facilitated by the Basic Format and Content Checker, as described above. Details of the types of format checks can be found in the User Guide for the Basic Format and Content Checker located at <http://www.epa.gov/ttn/chief/nif/index.html#qa>.

What procedures will we follow prior to, or as part of, sending data submittals to our contractors?

After downloading submittals from CDX, we will check the working copy of the S/L/T data file using the Basic Format and Content Checker. We will send error reports to the specified inventory development contractor, along with the data and chain of custody files. The contractors will make the decision on whether data can be used. We **will not** make any corrections to the NIF files prior to transmitting the files to the data development contractors.

If we receive files in XML format, we will forward them electronically, in a secure manner, to the data development contractors to be converted through our NEI data system XML translator to NIF Version 3.0 format. Procedures will be identified for notifying the data submitter in the event that incorrect XML formatting is detected. Quality control checks of the XML files actually start at their origin of generation as the schema contains tags that define for each data element some basic format attributes such as field length and data type. During the preparation of the

XML data set, error reports will indicate if the basic attributes are incorrect. In addition, for the XML files, as they are passed through the EPA's CDX, schema validation software will automatically check to ensure conformance of the data to the tag definitions. If the software detects errors, an automated message will be sent from CDX to the data submitter, indicating an inability to accept and process the file due to formatting errors.

What procedures will our contractors follow to check the quality of the data?

The contractors will contact each submitter for which format discrepancies were identified. The contractors will request that the agency resubmit the corrected file, via CDX, within a specified amount of time. If a file is not resubmitted within the specified time, the effected data will be removed from the inventory and replaced with data from the 2002 Preliminary NEI. When we prepare the memos to go out to the data providers about incorporation of their data, we will highlight the types of problems, including format problems, that we found so that they may be corrected in future submissions.

The Contractors will also:

- a. Convert any files to NIF 3.0 that are provided by other EPA offices;
- Identify and remove duplicate records and keep a copy of the records removed from each table;
- Identify and analyze referential integrity issues (i.e., "orphan" and "widow" or "childless parent" records), fill in certain missing mandatory data to reduce referential integrity problems, and add parent records for orphans. When adding parent records for orphans, the contractors will add the minimum required data to make the record valid and "code" the generated records with "E" for EPA to identify that the record was added by EPA. Widow records will be removed from tables, and stored in a separate file. A final analysis will confirm compliance with referential integrity;
- Manually set primary keys on each table in Access;
- Screen for records that contain emissions for pollutants that are not HAPs or CAPs;
- Compare to previous emission inventories (2002 draft, latest trends report, latest NEI (2001, 2000, 1999) by pollutant (all criteria and 33 HAPs), by category (starting at sector level down to Tier 3), and by geographic area (starting at national and working down to state level). Differences greater than 10% will be investigated to either confirm that the differences are justified, documenting the reason for the difference, or to identify errors causing the difference and correct them.
- Create "Top 10" lists by pollutant and source category - includes highest emitting facilities and source categories within a state and county or region;
- Prepare emissions ranked lists and maps by pollutant - includes counties within a state ranked by pollutant and source category emissions;
- Create emission density lists and maps - prepare population-based county level emissions and area-based county level emissions tables by pollutant and source category;
- Prior to sending corrected files to EPA for review, run the dataset through the Basic Format and Content Checker, or through the same checks in an alternative system, for identification of any uncorrected problems.

The contractors will correct any errors found or provide an explanation for any remaining

discrepancies.

For data not provided in NIF (generally EPA provided data such as TRI data, CAMD data, generated data for nonpoint categories, etc.), the contractors will convert data to NIF and conduct format QC on these files. All EPA data will meet the same requirements for referential integrity and data field values that external data are required to meet for incorporation into the NEI.

The individual data element checks and data augmentation procedures will be described in the sector augmentation documents, to be posted on the CHIEF web site. [URLs for the documents will be added to this plan when they are available.]

What QC procedures will we follow after the contractors complete their work but prior to publication of the draft NEI?

After the contractors complete the data quality and completeness checks, corrections, gap filling, and reconciling point and nonpoint source data, we will conduct a variety of content QC on an internal draft inventory prior to preparing the draft inventory files for external review. For the most part, we will identify outliers with very high emission estimates. We have developed the following series of internal QC activities to target outliers and duplicate emissions. Many of these activities can be done through the use of the Extended QC tool.

- Evaluate significant changes between the 1999 NEI and 2002 NEI data, and/or extreme variation within the 2002 data. This will include comparing 1999 emission estimates to 2002 estimates for each facility, source category, county, and state. These big picture summaries will highlight source categories, states, and facilities with potential problems.
- Identify top emitters for each pollutant nationwide, ranking each facility or nonpoint source category based on its emissions of each pollutant on a national basis, and listing the top emitters for pollutant/source category combination nationwide. Statistical tools and GIS software will be used to perform these activities.

Outliers are usually difficult to spot - what appears to be a high emissions value may in fact be acceptable for a particular facility or source category. To aid in detecting these errors, the emissions data will be compared to the range of values in the NEI and the percent contribution to total emissions. A summary table with the list of facilities that appear multiple times as top emitters for different pollutants will help identify point sources and nonpoint source categories with outliers. These high values may be due to a series of outliers or duplicated emission records. The high emissions may also be correct for that facility and category. Thus, these summary data will need to be closely reviewed before any records are recommended for deletion. In some cases, the state/local agency submitting the data will be contacted to discuss the quality of the estimates.

Comparative QC will include analyses such as:

- "Top 10" lists by pollutant and source category - includes highest emitting facilities and source categories within a state and county or region;
- Emissions ranked lists and maps by pollutant - includes counties within a state ranked by pollutant and source category emissions;
- Emission density lists and maps - prepare population based county level emissions and area based county level emissions tables by pollutant and source category; and
- Previous Year Comparison - compare pollutant emissions for national, state, county, source category and facility from previous inventories with the draft 2002 NEI

Statistical QC will include analyses such as:

- Statistical tables (min, max, average, standard deviation, and percentile values) for each pollutant at the county, state and national level - compare pollutant specific statistical values for state to individual counties for source categories and individual facilities;
- Emissions distribution by pollutant - prepare source category, county, state and national level emissions values frequency histogram; and
- Contribution of facilities and source categories to multiple pollutants - compare list of pollutants emitted from a source category to individual facilities and nonpoint categories.

The specific files to help QC the data will include the following reports:

- 1999-2002 Site List - provides a list of sites found in the 1999 base year NEI, Version 3, the 2002 draft, and sites common to both the 1999 and 2002 inventories. By sorting this list by state, county, and facility name, we will evaluate the sites listed in each county and detect potential duplicates. If a site is found in the 1999 version but is not in the 2002 draft, EFIG reviewers can verify that the facility closed.
- 2002 County Emissions Summary - will provide a snapshot of the emissions for each pollutant in each county by sector. These summary tables will enable us to target states and counties for detailed evaluation and to prioritize its efforts in content QC.
- 2002 Facility Emissions Summary- summarizes emissions of each pollutant emitted from each facility. These summary tables will enable us to target facilities for detailed evaluation.
- 2002 Facility Data Source Summary - provides summary emissions data for each facility/pollutant/source category combination where more than one data source (state or local agency, TRI, ESD, etc.) was available. The report indicates which source was selected for inclusion in the draft. It will allow us to quickly compare data from the different sources.
- GIS plots by nonpoint category and pollutant - provides visual distribution of emissions by county.

Upon receipt of our comments, the data development contractors will review the comments and modify the data set to address the comments. A draft inventory will then be produced for posting and external review.

What additional QC procedures will we conduct during the draft comment period, if time permits?

During the external review period, we will continue with enhanced review of the data. The following items are possible routines which could be run during this period:

- SCC - File to file comparison to see if the same SCCs are being reported for nonpoint sources and point sources
- Site Count Comparison for point sources
- SCC /SIC- Are all pollutants being reported per SCC/SIC for all source sectors.

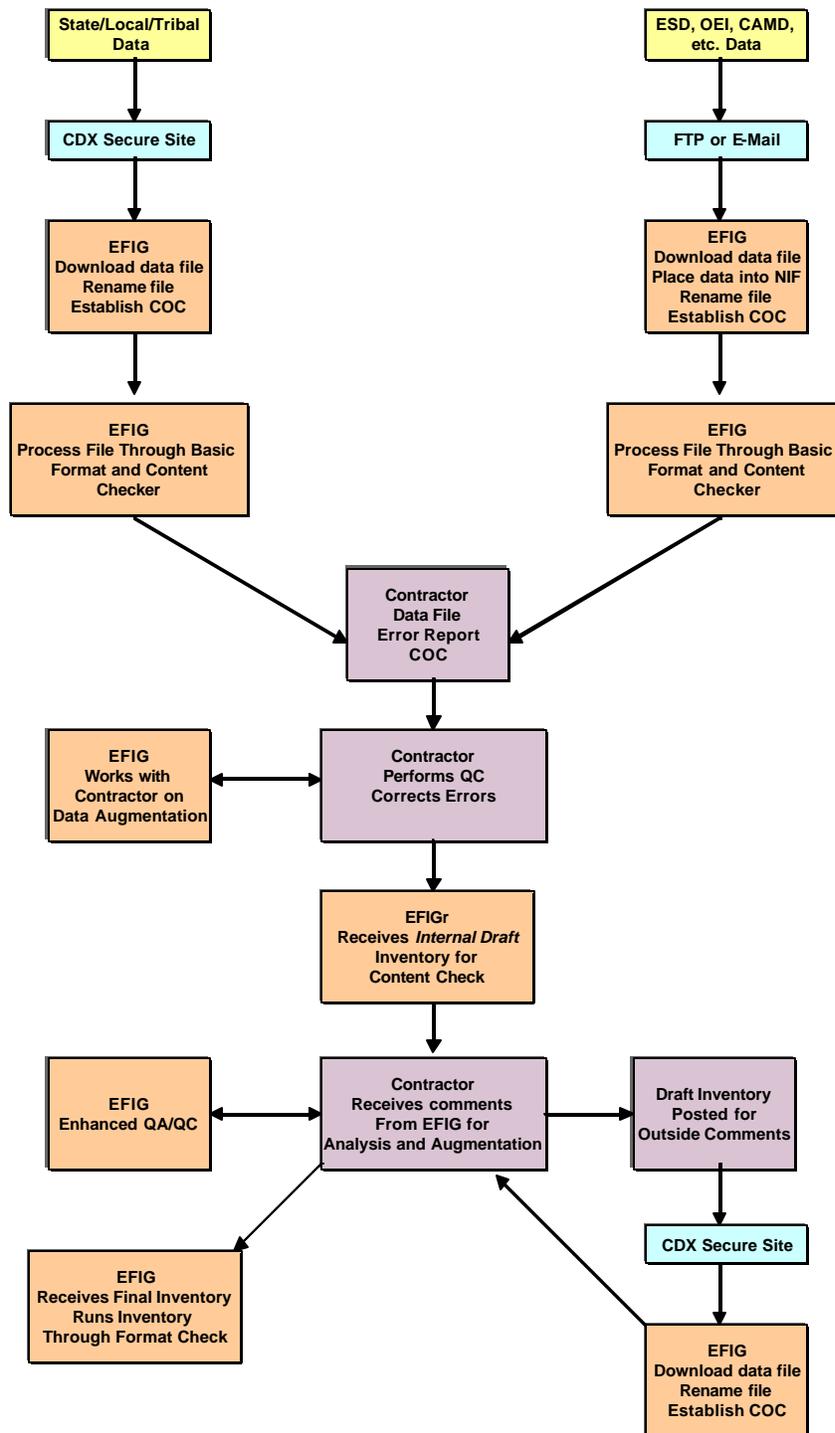
- Using the national emission, identify where sources may be outside the expected maximum emissions by SCC, SIC or MACT using a “look-up’ table from the maximum values in the final QC’d 99 file.

What steps will we follow to process the draft inventory revisions and comments?

Comments and revisions concerning the 2002 Draft Inventory will be submitted through CDX. Files will be downloaded from the CDX website and filed under the directory called “2002 Inventory - Comments.” Subdirectories will be established for each State/local/tribal agency with the original file locked and a working copy made.

A new chain of custody file will be completed and the data files forward to the appropriate data development contractor for review and augmentation.

QC Steps in Data Preparation



Chapter 7: How We Will Peer Review the NEI

In this chapter, we describe the plans for achieving peer review of the NEI process and products.

Why is the NEI undergoing peer review?

A peer review of the development process will be conducted because the NEI is considered a significant and important technical work product that may also support regulatory decisions and strategies for air quality control programs. The NEI data is also considered an important technical product due to the applicability to a broad spectrum of regulated entities and other stakeholders and users through published emission release data. While the NEI is national in scope, it may also have possible significant consequences on a small geographic scale such as input to air quality modeling performed by the Regional Planning Organizations (RPOs) to predict the affect of regional control measures; and community assessments of air toxic pollutant exposures. On a routine basis, the NEI is used to generate and track emission trends which may be associated with current and predicted air quality trends, control policies, and program guidance.

At what milestones in the NEI preparation process will peer review be conducted?

The following milestone dates describe the stages of the EPA's data development process and some of the interim products expected that will also be considered during the peer review. A thorough discussion of these process steps and the anticipated data sources may be found throughout this Plan..

Mar 2004

- Release Preliminary 2002 NEI data, documentation, and message regarding information quality and review instructions to intended audience.
- Distribute Draft 2002 NEI Preparation Plan.
- Contractor support work plans in place.

Jun 2004

- 2002 data submittals due from S/L/Ts.
- Distribute Final 2002 NEI Preparation Plan.

Jun - Dec 2004

- Blend/ merge/ augment data across sources as described in the Preparation Plan and implemented in contractors' work plans.
- QC data as described in Preparation Plan and implemented in contractors' work plan.

Feb 2005

- Release Draft 2002 NEI data for public review with documentation and message regarding information quality and review instructions to intended audience.
- Period opens for S/L/T comment and data revisions.
- QC data.

May - Aug 2005

- Resolve data discrepancies and incorporate revisions.
- QC data.

Oct - Nov 2005

- Finalize 2002 NEI data sets and documentation.
- Develop IQ message.

Dec 2005

- Release Final 2002 NEI data with documentation and message regarding information quality.

What is the purpose of the NEI peer review?

The peer review process must recognize and evaluate the interim and final products as necessary to achieve the NEI peer review objectives. The purpose of the NEI peer review is to evaluate the following:

- Choice of data estimation methods
 - Are the methods used to produce the NEI emission estimates technically sound?
 - Is consideration of alternative methods apparent?
- Documentation
 - Is the documentation adequate for user needs?
 - Are the data calculations transparent in terms of inputs/ output parameters, applied assumptions, and extrapolation techniques?
 - Are data distribution messages adequate?
- Data quality
 - Is the quality assurance process reasonable and appropriate?
 - Has EPA correctly/ appropriately defined uncertainty?
 - Does the data quality message adequately satisfy the EPA's IQ Guidelines?
 - Has the QA process been executed as planned and gross errors corrected?

How will the NEI peer review be conducted?

The peer review process for the 2002 NEI will be conducted in phases. The scope and intended outcome of each phase is summarized in the following table including a target schedule for completing the overall peer review process.

Phase I <i>June-Sep 2004</i>	<ul style="list-style-type: none">• Develop Peer Review Plan that recognizes the overall 2002 NEI process and product cycle time line.• Conduct internal EPA review for agreement on Peer Review Plan.• Solicit external expert review.• Inform external stakeholders.• Draft the peer review charge statements.• Solicit and confirm participation of peer reviewers.
Phase II <i>Oct 2004 - Jan 2005</i>	<ul style="list-style-type: none">• Peer review the data development process & outcomes that occurred through Sep 30, 2004.• Summarize comments and recommendations.• Develop plan of response.
Phase III <i>Feb - Dec 2005</i>	<ul style="list-style-type: none">• Peer review the data development process and documentation that occurred after Sep 2004 and through the final stage of product cycle.• Summarize comments and recommendations.• Develop report describing the outcome of final QA execution.• Develop plan of response.

The Peer review Plan that results from Phase I is expected to address operational issues including the following:

- Validate the need and scope of the peer review;
- Define the level of the peer review needed -that matches the impact and complexity of the NEI;
- Describe an appropriate and efficient way to implement the peer review given its scope, i.e., focus by major source sector with specific examples within sector, do integrated review in sector for criteria and HAP pollutants together;
- Specify the items that will be reviewed;
- The charge questions and how they were developed;
- Selection criteria for choosing peer reviewers that will best accommodate the implementation plan and focus, including the type of peer reviewer that is not appropriate;
- How conflict of interests issues will be identified and managed;
- How payments to peer reviewers, if required, will be handled, and anticipated costs and schedule for payment;
- Identify what NEI information and materials are appropriate and intended to be sent to each of the peer reviewers, this should identify information that is considered not readily available but which EPA may consider preparing in order to facilitate a robust review;
- The electronic format that will be used to help reviewers organize their comments and help the contractor summarize responses;
- Identify specific logistical needs of peer reviewers that may need to be closely managed;
- A prioritized and systematic process for organizing and replying to peer review comments and recommendations;
- How peer review results will be documented for the public record;
- Time schedules, constraints, and contingency plans;
- Costs and performance time lines throughout the approach.
- A tracking mechanism, such as a checklist, to ensure coordination and completion of progressive and related peer review events and successful orchestration of the overall peer review process.

Chapter 8: NEI Data Management System Flow

The data management system which we will use to facilitate compilation of 2002 NEI can be described in terms of three functional areas - collection; processing/ storage; and distribution. The purpose of this chapter is to describe the role anticipated for each functional area and the operational components involved and the sequence of data flow through each component. This view of the overall system flow will enable us to: determine points of necessary coordination and efficiencies, including chain of custody procedures; identify operation managers; anticipate necessary operation and maintenance of hardware/software including applied programming and routines; and identify the critical path elements necessary for successful data flow. Successful data flow is considered to include timely internal tracking and external distribution of reliable and documented 2002 NEI data.

Our strategy for complying with the Agency's data standards is implemented throughout the data management system. A summary of that strategy as it relates to each functional area of the system is included at the end of this chapter and specifically highlights by relevant standard, the form in which we will implement the standard during the data collection, the storage of the data, and then for data distribution.

The data flow through each functional area is described by source sector, e.g., point, nonpoint stationary, and mobile source sectors.

What are the operational components, and sequence of data flow for the Data Collection component?

For the data we expect to handle through the data collection functional area of the system, the following summary highlights by source sector:

- the data source and how that information is expected to flow through the applicable components,
- operations performed in this system area and by whom, and
- how the data moves out of this functional area into the next system area.

In addition, the primary objectives and endpoints for data flow through this functional area are described. The primary functions that will occur as part of data collection are:

- secure data handling;
- transformation of source formats to EFIG NIF format as necessary;
- confirmation of correct formatting for NIF and XML files received from S/L/Ts; and
- application of some basic data content and quality assurance routines.

Quality control and assurance routines, as well as chain of custody procedures will play an integral part of the data collection process.

Objectives

- Retrieve data for use in compiling the NEI.
- Use secure data handling procedures as necessary.
- Handle the volume of S/L/T files received in NIF or XML in efficient and timely manner.
- Maintain internal tracking system to identify chain-of-custody status for handling

S/L/T files.

- Determine if S/L/T files in NIF or XML are formatted correctly.
- Determine if other data source files if received in NIF, are formatted correctly.
- Interact rapidly with S/L/T data submitters to report if NIF or XML format errors or success.
- Transform into NIF those data collected in non-NIF formats.
- Determine a set of NIF and XML format checking and feedback procedures that may be further automated.
- Consider a flexible approach that may readily integrate node to node transfer of XML files over NEIEN.
- Identify those data collection operations that are part of our approach to comply with relevant Agency data standards and implement an internal status tracking mechanism.

Endpoints

- Primary data sources, formats, and expected flow are identified by source sector.
- Chain-of-custody/file handling procedures are implemented and convenient internal real-time status tracking is achieved.
- Error detection in NIF or XML files from S/L/T's receive swift response by us including direct interaction with data submitter to resolve problems.
- Majority of S/L/T files received in NIF and XML are made readily available to data processing steps.
- Majority of data files received in non-NIF format(s) are made readily available to data processing steps.
- Much of file handling and format checking procedures are automated.
- Data collection approach for complying with data standards is achieved.

What are the operational components, and sequence of data flow for the Data Processing and Storage component?

Objectives

- Automate as much of data processing steps as possible.
- Use secure data file handling procedures as necessary.
- Track origin of data and estimation methods used.
- Maintain internal tracking system by sector to identify status of data processing.
- Identify those data storage operations that are part of our approach to comply with

relevant Agency data standards and implement an internal mechanism to track status.

- Complete data processing steps and deliver resultant files within target schedule.
- Develop and apply procedure to ensure elimination of double-counting in point and nonpoint sector data.
- Identify critical schedule contingencies that may delay data flow.
- Specify file format for data compilers to use in delivering data results to EFIG, e.g., NIF V3(+) additional documentation fields = 'NEI format'.
- Establish automated routine for data compilers to apply to ensure that data files are correctly formatted prior to delivery to EFIG.
- For compiling sector updates to historical year data, initiate a process for contractors to start with existing year data stored and distributed centrally from NEI data system.
- Prescribe for data compilers the file format for delivering specific historical year data to EFIG.
- Proper file formatting is confirmed by our Oracle DBA prior to loading into NEI production database.

Endpoints

- Inventory sector data files are delivered by data development contractors in proper format for storage.
- All NEI criteria and HAP data, for current and historical years - are stored centrally in our NEI production database.
- 2002 NEI data is loaded into our system and readily available for distribution within targeted schedule.
- Data processing and storage approach for complying with data standards is achieved.

What are the operational components, and sequence of data flow for the Data Distribution component?

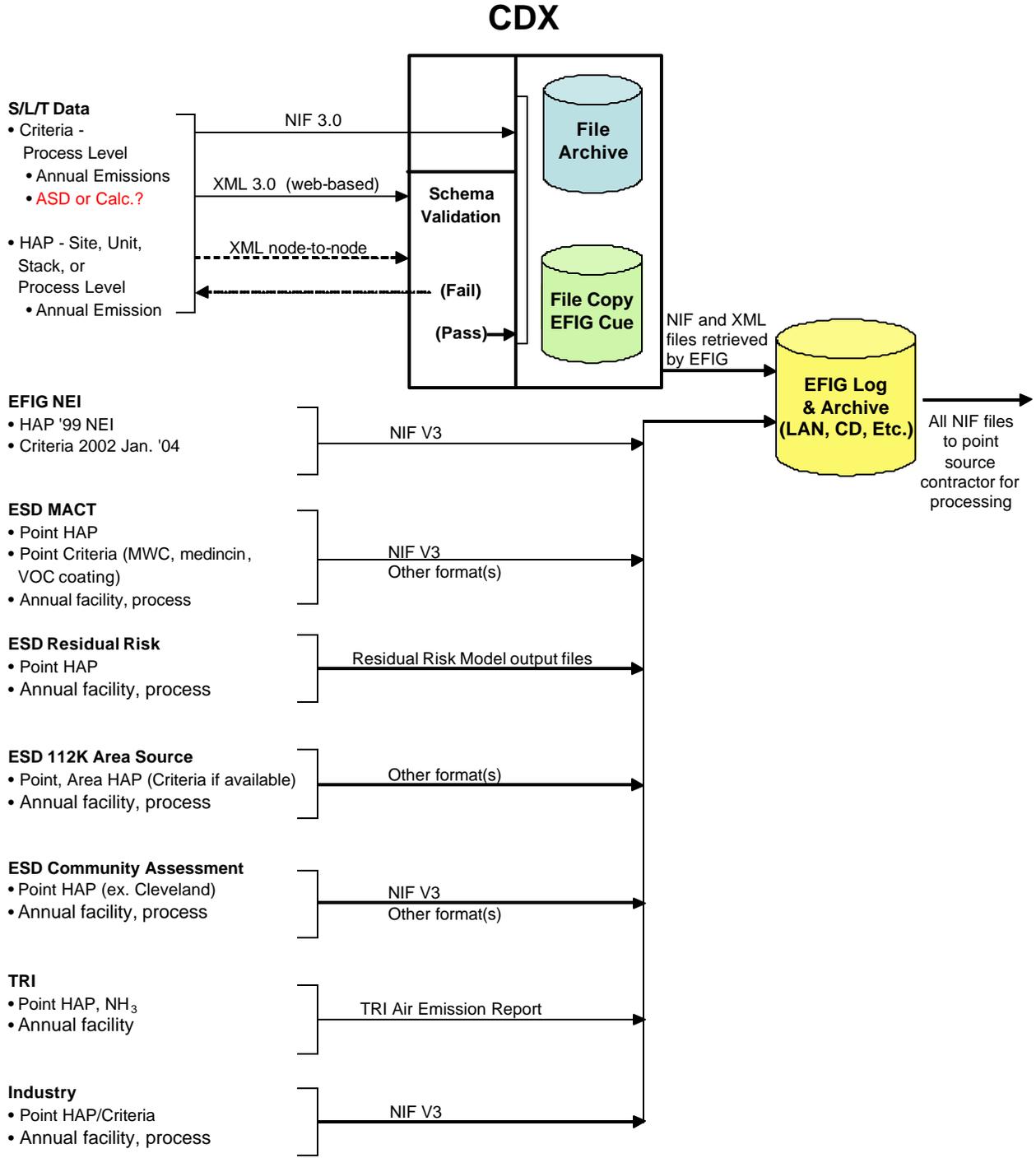
Objectives

- Identify all necessary distribution purposes, reports and files, formats, and target schedule.
- Identify, develop, and maintain queries and tools necessary to achieve desired criteria and HAP data reports.
- Identify those data distribution operations that are part of our approach to comply with relevant Agency data standards and implement an internal mechanism to track status.
- Establish and apply set of QC/QA checks as necessary to check for possible report errors prior to NEI system distribution, i.e., correct pollutant totals at specific data levels, by aggregate tier, sector, or sub-sector.
- Develop a method to determine evolving NEI data access needs by stakeholders.

Endpoints

- All NEI criteria and HAP data for current years and historical years are distributed centrally from our NEI data system.
- Data reports and files include the additional documentation elements intended and are available for distribution within targeted schedule.
- Data distribution approach for complying with data standards is achieved.
- Reports are distributed free of basic errors.
- Public access and internal EPA access to NEI criteria and HAP data is successful.

Collection – 2002 NEI – Criteria/HAP Point - Non-EGU

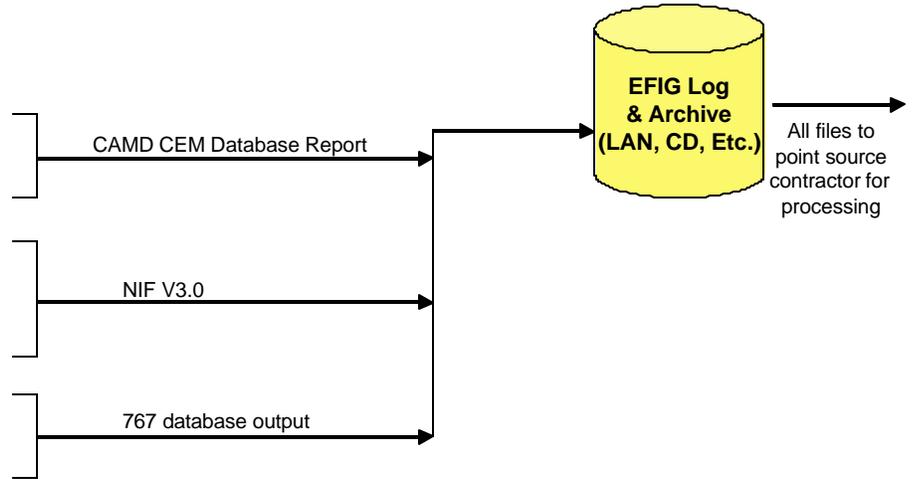


Collection – 2002 NEI – Criteria/HAP Point - EGU

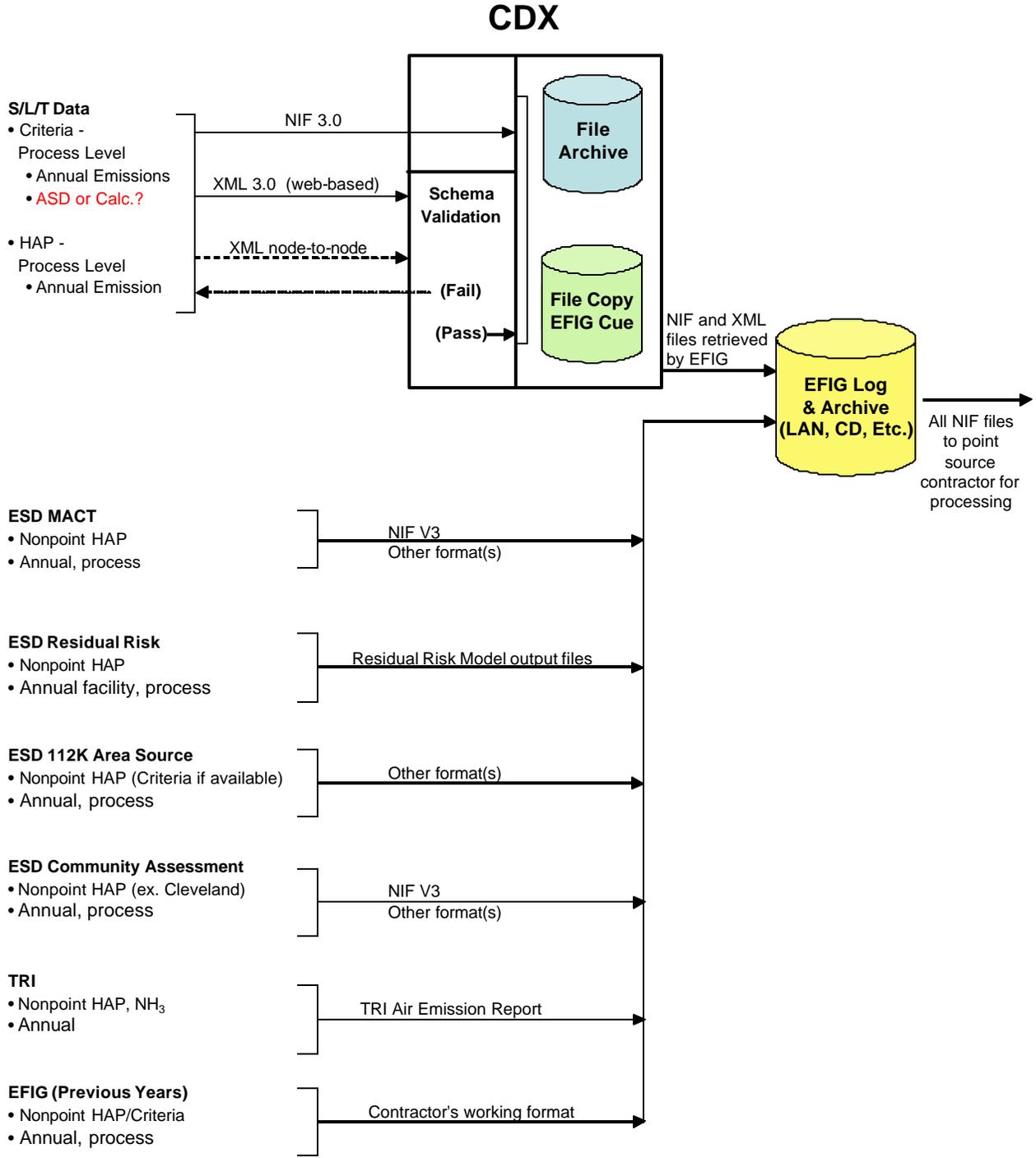
- CAMD CEM**
- Point Criteria
 - SO₂, NO_x

- ESD Hg Utility Model**
- Point HAP (Hg)
 - Annual facility ??

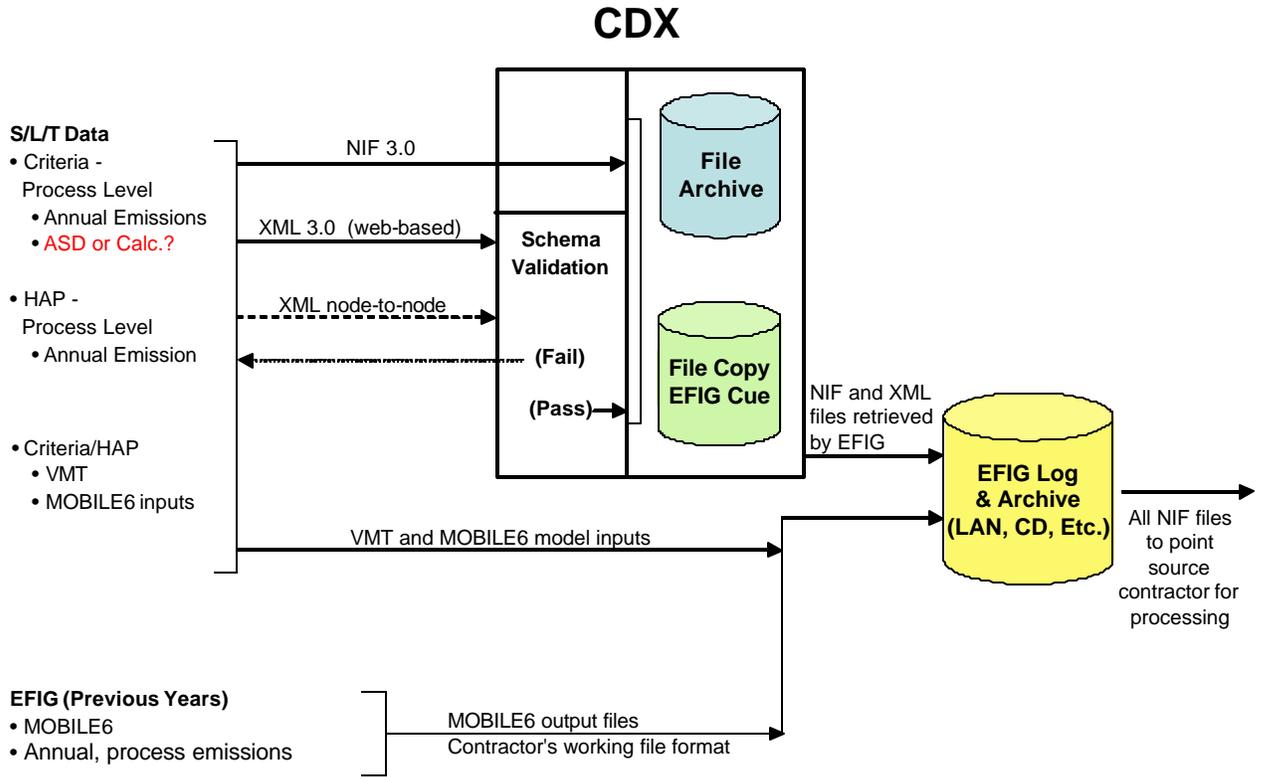
- DOE EIA 767**
- Point Criteria/HAP
 - Annual facility ??



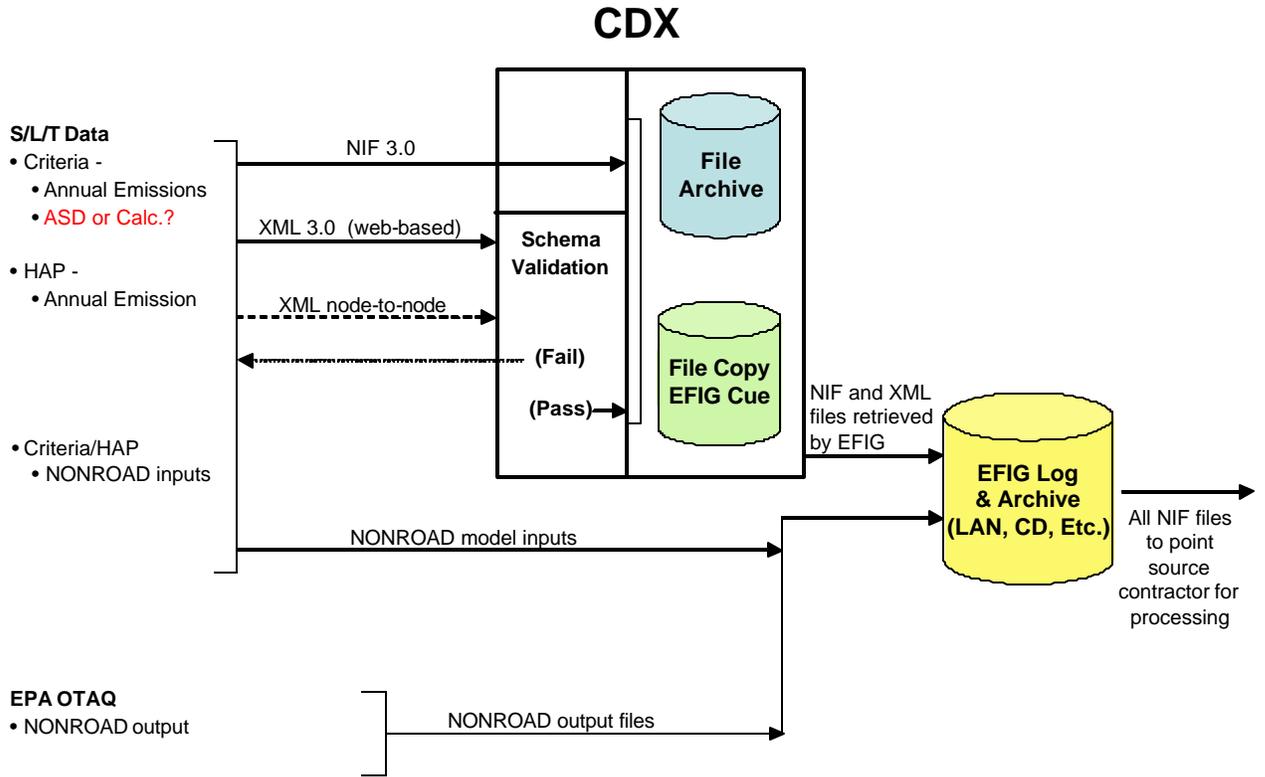
Collection – 2002 NEI – Criteria/HAP Nonpoint



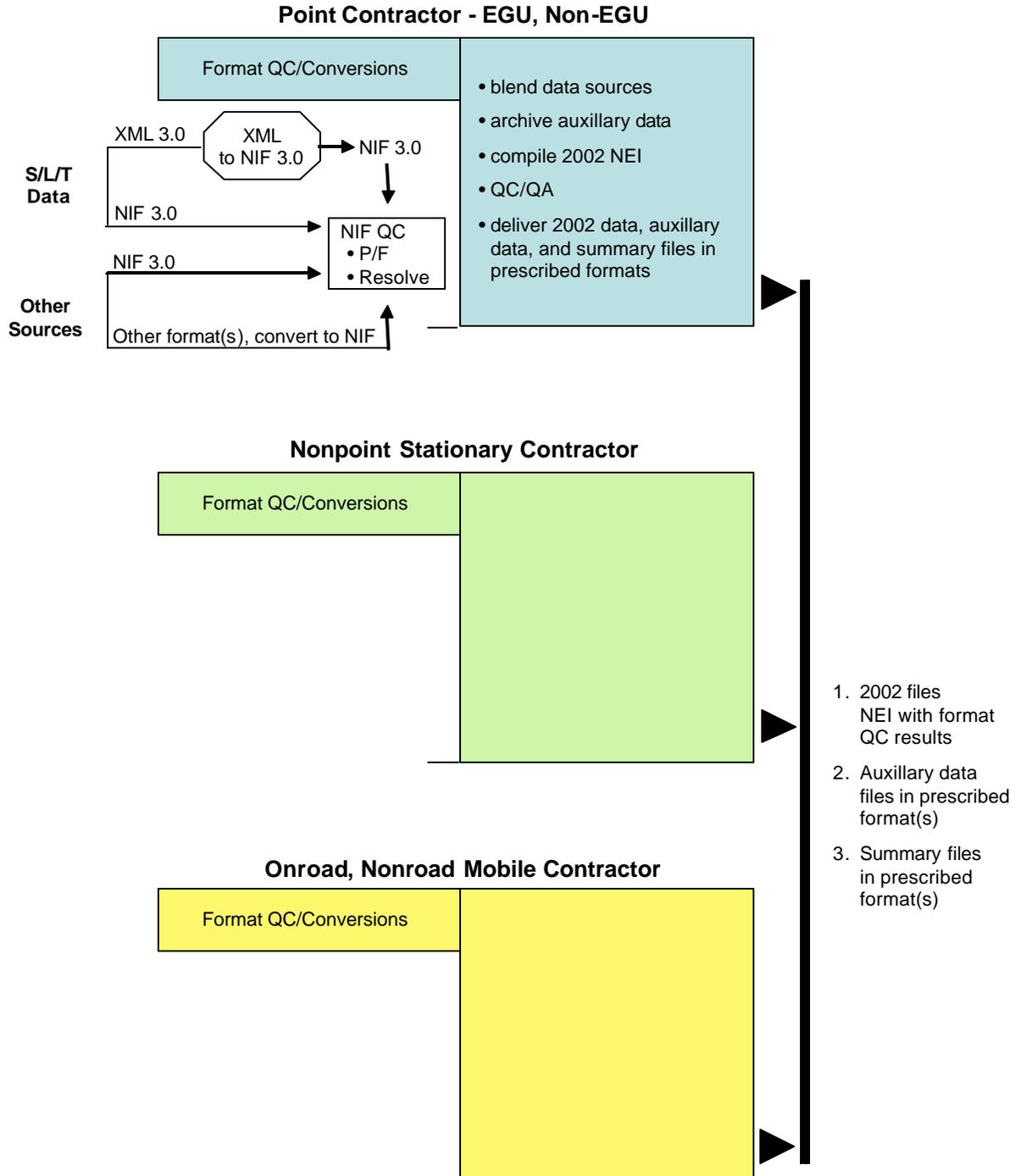
Collection – 2002 NEI – Criteria/HAP Onroad



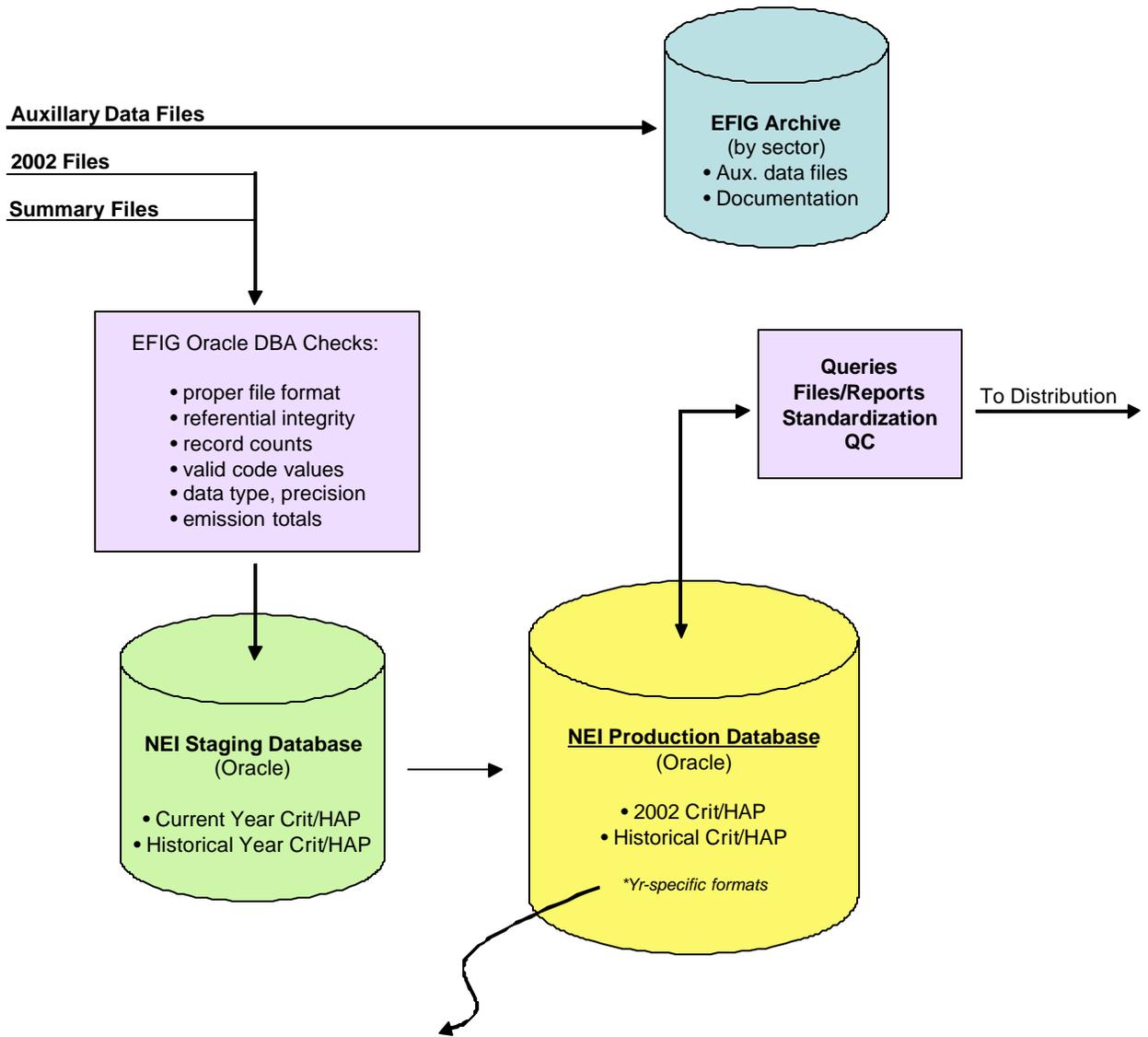
Collection – 2002 NEI – Criteria/HAP Nonroad



Compilation/Processing – 2002 NEI – Criteria/HAP



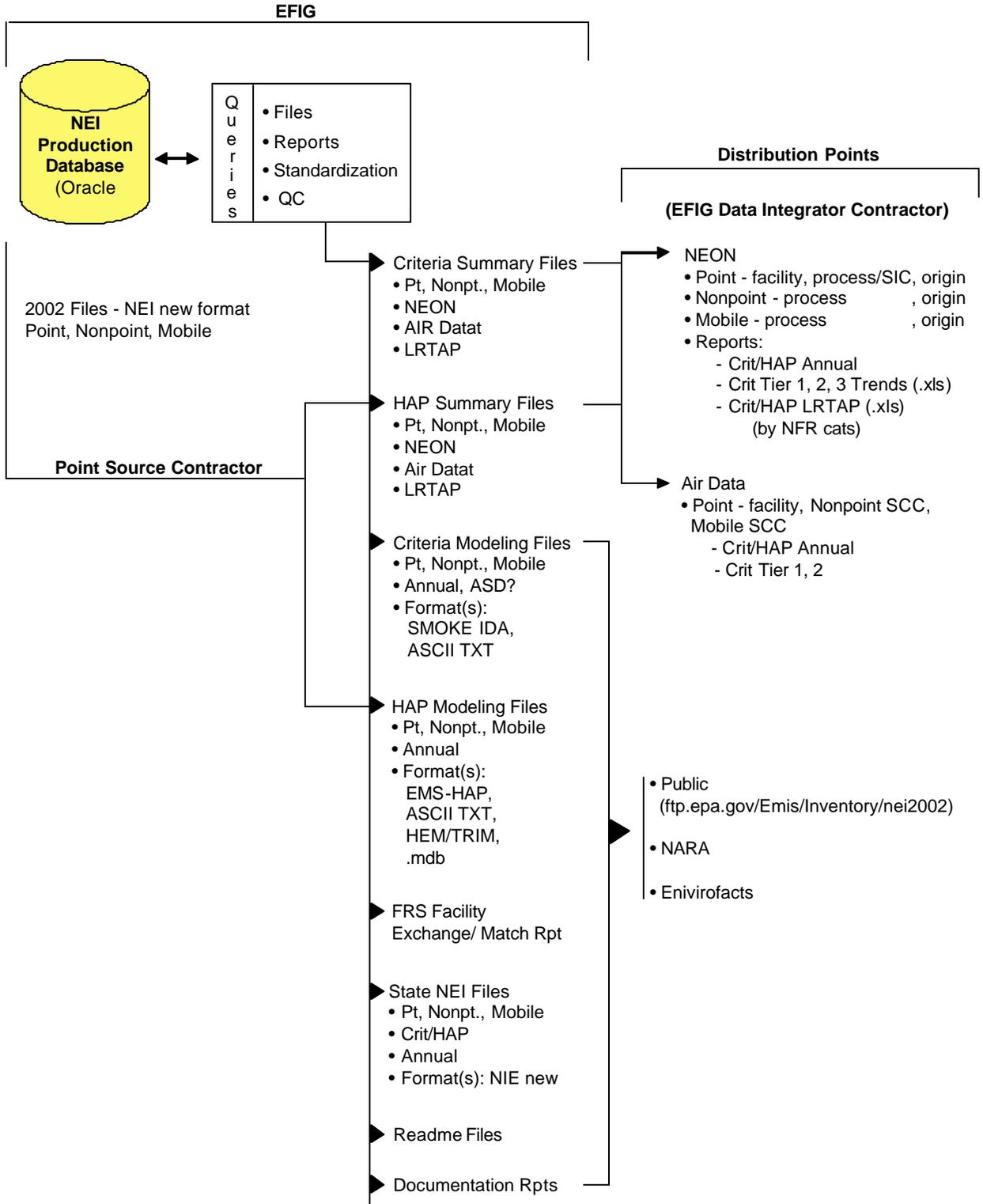
Storage – 2002 NEI – Criteria/HAP



** NEI Production Format(s) By Year*

Yr(s)	Crit	HAP	Format	Merged/ Separate
2002	✓	✓	NEI new	Merged
01-00	✓			N/A
1999	✓	✓	Crit NEI 90-99; NTI 99	Separate
98-97	✓		Crit NEI 90-99	N/A
96	✓	✓	Crit NEI 90-99; NTI 96	Separate
95-91	✓		Crit NEI 90-99	N/A
90	✓	✓	Crit NEI 90-99; NTI 90	Separate

Distribution – 2002 NEI – Criteria/HAP



Chapter 9: Special Issues

In this chapter, we describe special issues related to certain source categories.

Feedlot Dust

Because of the large uncertainty concerning the emission factor used previously to estimate feedlot PM emissions, we will not calculate emissions for PM from cattle feedlots in the 2002 NEI. Instead, we will rely on S/L/T's to provide locally derived emission estimates for this category.

Geogenic Dust and Agricultural Burning

Similar to Feedlot Dust, we will not calculate emissions for geogenic dust and agricultural burning. We may handle other categories similarly. We will note these in the Plan when they are identified.

Commercial Cooking

We are adding to the 2002 NEI emissions from commercial cooking. We will include emissions for VOC, PM and CO. The SCC's for this category are:

2302002100: Conveyorized Charbroiling (CO, VOC, PM10, PM25)
2302002200: Under-fired Charbroiling (CO, VOC, PM10, PM25)
2302003100: Flat Griddle Frying (CO, VOC, PM10, PM25)
2302003200: Clamshell Griddle Frying (VOC, PM10, PM25)
2302003000: Deep Fat Frying (VOC)

This inventory is adding the following approximate levels of emissions to the NEI:

CO : 33,000 tons/year
VOC : 11,600 tons/year
PM10 : 85,600 tons/year
PM25 : 71,100 tons/year

Although these represent moderate contributions at a national level, there are some counties that have substantial emissions. The maximum county emissions for all pollutants is in Los Angeles county, California.

Biogenic Emission Inventories

The Consolidated Emissions Reporting Rule (CERR) indicates that States should submit county-level biogenic inventories for 2002. This requirement is perceived as difficult to meet for many States and not useful for air quality modeling purposes, for which biogenic emissions are usually calculated with grid-based approaches using the BEIS2, BEIS3, or other biogenic models.

To help the States meet the CERR biogenics requirement, EFIG has prepared estimated 2002 county-level biogenic emissions using BEIS3 for the entire U.S. and has posted the data on our web site (www.epa.gov/ttn/chief/net/2002inventory.html#biogenic) for access by the States. To meet the CERR biogenics requirements, States have the option to use the EPA-provided biogenic emissions estimate.

To do this, States would need to inform EFIG in writing that the estimates are acceptable to the States for the purposes of the CERR. This would be in lieu of submitting county level biogenic inventories. Notification should come from the CDX account holder who would otherwise submit data to EPA via CDX. Notification should be addressed to:

Phil Lorang
US/EPA OAQPS (D205-01)
RTP, NC 27711
e-mail: lorang.phil@epa.gov

States may also choose to create new or revised estimates and submit these instead.

If a State accepts the estimates, States are not endorsing the emissions values for all purposes, only for submittal to the CERR. We will use the CERR data to provide public information about the biogenic emissions, but we will not use it for modeling.

The approach described below has some limitations of which we are already aware. We are not concerned about these limitations because they primarily affect modeling issues. This approach will provide ballpark estimates that are suitable for general information purposes. The data can also be used as a reference point for rough checking of more detailed 2002 estimates that will be used for modeling.

Below is a description of the approach that we will use to generate these estimates.

The approach EPA is using to create the county-total, annual 2002 biogenic emissions estimates is as follows:

* Run SMOKE BEIS3 (version 3.11) with EPA's final 2001 MM5 based 36-km national, annual meteorology data. This implies the use of BELD3 data at 36-km resolution.

* SMOKE will compute daily county-total biogenic emissions values, which will then be summed for the year for each county.

* We will calculate the emissions in tons/year for the following CMAQ CB-IV model species:

- o NO
- o ALD2
- o ETH
- o FORM
- o ISOP
- o OLE
- o PAR
- o TOL
- o XYL
- o NR
- o TERPB

* The emissions will also be calculated for total VOC mass as:

- o VOC = ALD2 + ETH + FORM + ISOP + OLE + PAR + TOL + XYL + NR

Note that there are two open issues:

(1) The format of the data has been proposed to be in MS Excel format. We have asked the States to notify us if this is not acceptable, and if not, to propose an alternative format.

(2) We are planning to generate annual totals by county. We have asked the

States to notify us If more detailed (e.g., monthly totals) or State totals are also needed, and if so, to indicate what this is needed for. This information will help us in our efforts to accommodate such requests.

Appendix A: Reporting of HAP Groups

Section 112(b) of the CAA currently identifies a list of 188 pollutants as HAPs (<http://www.epa.gov/ttn/uatw/pollsour.html>). EPA's Unified Air Toxics Web Site (UATW) presents more information on HAPs, their effects, and programs to reduce HAPs (<http://www.epa.gov/ttn/uatw/basicfac.html>).

In addition to numerous specific chemical species and compounds, the list of 188 HAPs includes seventeen compound groups (e.g., individual metals and their compounds, polycyclic organic matter (POM), and glycol ethers). Many of the uses of the NEI depend upon data for individual compounds within these groups rather than aggregated data for the groups. The 1999 NEI for HAPs contains more than 500 individual compounds.

One of the major issues encountered in compiling the 1999 NEI was the reporting of information on HAP groups. If aggregated emissions were reported for the HAP groups, various assumptions had to be made about the HAP compound group which introduced uncertainties in the use of the 1999 NEI data for air quality modeling and risk characterization. Individual compounds in some of the HAP groups (e.g., mercury compounds) have substantially different fate and transport characteristics which strongly affect the modeling results. For example, in order to model the atmospheric deposition of mercury, we need to include in the NEI three different species of mercury because they do not all transport or react the same once they are in the atmosphere. Likewise, HAP exposures and associated human health effects can vary substantially among the specific compounds within many of the HAP groups. For example, some compounds within the polycyclic organic matter (POM) group are relatively non-toxic, while others are highly potent carcinogens.

To reduce uncertainties and potential overestimation of risk in future NEI-based assessments that might result from default disaggregation of HAP groups into individual compound emissions, we will request that organizations providing data for the 2002 NEI report emissions for specific compounds, both for individual HAP species and for HAPs within compound groups. If emissions of individual pollutants within HAP groups are not reported, we will accept aggregated compound group emissions. However, we will have to use simplifying assumptions regarding speciation in order to use these data as inputs to models.

In "Questions and Answers for the 1999 NTI," (<http://www.epa.gov/ttn/chief/eidocs/ntiq&a.pdf>) recommendations for reporting data for specific groups of compounds are summarized in a hierarchy of most preferred method to least preferred. For pollutant groups, only one reporting strategy per HAP group per source should be used to avoid potential overestimation of emission levels and risk. The preferred methodology for all compound groups is to report emissions and associated CAS numbers of all individual species (e.g., report emissions and associated CAS numbers of arsenic oxide, lead arsenate, etc., rather than emissions of arsenic compounds as a whole.).

What are our recommendations for reporting specific Groups of Compounds?

Our recommendations are summarized in the following table:

Preferred Method	Alternate Method
Metals	
<p>If individual metal compounds can be reported:</p> <ul style="list-style-type: none"> • Report all individual compounds as the mass of the total compound, not just the metal within the compound 	<p>If individual metal compounds cannot be reported:</p> <ul style="list-style-type: none"> • Report the mass of emissions of the metal, not of the entire metal compound • Chromium - Separate chromium compounds into hexavalent chromium (CAS #18540299), trivalent (CAS #1606583) and other chromium. • Lead - Separate lead compounds into organic and inorganic. • Mercury - Separate mercury compounds into particulate divalent, gaseous elemental, and gaseous divalent. • Nickel - Separate nickel compounds into nickel subsulfide (CAS #12035722) and other nickel (CAS #7440020)
POM	
<p>If emissions of all individual PAH can be reported:</p> <ul style="list-style-type: none"> • Report emissions of the individual compounds rather than emissions of total polycyclic aromatic hydrocarbons (PAH) or total POM. • The most important PAH compounds to report individually are the 7-PAH compounds listed in the table below. • Report all other individual POM compounds for which cancer assessments are available. 	<p>If emissions of all individual PAH cannot be reported:</p> <ul style="list-style-type: none"> • Report 7-PAH as a subgroup. • Since naphthalene is listed individually as a HAP, do not include any individually-reported naphthalene as total POM.
Dioxins/Furans	
<p>If emissions of individual CDD and CDF congeners can be reported:</p> <ul style="list-style-type: none"> • Report mass emissions and associated CAS numbers of all individual congeners of both chlorinated dibenzodioxins (CDDs) and chlorinated dibenzofurans (CDFs). 	<p>If emissions of individual CDD and CDF congeners cannot be reported:</p> <ul style="list-style-type: none"> • Report dioxins and furans as 2,3,7,8-tetrachlorodibenzodioxin (TCDD) toxic equivalents (TEQ). • If dioxin TEQ estimates included, use the correct NEI code indicating which TEQ is used. • Dioxin TEQ can be reported either using WHO-toxic equivalent factors or I (NATO) toxic equivalent factors.

Preferred Method	Alternate Method
Xylenes, Cresols, and Glycol Ethers	
<p>If individual emissions of xylenes or cresols can be reported:</p> <ul style="list-style-type: none"> Report mass emissions and associated CAS numbers for individual xylene and cresol isomers and glycol ethers. 	<p>If individual emissions of xylenes or cresols cannot be reported:</p> <ul style="list-style-type: none"> Report total emissions of xylenes or cresols as a group under "xylenes (mixture of o, m, and p isomers)" (CAS #1330207) or "cresols/cresylic acids". For glycol ethers, use EPA guidance on glycol ethers to identify compounds that are glycol ethers. Ethylene glycol butyl ether (EGBE) has been delisted as a HAP and should not be included in the glycol ethers compound group.

What are the most important PAH compounds to report individually?

As mentioned in the table above, the most important PAH compounds to report individually are the 7-PAH compounds, listed in the table below.

7-PAH	POM Compounds for which we have cancer assessments - includes 7-PAH
Benz(a)anthracene Benzo(a)pyrene Benzo(b)fluoranthene Benzo(k)fluoranthene Chrysene Dibenz(a, h)anthracene Indeno(1,2,3-cd)pyrene	Carbazole Dibenz[a,h]acridine Dibenz[a,j]acridine 7H-Dibenzo[c,g]carbazole Dibenzo[a,e]pyrene Dibenzo[a,i]pyrene Dibenzo[a,l]pyrene 7,12-Dimethylbenz[a]anthracene 1,6-Dinitropyrene 1,8-Dinitropyrene 3-Methylcholanthrene 5-Methylchrysene 5-Nitroacenaphthene 6-Nitrochrysene 2-Nitrofluorene 2-Nitrofluorene 1-Nitropyrene 4-Nitropyrene

Appendix B: Source Categories in the NEI

The NEI contains emission estimates for point, nonpoint, and mobile onroad and nonroad source sectors. We describe below the types of sources that will be included in each of the sectors. Also, we define some important related terms.

What sources or source categories are included in each of the sectors (Point, Nonpoint, Onroad, Nonroad) of the NEI?

Point - We will include in this sector sources for which the specific location is identified.

For HAPs, point sources include major and area sources as defined in section 112 of the CAA (and defined later in this appendix). The NEI for HAPs includes a data field that identifies emissions as major, area/other, or mobile.

For Criteria pollutants, as defined in the CERR, point sources include:

- (a) For PM₁₀, PM_{2.5}, ammonia, sulfur oxides, VOC, and nitrogen oxides, any plant that emits at least 100 tons per year of any of these pollutants, and
- (b) For carbon monoxide, any plant that emits at least 1000 tons per year.

For dispersion and exposure modeling, it is preferable to model all stationary sources as point sources. To the extent possible, for HAPs and Criteria, we encourage organizations to provide facility-specific emissions data for all point sources, regardless of size, where they are already included in the S/L/T emission inventory. At a minimum, we request that all major HAP sources, including both MACT and non-MACT sources be reported as point sources in the 2002 NEI. We also encourage organizations to report facility-specific emissions data for all MACT source categories regardless of whether a facility is classified as major or area. The NEI retains all facility-specific data as point sources, regardless of the magnitude of the emissions, to facilitate more comprehensive assessments. Therefore, no reporting thresholds exist for point sources in the NEI.

If an agency or tribe submits a merged HAP and criteria inventory, then they should include all pollutants for each facility in the submission. For example, if an agency includes HAP data for a crematory in a merged HAP and criteria submission, then it should also report criteria emissions for the same facility.

Nonpoint - For HAPs, we will include in this sector area sources and other stationary sources that may be more appropriately addressed by other programs rather than through regulations developed under sections 112 or 129 of the CAA. Nonpoint sources emitting HAPs in the NEI also include area sources that are not identified as point sources. Nonpoint and mobile source emissions are reported as aggregates at the county level in the NEI.

For Criteria pollutants, we will include all stationary sources not included in the Point sector.

Onroad - Onroad mobile sources include all onroad vehicles such as cars, trucks, and buses. The NEI includes aggregated emission estimates at the county level for Onroad and Nonroad mobile sources.

Nonroad - Nonroad mobile sources include non-road 2- and 4- stroke and diesel engines, off road vehicles such as construction equipment, aircraft, locomotives, and commercial marine vessels.

For HAPs, what are major sources, area sources and other sources?

Major sources for HAPs - As defined by section 112 of the CAA, these are stationary sources that emit or have the potential to emit 10 tons per year or more of any listed HAP or a 25 tons per year or more of a combination of listed HAPs (<http://www.epa.gov/ttn/uatw/pollsour.html>). When estimates of potential emissions are not available, the NEI identifies point sources emitting HAPs as major based on reported emissions being at or above 10/25 tons per year. Examples of major sources include electric utility plants, chemical plants, steel mills, oil refineries, and hazardous waste incinerators.

Area sources for HAPs - As defined by section 112 of the CAA, these are stationary sources that emit or have the potential to emit less than 10 tons per year of a single HAP or less than 25 tons per year of a combination of HAPs (<http://www.epa.gov/ttn/uatw/pollsour.html>). When estimates of potential emissions are not available, reported emissions are used. The NEI includes facility data for some area sources and aggregated nonpoint emission estimates at the county level for the remaining area sources. Area sources are regulated under toxics provisions in the CAA. Examples of area sources include neighborhood dry cleaners and gas stations. Though emissions from individual area sources are often relatively small, collectively their emissions can be of concern particularly where large numbers of sources are located in heavily populated areas.

Other stationary sources for HAPs - These are sources that may be more appropriately addressed by other programs rather than through regulations developed under air toxics provisions such as sections 112 or 129 in the CAA. Examples of other sources include wildfires and prescribed burning whose emissions are being addressed through the burning policy agreed to by EPA and the U.S. Department of Agriculture (USDA). The NEI includes aggregated nonpoint emission estimates at the county level for these other sources.

Which Nonpoint Sector categories (SCC level) will EPA calculate emissions for?

EPA will calculate emissions for the following Nonpoint categories:

Source Category	Source Classification Code	MACT Category	Emissions for Puerto Rico and U.S. Virgin Islands
Agricultural Tilling	2801000003		
Aviation Gasoline Distribution: Stage I	2501080050	601	
Aviation Gasoline Distribution: Stage II	2501080100		
Cotton Ginning	2801000000		
Fluorescent Lamp Recycling	31301200		
General Laboratory Activities	31503001		Puerto Rico and U.S. Virgin Islands
Lamp Breakage	2861000000		U.S. Virgin Islands
Open Burning - Scrap Tires	2830000000		
Open Burning - Household Waste	2610030000		
Open Burning - Yard Waste - Leaves	2610000100		
Open Burning - Yard Waste - Brush	2610000400		
Publicly Owned Treatment Works	2630020000	803	Puerto Rico and U.S. Virgin Islands
Residential Heating: Anthracite Coal	2104001000		
Residential Heating: Bituminous & Lignite	2104002000		
Residential Heating: Distillate Oil	2104004000		
Residential Heating: Natural Gas	2104006000		
Residential Heating: Wood/Wood Residue	2.1040800012e+69		
Residential Construction	2311010000		
Surface Coating: Architectural	2401001000		Puerto Rico and U.S. Virgin Islands

Source Category	Source Classification Code	MACT Category	Emissions for Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Adhesive and Sealants	2460600000		Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Automotive Aftermarket Products	2460400000		Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Coatings and Related Products	2460500000		Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Household Products	2460200000		Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Miscellaneous	2460900000		Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Personal Care Products	2460100000		Puerto Rico and U.S. Virgin Islands
Commercial and Consumer Product Usage: Regulated Products	2460800000		Puerto Rico and U.S. Virgin Islands

Which Nonpoint Sector categories (SCC level) will EPA carry forward emissions from the Version 3 of the 1999 NEI?

EPA will carry forward 1999 emissions for the following Nonpoint categories:

Source Category	SCC	Rationale for Carrying Forward	MACT Category
Agricultural Burning	2801500000		
Animal Husbandry	2.805020e+59	2002 Census of Agriculture activity data will not be available until Spring 2004	
Asphalt Paving	2461021000	2002 consumption estimates not available until Fall 2003	
Asphalt Roofing Materials Manufacturing	2306010000	ESD estimates	418
Asphalt Concrete Manufacturing - Rotary Dryer at a Conventional Plant	305002011	State/Local/Tribal agency estimates	
Autobody Refinishing Paint Application	2401005000	County Business Patterns (CBP) data not available until April 2004 ²	
Beef Cattle Feedlots	2805001000	2002 Census of Agriculture activity data will not be available until Spring 2004	
Chromic Acid Anodizing	2309100050	CBP not available until April 2004	1607
Commercial/Institutional Fuel Combustion: Anthracite Coal (Total: All Boiler Types)	2103001000	CBP not available until April 2004	107
Commercial/Institutional Fuel Combustion: Bituminous/ Subbituminous Coal (Total: All Boiler Types)	2103002000	CBP not available until April 2004	107
Commercial/Institutional Fuel Combustion: Distillate Oil (Total: Boilers and IC Engines)	2103004000	CBP not available until April 2004	107
Commercial/Institutional Fuel Combustion: Residual Oil (Total: All Boiler Types)	2103005000	CBP not available until April 2004	107
Commercial/Institutional Fuel Combustion: Natural Gas (Total: Boilers and IC Engines)	2103006000	CBP not available until April 2004	107
Cremation - Human	2810060100	2002 percentage of cremations data from CANA not available	1807-2
Cremation - Animal	2810060200	CBP not available until April 2004	1807-2
Decorative Chromium Electroplating	2309100030	CBP not available until April 2004	1610
Dental Preparation and Use	315025001	CBP not available until April 2004	
Drum and Barrel Reclamation	2461160000	No change from 1999 national activity	
Fertilizers	2.801700e+99	2002 fertilizer usage data not available	
Flexible Polyurethane Foam Fabrication Operations	301018801	ESD estimates	1341
Gasoline Distribution Stage I - Bulk terminals, Plants and Pipelines	2501050120	CBP not available until April 2004	
Gasoline Distribution Stage I - Service Station Storage Tanks	2501060053	State gasoline sales from EIA not available until Fall 2003	601
Gasoline Distribution Stage II	2501060100	State gasoline sales from EIA not available until Fall 2003	

Source Category	SCC	Rationale for Carrying Forward	MACT Category
Grain Elevators: Terminal	302005121	2002 <i>Agricultural Chemical Usage</i> reports not available	
Graphic Arts	2425000000	CBP not available until April 2004	
Halogenated Solvent Cleaners	2415000000	CBP not available until April 2004	1614
Hard Chromium Electroplating	2309100010	CBP not available until April 2004	1615
Hospital Sterilization	2850000010	CBP not available until April 2004	1644
Industrial Fuel Combustion: Bituminous/Subbituminous Coal (Total: All Boiler Types)	2102002000	CBP not available until April 2004	107
Industrial Fuel Combustion: Distillate Oil (Total: Boilers and IC Engines)	2102004000	CBP not available until April 2004	107
Industrial Fuel Combustion: Residual Oil (Total: All Boiler Types)	2102005000	CBP not available until April 2004	107
Industrial Fuel Combustion: Natural Gas (Total: Boilers and IC Engines)	2102006000	CBP not available until April 2004	107
Industrial Fuel Combustion: Natural Gas (All Boiler Types)	2102006001	CBP not available until April 2004	107
Industrial Fuel Combustion: Natural Gas (All IC Engine Types)	2102006002	CBP not available until April 2004	105
Industrial Fuel Combustion: Liquefied Petroleum Gas (LPG) (Total: All Boiler Types)	2102007000	CBP not available until April 2004	107
Institutional/Commercial Heating: Public Owned Treatment Works (POTW) Digester Gas	103007011	ESD estimates	107
Mining and Quarrying	2325000000	2002 EIA coal mined data not available.	
Miscellaneous Organic Chemical Processes ³	301999991	ESD estimates	1640
Natural Gas Transmission and Storage	310002991	ESD estimates	504
Non-Residential Construction	2311020000	2002 Covered Employment Wages data from BLS not available until Fall 2003	
Oil and Natural Gas Production	2310000000, 2310010000, 2310020000, 2310030000	ESD estimates	501
Open Burning - Land Clearing Debris	2610000500	2002 Federal Highway Administration (FHWA) State expenditure data not available	
Paint Stripping Operations	682400591	ESD estimates	1621
Perchloroethylene Dry Cleaning	2420000055	CBP not available until April 2004	1643
Pesticide Application	2461850000	2002 Pesticide sales from Census not available until Fall 2003	
Refractories Manufacturing	2305000000	ESD estimates	406
Roadway Construction	2311030000	2002 FHWA data not available	
Steel Pickling with Hydrochloric Acid (HCL)	2303000000	ESD estimates	310
Structure Fires	2810030000	National number of fires for 2002 will not be available from National Fire Data Center (NFDC) until Fall 2003	
Surface Coating: Traffic Markings	2401008000	2002 Federal highway money data are not available from FHWA until 11/03	
Surface Coating: Industrial Maintenance	2401100000	CBP not available until April 2004	
Swimming Pools	2862000000	CBP not available until April 2004	

¹ EPA did not carry forward 1999 criteria pollutant emissions for these point source categories because it is unclear that these emissions should be incorporated into the nonpoint source inventory. To estimate nonpoint source emissions for these categories, EPA will first estimate total (point and nonpoint) year 2002 emissions activity. Then EPA will estimate 2002 nonpoint source emissions activity by subtracting year 2002 point source emissions activity from 2002 total emissions activity.

² CBP = County Business Patterns, U.S. Census Bureau, <http://www.census.gov/epcd/cbp/view/cbpview.html>.

³ For miscellaneous organic chemical processes, MACT code 1640 was assigned to the original set of data. Subsequently, the category was split between MACT codes 1641 and 1642 and MACT code 1640 was removed from the NEI Input Format (NIF) code table. For the 1999 emissions included in the 2002 NEI, it is unclear how to divide the emissions between MACT codes 1641 and 1642; consequently, the original MACT code 1640 has been maintained in the 2002 NEI.

Appendix C: Data Elements in the NEI

We have included below a list of data elements, extracted from the NIF 3.0 guidance (www.epa.gov/ttn/chief/nif/index.html#ver3), that we will collect or augment in the NEI. This list is the subset of the NIF data elements that are mandatory or necessary. Mandatory data elements are those that S/L/Ts must complete in their data submittal in order for us to incorporate the data submittal into the NEI. Necessary data elements are those that we will complete if the S/L/Ts do not include them in their data submittal. We include in the NIF guidance additional data elements that we are seeking from S/L/T agencies.

CAPs	HAPs	Data Element	Data Definitions
Site Level			
M	M	RECORD TYPE	A code that identifies the type of Record
M	M	STATE AND COUNTY FIPS CODE	The FIPS code for the state and county.
M	M	STATE FACILITY IDENTIFIER	Unique ID number used by a state/loca/tribal agency to identify a facility.
	M	FACILITY CATEGORY	Indicates if HAP emitting facility is MAJOR or AREA.
N	N	SIC PRIMARY	Standard Industrial Classification code system.
M	M	NAICS PRIMARY	North American Industry Classification code.
M	M	FACILITY NAME	The name of the facility.
M	M	LOCATION ADDRESS	Physical location of the front door / main entrance of the facility site.
M	M	CITY	The name of the city.
M	M	STATE	State abbreviation.
M	M	ZIPCODE	The U.S. Postal Service zip code.
M	M	TRIBAL CODE	Codes that represent American Indian tribes and Alaskan Native entities.
Unit Level			
M	*M*	EMISSION UNIT ID	Unique ID reported consistently over time by state/ local/ tribal agency.
Process Level			
M	*M*	PROCESS ID	Unique ID reported consistently over time by state/ local/ tribal agency.
M		SCC	EPA Source Category Code for Point Sources.
	N	PROCESS MACT CODE	Maximum Achievable Control Technology for HAP regulated sources
N		WINTER THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		SPRING THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		SUMMER THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		FALL THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		ANNUAL AVG DAYS PER	Average number of days per week an emission process is active within

		WEEK	year.
N		ANNUAL AVG WEEKS PER YEAR	Average number of weeks per year an emission process is active.
N		ANNUAL AVG HOURS PER DAY	Average number of hours per day an emission process is active within year.
N		ANNUAL AVG HOURS PER YEAR	Average number of hours per year an emission process is active.
N		HEAT CONTENT	The heat content of a fuel in million BTU's per Ton of coal, 1000 Gals of oil, or million SCF gas.
N		SULFUR CONTENT	The sulfur content of a fuel (mass percent).
N		ASH CONTENT	The ash content of a fuel (mass percent).
	N	PROCESS MACT COMPLIANCE STATUS	Major/Area classification and status under CAAA Sections 112&129.
Period Level			
M	M	<i>START DATE</i>	Start date of the period in which reported emissions occur.
M	M	<i>END DATE</i>	End date of the period in which reported emissions occur.
M	M	<i>POLLUTANT CODE</i>	Pollutant Code
	M	<i>EMISSION RELEASE POINT ID</i>	State/ local/ tribal ID for point / location where emissions are released to ambient air.
M	M	<i>START DATE</i>	Start date of the period in which reported emissions occur.
M	M	<i>END DATE</i>	End date of the period in which reported emissions occur.
Emission Level			
M	M	EMISSION NUMERIC VALUE	Numeric value of emission.
M	M	EMISSION UNIT NUMERATOR	Unit of measure for reported emissions value. If criteria emissions, report unit: annual = TON; average season day = LBS.
M	M	<i>EMISSION TYPE</i>	Code describing temporal designation of emissions reported, i.e., Entire Period, Average Weekday, etc.
N		FACTOR NUMERIC VALUE	The numeric value of the emission factor.
N		FACTOR UNIT NUMERATOR	Unit of measure for emission factor numerator.
N		FACTOR UNIT DENOMINATOR	Unit of measure for emission factor denominator.
N		MATERIAL	Material code for material processed.
N		MATERIAL I/O	A descriptor indicating whether material is used or produced.
N		RULE EFFECTIVENESS	Measure of the percent effectiveness of the control strategy.
	M	HAP EMISSIONS PERFORMANCE LEVEL	Code that represents the performance level, or operating scenario, for the HAP emissions reported.
	M	CONTROL STATUS	Indicates if reported emissions are controlled or uncontrolled.

	M	EMISSION DATA LEVEL	The level of site emissions disaggregation reported on the emission record.
Control Level			
N	N	PRIMARY PCT CONTROL EFFICIENCY	The percent effectiveness of primary control device.
N	N	PCT CAPTURE EFFICIENCY	Numeric value for percentage capture efficiency of control system.
N	N	TOTAL CAPTURE CONTROL EFFICIENCY	Collective (aggregate) value for all controls.
M	M	PRIMARY DEVICE TYPE CODE	The primary type of control equipment used.
Release Point Level			
M	M	EMISSION RELEASE POINT ID	State/ local/ tribal ID for point / location where emissions are released to ambient air.
M	M	EMISSION RELEASE POINT TYPE	The code for physical configuration of the release point.
N	N	STACK HEIGHT	The height (in feet) of a stack.
N	N	STACK DIAMETER	The diameter (in feet) of a stack.
N	N	EXIT GAS TEMPERATURE	The temperature of an exit gas stream (degree Fahrenheit).
N	N	EXIT GAS VELOCITY	The velocity of an exit gas stream (feet per second).
N	N	EXIT GAS FLOW RATE	Numeric value of stack gas flow rate in actual cubic feet per second.
M	M	X COORDINATE	Longitude measure in decimal degrees of the angular distance on a meridian east or west of the prime meridian. Negative (-) data point for N America. Include (-) sign, Ex. -123.234561; or UTM Easting in kilometers.
M	M	Y COORDINATE	Latitude measure in decimal degrees of the angular distance on a meridian north or south of the equator. Positive (+) data point for N America. Include (+) sign, Ex. +78.123456; or UTM Northing in kilometers.
M	M	UTM ZONE	Zone number in UTM coordinate system.
M	M	XY COORDINATE TYPE	Type of coordinate system used, eg., LATLON or UTM.
M	M	HORIZONTAL COLLECTION METHOD CODE	Method used to determine the lat and lon coordinates for a point on the earth.
M	M	HORIZONTAL ACCURACY MEASURE	The measure of accuracy (in meters) of the lat and lon coordinates.
M	M	HORIZONTAL REFERENCE DATUM CODE	Code that represents the reference datum used to determine the lat / lon coordinates.
M	M	REFERENCE POINT CODE	The code that represents the place for which geographic coordinates were established. Code value should be 106 (e.g., Point where substance is released).

Nonpoint Stationary and Nonroad Mobile Sources

CAPs	HAPs	Data Element	Data Definitions
Process Level			
M	M	RECORD TYPE	A code that identifies the type of Record
M	M	<i>STATE AND COUNTY FIPS CODE</i>	The FIPS code for the state and county.
M	M	<i>SCC</i>	EPA Source Category Code for Area and Mobile Sources.
	N	PROCESS MACT CODE	Maximum Achievable Control Technology for HAP regulated sources.
N		WINTER THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		SPRING THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		SUMMER THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		FALL THROUGHPUT PCT	The percentage a process operates during the spring months. Whole number between 0 to 100.
N		ANNUAL AVG DAYS PER WEEK	Average number of days per week an emission process is active within year.
N		ANNUAL AVG WEEKS PER YEAR	Average number of weeks per year an emission process is active.
N		ANNUAL AVG HOURS PER DAY	Average number of hours per day an emission process is active within year.
N		ANNUAL AVG HOURS PER YEAR	Average number of hours per year an emission process is active.
	N	PROCESS MACT COMPLIANCE STATUS	Major/Area classification and status under CAAA Sections 112&129.
M	M	<i>TRIBAL CODE</i>	Codes that represent American Indian tribes and Alaskan Native entities.

Period Level			
M	M	START DATE	Start date of the period in which reported emissions occur.
M	M	END DATE	End date of the period in which reported emissions occur.
N		ACTUAL THROUGHPUT	Numeric value of process activity.
N		THROUGHPUT UNIT NUMERATOR	Throughput unit of measure.
N		MATERIAL	Material code for material processed.
N		MATERIAL I/O	A descriptor indicating whether material is used or produced.
N		PERIOD DAYS PER WEEK	Avg no. days/wk the process is active within the time period specified.
N		PERIOD WEEKS PER PERIOD	Avg no. wks/period the process is active within the time period specified.
N		PERIOD HOURS PER DAY	Avg no. hrs/day the process is active within the time period specified.
N		PERIOD HOURS PER PERIOD	Avg no. hrs/period the process is active within the time period specified.
Emission Level			
M	M	EMISSION NUMERIC VALUE	Numeric value of emission.
M	M	EMISSION UNIT NUMERATOR	Unit of measure for reported emissions value. If criteria emissions, report unit: annual = TON; average season day = LBS.
M	M	EMISSION TYPE	Code describing temporal designation of emissions reported, i.e., Entire Period, Average Weekday, etc.
N		FACTOR NUMERIC VALUE	The numeric value of the emission factor.
N		FACTOR UNIT NUMERATOR	Unit of measure for emission factor numerator.
N		FACTOR UNIT DENOMINATOR	Unit of measure for emission factor denominator.
N		MATERIAL	Material code for material processed.
N		MATERIAL I/O	A descriptor indicating whether material is used or produced.
N		RULE EFFECTIVENESS	Measure of the percent effectiveness of the control strategy.
Control Level			
N	N	PRIMARY PCT CONTROL EFFICIENCY	The percent effectiveness of primary control device.
N	N	PCT CAPTURE EFFICIENCY	Numeric value for percentage capture efficiency of control system.
N	N	TOTAL CAPTURE CONTROL EFFICIENCY	Collective (aggregate) value for all controls.
M	M	PRIMARY DEVICE TYPE CODE	The primary type of control equipment used.

Appendix D: Emissions Data Needs for Air Quality Modeling

Emission inventories are processed with emissions models (e.g., SMOKE) so that they can be used as an input to air quality models. Because air quality modeling is used to develop regulations that intend to improve air quality in a sensible, effective, and cost-efficient way, it is critical that emission inventories are accurate and include the necessary information to support air quality modeling analyses. Because early versions of the NEI (and its predecessor) were designed prior to the development of many emissions modeling tools, the emission inventory data that are needed for emissions modeling can be different from what is collected and maintained in the NEI. In this appendix, we describe the general data needs for emissions modeling and which data are most useful for modeling purposes based on the latest emissions processors and future anticipated emissions modeling capabilities.

The reasons for much of the data suggested in this appendix are the following:

- Include data that allows consistent emissions estimates to be developed (e.g., if the activity is known for all sources in a particular category, a consistent approach can be used to optionally estimate emissions when problems are identified in the inventory). It is possible to QC the data in the NEI by comparing the emissions to emissions calculated from a consistent emissions estimation methodology.
- Include local-specific information that effects the spatial, temporal, and chemical distribution of emissions for air quality modeling.
- Include additional information that helps emissions modelers understand the origin and meaning of the data when data provided are not consistent with one of the reporting guidelines. For example, additional information that would explain what an SCC means that is not on the official SCC list.
- Include additional information to use emission emissions models that may not be used when developing an national emission inventory (e.g., S/L/T inputs to models such as MOBILE6 and NONROAD).

What are the needs concerning emissions and emissions-associated data?

While nearly all data that one can provide in the NIF3.0 format is useful for emissions modeling purposes, some types of data are more useful than others. We have organized this appendix by major source sector; for each, we have described (1) which data fields from the NIF3.0 format are critical for optimal air quality modeling performance and why, and (2) which additional data could be used for emissions modeling. We do not include the NIF3.0 primary keys in this list of critical inputs since these fields are required by the NIF3.0 format already and their accuracy is essential for all aspects of the inventory.

All Source Categories

The following list provides suggestions that generally apply to inventory submittals for all source categories. The items here reflect the recommended handling of fields that are part of NIF3.0 and why the recommended approach is useful for air quality modeling purposes.

- **Emission process seasonal throughput data for all seasons:** This can be used in emissions modeling as a simple way to derive emissions for all seasons, without the complexity of reporting average-day seasonal emissions for just one season. Since many modeling efforts are now annual modeling (particularly for PM), providing emissions for only a single season is not sufficient.

- **Emission process heat input:** This field allows emissions models to QC the combustion source categories by comparing the heat input with the emissions. It also optionally allows emissions models to recalculate emissions based on heat input, emission factors for the SCC of interest, and control device and rule effectiveness information.
- **Emission process MACT compliance status (where applicable):** Useful when modeling control strategies to determine whether MACT controls have already been applied to the source or not.
- **Emission period throughput and units:** Important for QC purposes and for calculating emission values when not reported but are known to exist.
- **Emission period start and end dates of the period in which the emissions occur:** Useful for characterizing the temporal distribution of emissions over the course of the year, for running the hour-based air quality model.
- **Control equipment control efficiencies and method; Control device types; Emission rule effectiveness and method:** Useful for developing emission control strategies using air quality models, to know whether sources already have controls and whether additional controls might be considered for a source. It is undesirable to EPA, S/L/T's, and industry to develop control strategies that "over-control" a source category past what is possible in reality; but EPA cannot prevent this error unless accurate information is provided.
- **Emissions reporting units (e.g., tons or lbs):** Essential for preventing the inadvertent overestimation or underestimation of emissions at a source.
- **Emissions:** It is important for S/L/T's to provide emissions for all pollutants, including zeros where zero emissions are intended. It is also important to notify EPA that zero emissions values means that the submitting S/L/T has confirmed that the emissions are really zero, not just a null value. The reason for this is that if inventory submittals are incomplete or imprecise, inventory developers may fill in emissions values if the intent of the submittal is unclear. We need to minimize the number of cases where emission values are not reported or where inaccurate values are reported. This is because errors may be introduced when we fill in missing or inaccurate emissions values and air quality modeling results are negatively impacted. Thus, this should be avoided whenever possible.
- **Emissions code description of method used to derive emissions:** This is important for QC purposes in identifying why States near one another may have drastically different emissions for a particular source category. These discrepancies are useful for identifying inaccuracies in the inventory that can negatively affect air quality modeling results.

Additionally, there are pieces of information that are useful for interpreting what has been submitted in the NIF3.0 format. This information would have to be provided as part of the transmittal records (i.e., in the "strTransactionComments" field) or as part of additional written correspondence. It is important to provide all such information in written form (email, memoranda, etc) for the most effective record-keeping purposes, though in practice EPA has approaches for documenting phone conversations and other interpersonal communications.

- **If adding an SCC that is not on the official list, provide a description of the SCC in the four-tier format.** This is useful for emissions modeling purposes to allow the inventory developers to identify if in fact an SCC is a new SCC or if there was an error made in the submittal. For new SCCs, this description allows the most appropriate selection of temporal, spatial, and speciation approaches to use

in the emissions model, as well as to add to reports produced during emissions modeling to know the meaning of the emissions data.

- **If using pollutant code that is not on the official list, a description of the chemical makeup of that pollutant.** The reasons for providing this information are the same as for the information for new SCCs just described.
- **If providing average-day data for any period or annual data for a subset of the year, make sure to indicate the time period that it applies and whether the time periods not provided are intended to be zero or if no information is known about them.** Most purposes of the NEI use the annual emissions values. Air quality modeling benefits from more specific temporal information about emissions; therefore, emissions models would like to use this information when it is available. However, it is difficult to use the data for one period when information about the other periods is not available. If information is not provided about the emissions during those other periods, emissions modelers must make assumptions about those other periods, which can lead to incorrect emissions being used in air quality modeling and the associated possible problems (e.g., bad model performance, improper control strategy assessments, additional required analysis and sensitivity testing).

Point Sources

The discussion above under the heading “All Source Categories” apply to Point and Nonpoint sources. Additionally, there are some pieces of information that are important specifically for Point sources. These are listed below.

- **Site ORIS facility code; Emission release point ORIS boiler ID:** These codes are used to match the records to the CEM database, which is used to populate the annual emissions for EGUs and other sources covered by CEM. The SMOKE model can also use these fields to automatically match records to the hourly CEM data, but they need to be filled in the NIF for that to happen. This allows hourly high-quality data to be used for these major sources.
- **Site SIC code:** In emissions modeling, SIC is used for assigning factors for growth of point sources. If no SIC is supplied, then SCC must be used to assign growth factors, which is not always a good way to assign factors.
- **Site NAICS code:** NAICS codes are replacing SIC codes, but emissions modeling has not taken this step because the NAICS codes have not been commonly available in the inventories. The sooner these codes are populated the sooner we can take advantage of the benefits of this improved code.
- **Emissions unit design capacity and units:** When growing emissions, its useful to know the design capacity of the facility. If these data were provided, emissions models (or inventory preparers) could check to ensure that existing units are not grown past their design capacity, which helps prevent overestimation of future year emissions.
- **Emission release point stack height, diameter, exit gas temperature, and exit gas velocity or release height (for fugitives):** Emissions models predict the elevated plume rise of emissions from point sources, which causes emissions to be in the elevated model layers of the air quality model. This distribution of emissions is important for preventing overestimation of layer-1 emissions and for modeling long-range transport of pollutants. The stack parameters are used by the emissions models and air quality models to estimate plume rise for more realistic modeling.

- **Emission release point x and y coordinates:** By definition, point source emissions occur at a point (or approximately so) on the face of the earth. The location of the facility in the inventory will determine to which model grid cell the emissions from a facility will go in the air quality model. This in turn significantly affects the air quality modeling results. Without accurate latitude and longitudes, the point sources cannot be properly modeled and could cause air quality models to give incorrect results. For emissions and air quality modeling, its useful to have the location information with as many significant figures as possible.
- **Emission process MACT code.** Particularly for toxics, this code is critical for applying many of the toxics emissions modeling processing steps, including controls programs. Without this code, some toxics facilities will not be treated accurately in the air quality modeling, possibly leading to erroneous results.

In addition to the correct NIF3.0 fields being provided as described above, additional information can be useful for interpreting and using the NIF data.

- **Whether ORIS ID and ORIS Boiler ID match what is in the CEM database, and if not, what the codes are in the CEM database that correspond to this Site and Emission unit.** Because the ORIS IDs used for the CEM program have been fixed since 1995, but the official ORIS IDs have in fact changed over time, there can be discrepancies between the CEM codes and the current codes. If these differences are known, it is useful to have this information for matching between the inventory submittals and the CEM database.
- **Day-specific emissions data:** Emissions models can use day-specific emissions data for one or more pollutants for point sources. These day-specific (not average day) data are used to override the annual inventory adjusted by monthly profiles. The day-specific emissions are adjusted to hourly values using the diurnal temporal profiles.
- **Hour-specific emissions and velocity, flow, or exist gas temperature data:** Emissions models can input hour-specific emissions data for one or more pollutants for point sources. Hour-specific stack parameters can also be used by some emissions models for estimating hourly plume rise. When these data are available in the inventory, they override less specific annual or other numbers in the inventory.

Nonpoint (Stationary) Sources

The discussion above under the heading “All Source Categories” apply to Point and Nonpoint sources. Additionally, there are some pieces of information that are important specifically for Nonpoint sources. These are listed below.

- **Whether tribal emissions have been included or excluded from county emissions:** When tribal data are provided, it raises a question about the emissions inventory for the States and counties that intersect that tribe. The NIF3.0 format does not provide a way for submittals to indicate whether the tribal emissions have been subtracted from the county emissions or not. Without knowing the situation, emissions models must either assume that the emissions have not been subtracted and subtract the tribal emissions that intersect the county or do nothing. Without an indication of how the estimates where developed, emissions models could double-count the tribal inventory or subtract the tribal inventory twice from the county totals.

- **Activity data, fuel loading, etc. for EPA-modeled source categories (e.g., fire inventories):** Air quality modelers like the consistency that a national approach to modeling provides, but realize that top-down estimates of activity information may lead to inaccurate results. Therefore, for source categories that EPA calculates, S/L/T activity data would provide a better basis for calculating these emissions.
- **Nonpoint sources that are handled by emissions models as point sources (e.g., airports, ports):** Because these sources are moved by the emissions models to point locations, it is important to know which point locations to include. If this information is not provided, the emissions can be calculated for one set of sources (e.g., 4 airports), but emissions moved to fewer locations (e.g., 2 airports). Furthermore, without correct information about the precise locations and county assumptions used, it can be impossible to identify a data source that uses the same information to use during spatial allocation for air quality modeling purposes.

Nonroad mobile

When NIF-based inventories are submitted, the useful fields are the same as those in described previously in the Plan. However, since the NONROAD model is used by EPA to compute county-specific nonroad mobile source emissions, local inputs to the NONROAD model can improve the accuracy of the calculated emissions. The NONROAD inputs include local-specific information about equipment population, activity, and seasonal activity patterns. Where S/L/Ts have better data than the EPA national defaults, it is preferable to use this information for use air quality modeling.

On-road mobile

The NEI annual on-road mobile estimates are not generally used for air quality modeling. Since on-road mobile emissions are highly dependent on temperature, it is important to have at least month-specific emissions calculated based on some average-monthly temperatures. Additionally, emissions models have the capability of running the MOBILE6 model using hourly, gridded meteorology to more accurately reflect the impact of temperature on emissions. In this latter case, additional data not included in the NEI can be useful for modeling purposes. These data are:

- **Vehicle Miles Traveled (VMT):** We need VMT at the county and SCC level at a minimum and based on travel demand models at the road link level if possible and the case warrants it. For national scale coarse-grid (e.g., 36-km) modeling, county-based data and road-based spatial surrogates are often sufficient. For local scale modeling (e.g., 4-km), particularly in nonattainment areas, link-based modeling is sometimes desirable. Link-based data must include the starting and ending coordinates of the links, which are straight road segments representing road locations and that do not intersect state or county lines.
- **Vehicle Mix:** Vehicle mix represents the mixture of MOBILE6 vehicle types on the road in each county. Since the emission factors are assigned by vehicle type (in addition to many other factors), it is helpful to have local-specific information about vehicle mix.
- **Average Speed:** Emissions models need at a minimum the estimated average speed for each SCC (road and vehicle type) in the inventory.
- **MOBILE6 Input/Output Files (and which counties share which MOBILE6 inputs):** The MOBILE6 model has dozens of inputs that may be provided to calculate the emission factors used for estimating emissions. Many of these inputs, such as fuel type, can have a significant impact on the emission factors. Inputs such as speed and temperature also have a large impact, but are not

provided via the MOBILE6 inputs to emission models. Because some of the inputs change on a seasonal basis (e.g., fuel information), it is also useful to provide these MOBILE6 inputs seasonally where necessary.

Biogenic and geogenic sources

For modeling purposes, EPA uses the BEIS3 model to estimate emissions, and will most likely use version 3.12 of this model during 2004. This model estimates VOC and NO emissions from vegetation and soil. Other biogenic and geogenic emissions are not estimated by BEIS3. To obtain the best biogenic inventory, we need additional information from S/L/T's:

- Whether and why the BEIS3.12 model is insufficient for their area for biogenic emissions from vegetation and soil sources.
- Alternative emission estimates on a fine-resolution gridded and hourly basis (e.g., 4-km or smaller) of model-species VOC and NO emissions.
- Emissions from major geogenic sources (e.g., SO₂ emissions from volcanic activity).

Ancillary data

In addition to inventory data, emissions models need data to adjust the inventory to the required temporal, spatial, and chemical resolution. Models need these adjustment factors, called profiles, as well as information about assigning the factors to the inventory, called cross-reference files, in the form of ASCII files. Note that for spatial allocation, the factors are called spatial surrogates instead of profiles. EPA has national defaults for these ancillary data that are not frequently updated and which could be inaccurate for air quality modeling purposes. Therefore, it is important for S/L/Ts to collect and provide these ancillary data specific to their region. Specifically, the key information that can be helpful for air quality modeling is:

- **Temporal profiles, particularly for hard-to-estimate and irregularly patterned activities such as fires and on-road VMT.** These data help to allocate emissions to more specific time periods (e.g., annual to seasonal, daily to hourly) of the modeling episode. They include the following subsets of data.
 - **Monthly profiles:** These profiles adjust annual emissions to average-day emissions that are specific to each month. It is also important for S/L/Ts to provide accurate seasonal throughputs, which can be used if the more detailed monthly data are not available.
 - **Weekly profiles:** These profiles adjust the average-day emissions to each day of the week (Sunday through Saturday).
 - **Diurnal profiles:** These profiles adjust the weekday-specific emissions to each hour of the day. If necessary, the diurnal profiles can be provided for all days of the week, weekdays and weekends separately, or each individual day of the week (Saturday through Sunday).
- **Spatial information.** EPA computes spatial surrogate information and makes available both the model-ready spatial surrogate files and the GIS shapefiles that were used to create the model-ready inputs. States can help EPA to better model the inventories by providing the following additional information.

- GIS shape files for animal husbandry activities in the state. This will help to spatially allocate this large ammonia source category for which data have traditionally been very difficult to obtain.
 - Any changes to county boundaries and the date of the change.
 - Any additional information about which spatial surrogates are most appropriate for certain SCCs. S/L/Ts would need to review the default spatial cross-reference files to decide whether those assignments make sense for their area.
- **Chemical speciation information.** Currently, we need better data regarding chemical speciation, though ongoing work seeks to improve this major concern for emissions modeling accuracy.
 - States can help improve the knowledge base by providing any speciation profiles developed for local modeling, the chemical mechanism to which the profiles apply, and the SCCs to which the profiles apply. The profiles should be based on Total Organic Gases (TOG) and should be at least for the CB4 chemical mechanism.
 - S/L/Ts can also review the existing profiles and speciation cross-reference assignments and provide feedback about changes that should be made, what those changes are, and why the changes are known to be an improvement.

Appendix E: Example Chain of Custody Form for Data Submittals

Chain of Custody

Submittor Name: CA

Submittor Contact Information: Andrew A. Alexis, (916) 323-1085, aalexis@arb.ca.gov

Date Submitted to CDX: 06092004

Date Downloaded from CDX: 06102004

Downloaded by: **Sally Dombrowski** - ali

Phone: **(919) 541-3269**

Initial File Information

File Name	File Type	File Size	Table Name	Row Count Per Table
neica04andrewalexis.mdb	Access	92.3 MB	TR	56
			CE	0
			EM	207,291
			EP	72,256
			ER	46,752
			EU	63,412
			PE	72,256
			SI	11,627

Transfer File Information

From	To	Date	File Name	File Size	Table Name	Row Count Per Table	Comments
Sally Dombrowski	ERG	06112004	CA2002PTCAP06092004.mdb	77.7 MB	TR	56	Submittal replaces previous
					CE	0	
					EM	207,291	
					EP	72,256	
					ER	46,752	
					EU	63,412	
					PE	72,256	
					SI	11,627	

Appendix F: Example Communications with Point Source Data Providers

Example Email to Data Submitter Agency by the EFIG Point Source Data Development Leader

From: Anne Pope

07/01/2004 09:23 AM

To: buzz.asselmeier@epa.state.il.us

cc: Suzanne King/R5/USEPA/US@EPA, Rafiu Dania/R5/USEPA/US@EPA, Charles Hatten/R5/USEPA/US@EPA, Regina Charles/R5/USEPA/US@EPA, Loretta Lehrman/R5/USEPA/US@EPA, stephanie.finn@erg.com, melodie.vines@erg.com, janssen@ladco.org, David Misenheimer/RTP/USEPA/US@EPA, Phil Lorang/RTP/USEPA/US@EPA, Douglas Solomon/RTP/USEPA/US@EPA

Subject: Illinois Submittal for the 2002 Point Source NEI

Hello Buzz,

Thank you for supplying data and comments for the point source National Emissions Inventory (NEI) for criteria pollutants (CAPs) and hazardous air pollutants (HAPs).

To date, we have received 1 submittals on point source CAPHAP. Your submittal contained the following information.

- 102 counties out of 102 counties in the state
- 140 HAPs, 8 CAPs
- 8,950 facilities
- 258,167 emission records

The submittals has been checked and 5 errors have been found. This QA/QC process includes most of the checks covered in the format checker and additional checks by ERG. There are 2 files attached to this e-mail for each submittal. The QC Report lists the errors associated with the submittal. The Access file contains tables with all of the errors and includes a README on how to correct the errors.

Please review the QC report and make your revisions in the provided Access file. If you need assistance, or will not be able to provide revisions within THREE working days of receipt of this email, please let me know. If you have any questions about specific data errors or you are unable to open the files, please contact Melodie Vines at 919-468-7823 or Stephanie Finn at 919-468-7889 for additional information.

All revisions should be returned directly to pope.anne@epa.gov.

Thank you,

Anne Pope
US EPA Office of Air Quality Planning and Standards
Emission Factor and Inventory Group
919-541-5373

[Attached files: IL QC Report.doc, Illinois Errors.mdb]

Example Point Source QC Report Emailed to Data Submitter Agency by the EFIG Point Source Data Development Leader

QC Report Filename: IL QC Report.doc
Access Database Listing Erroneous Records: Illinois Errors.mdb

2002 NEI QC Report

State/Local Agency/Tribe: Illinois
Submittal Date to CDX: May 27, 2004
Inventory Type: CRIT & HAP

Summary/Instructions:

The following report outlines the errors found within the data submittal. The report is in three sections: critical, important, and minor errors. The critical errors need to be addressed by the state, local, or tribal agencies. The important and minor errors can be addressed by ERG, although any agency assistance would be helpful. If the errors are not addressed then ERG will correct them the best way possible.

All of the records associated with the errors can be found in the Access database accompanying this report.

Critical Errors

Table	Error Type	Count
Referential Integrity	Processes > 1 Emission Release Point	2
Referential Integrity	Emission Release Point ID in EM not EP	1
Referential Integrity	Widows: B/W SI & EM	1822

Important Errors

Table	Error Type	Count
tblPointEM	CAP Emission Out of Range	124
tblPointER	Coordinates Outside County	143

Appendix G: Example Communications with NonPoint Source Data Providers

Example Email to Data Submitter Agencies by the EFIG NonPoint Source Data Development Leader

From: Roy Huntley

07/16/2004 03:50 PM

To: Madonna Narvaez/R10/USEPA/US@EPA, Robert Mcconnell/R1/USEPA/US@EPA, Susan Lancey/R1/USEPA/US@EPA, Raymond Forde/R2/USEPA/US@EPA, Alice Chow/R3/USEPA/US@EPA, Leonardo Ceron/R4/USEPA/US@EPA, Scott Martin/R4/USEPA/US@EPA, Charles Hatten/R5/USEPA/US@EPA, Loretta Lehrman/R5/USEPA/US@EPA, Rafiu Dania/R5/USEPA/US@EPA, Regina Charles/R5/USEPA/US@EPA, Suzanne King/R5/USEPA/US@EPA, Herb Sherrow/R6/USEPA/US@EPA, Michael Jay/ARTD/R7/USEPA/US@EPA, Mark Komp/P2/R8/USEPA/US@EPA, Daniel Webster/P2/R8/USEPA/US@EPA, Larry Biland/R9/USEPA/US@EPA, Pam Tsai/R9/USEPA/US@EPA, mschuster@marama.org, skayin@marama.org, gms@alpinegeophysics.com, pat.brewer@ncmail.net, sheila.holman@ncmail.net, elizabeth.byers@la.gov, janssen@ladco.org, mooret@cira.colostate.edu, david.holoman@pechan.edu, mschuster@marama.org, skayin@marama.org, gms@alpinegeophysics.com, pat.brewer@ncmail.net, sheila.holman@ncmail.net, elizabeth.byers@la.gov, janssen@ladco.org, mooret@cira.colostate.edu, david.holoman@pechan.edu, bdowning@mail.maricopa.gov, lbcole@adem.state.al.us, withers@adeq.state.ar.us, aalexis@arb.ca.gov, dale.wells@state.co.us, yi.zhu@dep.state.fl.us, bstoptei@co.pinellas.fl.us, scott_southwick@dnr.state.ga.us, greinbol@deq.state.id.us, sarah.kelly@nau.edu, buzz.asselmeier@epa.state.il.us, janssen@ladco.org, marnie.stein@dnr.state.IA.us, wvit@kdhe.state.ks.us, paul.lanka@loukymetro.org, elizabeth.byers@la.gov, ostranda@state.mi.us, elliot_bickerstaff@deq.state.ms.us, patricia.tighe@dnr.mo.gov, angelique.leudeker@nau.edu, deqnode@ndeq.state.ne.us, UHL@co.clark.nv.us, dhealy@des.state.nh.us, rothaj@rapca.org, fields.brian@deq.state.or.us, cheathcc@dhec.sc.gov, metorres@tceq.state.tx.us, sott461@ecy.wa.gov, tcfoster@deq.state.va.us, dporter@wvdep.org, patter@dnr.state.wi.us, johna@psccleanair.org

cc: David Misenheimer/RTP/USEPA/US@EPA, Douglas Solomon/RTP/USEPA/US@EPA, Laurel Driver/RTP/USEPA/US@EPA

Subject: Nonpoint EI summary Information from EPA

The attached Excel Workbook file (2002 State Nonpoint EI Tracking_July1604.xls) contains the tracking sheets for the State, Local, and Tribal (S/L/T) agency nonpoint inventories. We (via our contractor, Pechan and Associates) are working with the State/Local data submitters one on one to address questions and resolve errors identified during our initial review of the data files. This is mostly for your information. However, if you see something amiss, let us know ASAP.

We will not be providing such a summary of files received concerning Onroad and Nonroad Mobile Source files because of the complexity of the data files and formats.

The attached Excel workbook contains the following spreadsheets:

1. List of S_L_T Submittals: This spreadsheet identifies the S/L/T agencies that submitted inventories by June 1 and after June 1.
2. 2002 Nonpoint File Names: This spreadsheet identifies the names of the files that were received from a submitter or were generated by EPA's contractor (i.e., QA error files).
3. 2002 Nonpoint Summary: This spreadsheet shows the geographic and pollutant coverage of each inventory.
4. 2002 Nonpoint Emission Sums: This spreadsheet summarizes the temporal basis of the emissions in each inventory.

5. 2002 Nonpoint Error Summary: This spreadsheet contains a copy of the "SummaryStats" table from the QA program's output file.

Note that we have moved nonroad and onroad SCCs from the nonpoint inventories to the nonroad and onroad inventories. We have noted this in the spreadsheets where this is an issue. The following SCCs will remain in the nonpoint inventory: paved and unpaved road dust (SCCs starting with 229), unpaved air strips (SCC 2275085000) and aircraft refueling (SCCs starting with 22759).



2002 State Nonpoint EI Tracking_July1604.x

Roy Huntley
Environmental Engineer
Emission Factor and Inventory Group (D205-01)
Environmental Protection Agency
RTP, NC 27711
919 541-1060

Excerpt from 2002 Nonpoint Error Summary - Spreadsheet File (Filename: 2002 State Nonpoint EI Tracking_July1604.xls)

Tracking of 2002 State, Local, and Tribal Inventories Submitted to EPA
NonPoint Sources

Summary Stats Tables from EPA QA Program

Last Updated July 16, 2004; Crossed out entries mean files have been superceded by subsequent submittals

State/Local/ Tribal	Table	NumRecs	ErrRecs	Duplicate	Invalid Type	Mandator y Missing	Length Error	Range Error	Indeterminable	Lookup Error	Widow	Orphan	Notes
AL	TR	67	0	0	0	0	0	0	0	0	---	---	
AL	EP	8680	0	0	0	0	0	0	0	0	---	---	
AL	CE	0	0	0	0	0	0	0	0	0	---	---	
AL	PE	8680	0	0	0	0	0	0	0	2546	---	---	
AL	EM	22965	0	0	0	0	0	181	0	2814	---	---	
AL	TR/EP	8680	---	---	---	---	---	---	0	---	0	0	
AL	PE/EM	22965	---	---	---	---	---	---	0	---	0	0	
AL	EP/PE	8680	---	---	---	---	---	---	0	---	0	0	
AL	EP/CE	0	---	---	---	---	---	---	0	---	0	0	
AL	CE/EM	0	---	---	---	---	---	---	0	---	0	0	
AR	TR	75	0	0	0	0	0	0	0	0	---	---	
AR	EP	3013	0	0	0	0	0	0	0	0	---	---	
AR	CE	10548	0	0	0	0	0	0	0	0	---	---	
AR	PE	8552	0	0	0	0	0	0	0	0	---	---	
AR	EM	29422	0	0	0	0	0	77	0	368	---	---	
AR	TR/EP	3013	---	---	---	---	---	---	0	---	0	0	
AR	PE/EM	29422	---	---	---	---	---	---	0	---	0	0	
AR	EP/PE	8552	---	---	---	---	---	---	0	---	0	0	
AR	EP/CE	10548	---	---	---	---	---	---	0	---	0	0	
AR	CE/EM	29422	---	---	---	---	---	---	0	---	0	0	
AZ Maricopa	TR	1	0	0	0	0	0	0	0	0	---	---	
AZ Maricopa	EP	85	0	0	0	0	0	0	0	0	---	---	
AZ Maricopa	CE	0	0	0	0	0	0	0	0	0	---	---	
AZ Maricopa	PE	165	0	0	0	0	0	0	0	0	---	---	
AZ Maricopa	EM	480	0	0	0	0	0	19	0	0	---	---	
AZ Maricopa	TR/EP	85	---	---	---	---	---	---	0	---	0	0	
AZ Maricopa	PE/EM	343	---	---	---	---	---	---	0	---	0	0	
AZ Maricopa	EP/PE	165	---	---	---	---	---	---	0	---	0	0	
AZ Maricopa	EP/CE	0	---	---	---	---	---	---	0	---	0	0	
AZ Maricopa	CE/EM	0	---	---	---	---	---	---	0	---	0	0	