The Challenge of Meeting New EPA Data Standards and Information Quality Guidelines in the Development of the 2002 NEI Point Source Data for HAPs

Anne Pope
Emission Factor and Inventory Group, U.S. Environmental Protection Agency
D205-01, USEPA Mailroom, Research Triangle Park, NC 27711
pope.anne@epa.gov

Stephanie Finn and Darcy Wilson
Eastern Research Group, Inc., 1600 Perimeter Park, Morrisville, NC 27560
stephanie.finn@erg.com

ABSTRACT

The Environmental Protection Agency (EPA) has established new uniform data standards and the Office of Management and Budget (OMB) has established new information quality guidelines that will affect the development of the National Emission Inventory (NEI) for hazardous air pollutants (HAPs). In order to comply with the new data standards and the information quality guidelines, the EPA has revised its NEI for HAPs database management system for point sources. The EPA is now facing the challenge of how to compile 2002 HAP emission inventory data required to meet the new data standards and information quality guidelines. The new data standards and information quality guidelines will help improve the quality of the NEI for HAPs for its use in meeting the requirements of the Clean Air Act (CAA). The CAA has established the need for a high quality comprehensive HAP emissions inventory that can be used to: estimate risk, and track progress by the EPA over time in reducing HAPs in ambient air. To meet these needs, the EPA compiles the NEI for HAPs to provide a model-ready emissions inventory. The EPA has previously compiled a baseline 1990 and 1996 HAP inventory, is completing the development of the 1999 NEI for HAPs, and is currently planning for the development of the 2002 NEI for HAPs. This paper summarizes the new data standards and information quality guidelines, highlights the changes in the NEI for HAPs database management system, and discusses the process EPA will take to compile new data to meet the data standards and information quality guidelines.

INTRODUCTION

To ensure maximum objectivity, utility and integrity of data disseminated by federal agencies, the Office of Management and Budget (OMB) has required all federal agencies to issue their own information quality guidelines.[1] EPA's guidelines require data producers to closely adhere to existing EPA quality procedures and ensure the transparency of their information products. Data providers must include sufficient documentation such that potential end-users can assess the suitability of the data product for their own uses. Where a product is deemed "influential," the owner must guarantee its reproducibility. EPA's Emission Factor and Inventory Group (EFIG) will comply with these guidelines by more rigorous and thorough documentation, the assignment of a data rating to emission estimates, and the introduction of derived tables that track the history of each estimate in successive versions of the National Emission Inventory (NEI).

EPA's Office of Environmental Information (OEI) has created uniform data standards or elements which provide "meta" information on the standard NEI Input Format (NIF) fields.[2] These standards were developed by teams representing states, tribes, EPA and other federal agencies. The use of common data

standards among partners will foster consistently defined and formatted data elements and sets of data values, and provide public access to more meaningful data.[3] The standards relevant to the NEI are the: SIC/NAICS, Latitude/Longitude, Chemical Identification, Facility Identification and Contact Data Standards. Compliance with these standards requires addition of new columns to one or more NIF tables to accommodate the new elements, revision of some field names, and the replacement of some data elements with others. These structural changes will be relatively easy to implement. However, populating these new elements poses a challenge for both EFIG and state and local agencies.

Eastern Research Group, Inc. (ERG) compiles the point source NEI for hazardous air pollutants (HAPs) under the auspices of EFIG. Compliance with these guidelines and standards has challenged ERG to re-think its data handling processes. ERG's goal is to efficiently integrate these compliance activities into existing data management practices. These new processes will be tested on the 1999 NEI for HAPs and will be fully implemented during the creation of the 2002 NEI. Although the standards and guidelines introduce new complexity in the NEI creation process, they ultimately will benefit state and local agencies (and other users of the NEI) by providing a better, more transparent inventory.

1.0 INFORMATION QUALITY GUIDELINES

1.1 Overview

EPA developed its Information Quality Guidelines (IQG) in response to guidelines issued by OMB under Section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (P.L. 106-554; H.R. 5658). (See http://www.epa.gov/oei/qualityguidelines/ for more details.) The *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency* (the Guidelines) embody the following performance goals:

- Disseminated information should adhere to a basic standard of quality, including objectivity, utility, and integrity;

- Principles of information quality should be integrated into each step of EPA's development of information, including creation, collection, maintenance, and dissemination; and

- Administrative mechanisms for correction should be flexible, appropriate to the nature of and timeliness of the disseminated information and incorporated into EPA's processes.[4]

These guidelines apply to "information" that EPA "disseminates" to the public. Such information includes any communication or representation of knowledge such as facts or data, in any medium or form, including web sites, FTP sites, brochures, data flat files, scientific studies etc. However, the guidelines DO NOT apply to all products distributed by EPA. EPA must sponsor or initiate the distribution of the information and EPA must adopt or endorse this information:

- "EPA initiates a distribution of information if EPA prepares the information and distributes it to support or represent EPA's viewpoint, or to formulate or support a regulation, guidance or the agency decision or position."

- "EPA initiates a distribution of information or EPA distributes information prepared or submitted by an outside party in a manner that reasonably suggests that EPA endorses or agrees with it; if EPA indicates in its distribution that the information supports or represents

EPA's viewpoint; or if EPA in its distribution proposes to use or uses the information to formulate or support a regulation, guidance, policy, or other Agency decision or position."

- "Agency-sponsored distribution includes instances when EPA reviews and comments on information distributed by an outside party in a manner that indicates EPA is endorsing it, directs the outside party to disseminate it on EPA's behalf, or otherwise adopts or endorses it."[4]

1.2 Does It Follow Appropriate EPA Policies?

The EPA has prepared a "pre-dissemination" checklist that summarizes the important aspects of the IQG and provides a list of specific actions essential to satisfying the requirements of the guidelines. In general, the IQG (and the checklist) reinforce existing EPA quality procedures. Specifically, it asks these questions of the NEI:

- Is the NEI subject to EPA peer review policy? This policy covers "major" products as defined in the *Science Policy Council Peer Review Handbook*.

- Does the NEI use environmental data? If so, the NEI must have a Quality Assurance Plan (QAPP).

- Did the NEI conduct an assessment of existing data when used to support agency decisions or other secondary purposes? Is this data of sufficient quality and quantity to meet the objectives of the product?

- Does the NEI fall under the guidelines of the EPA Risk Characterization Policy Handbook?

Only two of the items listed above apply to the 1999 NEI - the NEI must have a QAPP and must assess the data to determine if it is of sufficient quality to meet NEI's objectives. These items are addressed below. EPA has determined that the 2002 inventory will be subject to peer review. This is discussed in the next section.

Our QAPP is part of the Inventory Preparation Plan (IPP). Since it has been two years since this document was drafted for the 1999 NEI, we plan to revisit the QAPP and revise it if necessary. As ERG continuously refines and improves its quality assurance/quality control (QA/QC) procedures, this document should reflect the scope and extent of our current QA/QC activities.

Before incorporating new data into the NEI, ERG follows a rigorous process of data review and standardization. All new data sets are logged, checked for database integrity, valid pollutant, unit and other codes, and uniqueness of records. These data sets must conform to NIF. Each new data set is compared to the current NEI to 1) assign NTI Unique IDs to facilities and 2) determine how the new set will be incorporated into the NEI. For example, if the state has only revised a few records, then only those records will be replaced. If however, the state has submitted a new inventory, it is tested for completeness. This includes reviewing: HAP, facility, source category, and county coverage. If any gaps are detected, the state submittal is treated as a partial replacement set. Often, the state agency must be consulted during this process. Finally, geographic coordinates are reviewed and "geocoded" or corrected where possible. Stack parameters are reviewed for completeness and consistency and calculated or defaulted where necessary.

In the process of creating the NEI, we routinely assess data gaps and completeness. Assessing the quality of the estimates provided by the different sources is a more difficult task. Although we do conduct outlier and comparison tests to detect aberrant data, outside reviewers provide most of the in-depth assessment of the data quality. To provide a more systematic assessment on the entire inventory, we have proposed the development of a data rating for the NEI (see Section 1.4).

1.3 Is The Information Influential?

The IQG specifies that if a product is deemed influential, the author must demonstrate that the product is capable of being reproduced by a qualified third party according to commonly accepted scientific and technical standards.

In 2002, the NEI will be subject to peer review and as such will meet the definition of an "influential" product. The 1999 NEI will not be peer reviewed. However, it has gone through extensive rounds of public review, comment, and revision. Furthermore, steps have been taken to make the inventory as reproducible as possible. The next section describes these efforts.

1.4 Is The Product Transparent?

ERG currently implements some procedures on the NEI to make it more transparent to third parties and reproducible by a qualified third party user:

- Logging and archiving of all original data submittals by states, industry, EPA and other parties;
- Archiving of "holding" tables which contain the compiled revisions to the inventory with indication as to date of submittal, type of revision, and source;
- Archiving and indexing of intermediate work products essential to the production of the final NEI;
- Incorporation of data elements in the output files which allow the user to determine the source of the record (EPA, state/local agency, Toxics Release Inventory (TRI), industry or trade association) and the nature of a revised record (i.e., addition since last version, revision to last version, or original record) as well as the origin of defaulted data; and
- Documentation which summarizes the targeted end-users of the inventory, its purpose, data sources, submittal contacts, and methodology.

Thus, the groundwork for a transparent, reproducible inventory is already in place. In this section, we revisit each of these items, highlighting additional refinements to better meet the objectives of the guidelines.

Archiving and Logging Refinements

Currently, relevant project notes and intermediate work products are copied to CD-ROM and placed in the project file. Usually, a brief summary memorandum accompanies these CDS such that a third party can open the files and follow the progression of work from raw data through compiled revisions through final product. This process is well-implemented and very useful. Some simple improvements, however, would make it much more accessible to a third party. We suggest the creation of a summary sheet which lists all of the

relevant work products and provides a brief description of each. Each item on the summary sheet should have an ID number that ties it directly to the hard-copy file folder or its electronic medium.

Although we archive all emails and or summaries of phone conversations with data submitters, at present these are saved to individual working files and not to a central project file. In order for a third party to reproduce all of our steps, they would need access to these emails/calls as many ad-hoc decisions with respect to the data are made in the course of these communications. Centralizing this information would make it more accessible. We propose setting up a central folder in our email system for such communications.

Additional Data Source Elements - Maintaining an Emissions History Table

We will create and maintain an emissions history table (Table 1) which tracks the changes in total HAP emissions for each facility/site/pollutant combination in iterative versions of the NEI. This table will enable users to see past and current total emissions for each HAP at 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;
- NTI Unique Facility ID or Federal Registry System (FRS) ID - Unique ID assigned to co-located sites in NEI;
- Facility Name - Plant Name;
- Pollutant Code - Chemical Abstract Service (CAS) number;
- Emissions Value 1 - Emissions in TPY associated with version 1;
- Data Source 1- Source of emissions value in version 1;
- Ver_Date 1 - Version number concatenated with date of release;
- Emissions Value 2 - Emissions in TPY associated with version 2;
- Data Source 2 - Source of emissions value in version 2;
- Ver_Date 2 - Version number concatenated with date of release.

This table will be aggregated to the site/pollutant code level and will provide a total HAP emission estimate 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 HAP 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 in. 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 submitter 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 HAP value from among multiple data sources for the same HAP/facility, we 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.

Establish a Data Rating for HAP Estimates

To give reviewers some sense of the reliability of an emission estimate, we will develop a simplified rating scheme so that a score can be assigned to each estimate (e.g., on a scale of 1-5). This enhances the transparency of the data and also satisfies the requirement that EPA do an assessment of the data. This rating scheme will not be in-depth as other systems such as DARS (Data Attribute Rating System), but will consider the following factors in assigning a score:

1. ***Completeness of data*** - Has the submitter provided enough information to enable the reviewer to repeat the calculation, assess emission factors and/or calculation methods? The fields that provide this information are:

   - Actual throughput
   - Throughput unit numerator
   - Emissions Reliability indicator
   - Factor Numeric Value
   - Factor Unit Numerator
   - Factor Unit Denominator
   - Emission Calculation Method
   - Emission Factor Reliability Indicator

2. ***Emission Calculation Method*** - Estimates based on continuous monitoring should receive higher scores than data based on less accurate methods, e.g., "engineering judgment."

3. ***Age of data*** - In some cases, we have an emissions estimate from an earlier year, not the current inventory year. Having "old" data is preferred to having a data gap. However, a 1996 estimate should have a lower rank than a 2002 estimate.

4. ***Qualitative Information*** - We have additional information with respect to several submittals that is not reflected in the database (e.g., municipal waste combustor estimates from EPA are based on source testing). We might also want to consider breadth of data, i.e, did the source of this estimate provide a large number of HAPs relative to other sources for this category? For example, does EPA refinery data have more HAPs per facility than data provided by other sources (state and local agencies, TRI and industry)?

5. ***Specificity of Data*** - An estimate which provides process level emissions is better than aggregated facility level HAP emissions.

We will devise a simple method for assigning each score based on the attributes listed above. This scoring should be kept simple and easy to replicate. We do not want to put undo importance on this score, as many of the fields listed above may be blank. However, it may help EPA make better decisions as to which data point to retain when there are multiple estimates for the same facility. It may also help us understand which source categories need improvement.

We will have to assign some relative weighting to the factors listed above to determine the overall score. For example, one simple method would be to give an estimate a high score ("5") if it satisfies certain of the requirements listed above (e.g., source test data with high completeness), but to subtract points if the data are old or fails to fulfill other attributes. Our first step will be to evaluate the list above, make sure it represents the attributes we want to measure and then come up with a matrix of possible scoring scenarios. The final scoring system should be simple, clear, and easy to program.

Documentation Changes

The documentation for the NEI is the obvious place to meet many of the objectives of the transparency guideline. To this end, we propose the inclusion of a section (Information Quality Guidelines Addendum) which elaborates on the following items: 1) purpose, 2) explanation of potential users, 3) product content - inputs, methodologies, and outputs, 4) product limitations and caveats, and 5) contact information. Although, much of this information is already summarized in the documentation, we plan to make this a stand-alone section which can be lifted from the overall document. This addendum will focus on those items that are under-represented in the current documentation.

2.0 DATA STANDARDS

According to the Environmental Data Registry website (http://oaspub.epa.gov/edr/epastd$.startup): "The vision for the Environmental Protection Agency's (EPA) data standards program is to promote the efficient sharing of environmental information among EPA, states, tribes, and other information partners through the cooperative development of data standards."[2]

The data standards program has implications for the NEI which must incorporate or revise several data elements based on these standards. These standards will be implemented in the new version of NIF (NIF 3.0). In the following sections, we discuss each standard in turn and our implementation plans for each standard.

2.1 SIC/NAICS Data Standard

This standard includes ways to classify business activities, including industry classifications, product classifications, and product codes.[5]

The Standard Industrial Classification (SIC) System has been used for many years to provide a code system for the identification of business activities. SIC codes range from two digit general categories to four digit specific categories. SIC codes are gradually being replaced by the North American Industry Classification System (NAICS) that was adopted by Canada, Mexico, and the United States in 1997. (See http://www.census.gov/epcd/www/naicsdev.htm).[6]

The NAICS code consists of 6 digits which are arranged hierarchically:

- ***Two digits*** - Economic sector (North American Industry Classification Sector Code)

- ***Three digits*** - Economic subsector (North American Industry Classification Subsector Code)

- ***Four digits*** - A group of related industries within the economy (North American Industry Classification Industry Code)

- ***Five digits*** - An industry within the economy (North American Industry Classification Industry Code)

- ***Six digits*** - A subdivision of an industry (North American Industry Classification Code)

For example the NAICS code 334611 can be broken down as follows:

33 = manufacturing
334 = computer and electronic manufacturing
3346 = manufacturing and reproduction of magnetic and optical media
334611 = manufacturing and reproduction of magnetic and optical media, reproduction of software

To populate the NAICS field, we must first create a crosswalk of SIC codes to NAICS codes. The Census Bureau has mapped the 2002 NAICs codes to the 1987 SIC codes.[7] We will evaluate this map and come up with a preferred scheme for the NEI. Where there is a one-to-one correspondence between NAICS and SIC codes, the assignment is straightforward. However, in those cases in which one SIC code maps to many NAICS codes, we will have to map the SIC code to a less specific NAICs code (i.e., 2, 3 or 4 digit).

We will use our crosswalk to populate the NAICS field in the 1999 NEI for HAPs. States will be directed to populate this field in their 2002 submittals. However, if a state has not filled in the NAICS we will apply our crosswalk. Additionally, we may be able to obtain facility-specific NAICS codes from third-party services such as PowerBusiness by InfoUsa. The PowerBusiness database is available on a subscription basis and contains census-derived data for over 14 million U.S. businesses and is searchable by name, address, phone number, business type and SIC and NAICS codes.

2.2 Latitude/Longitude Data Standard

The latitude/longitude standard consists of the group of data elements used for recording horizontal and vertical coordinates and associated metadata that define a point on earth. Incorporating this standard in the NEI, entails changing the name of some fields as well as adding new fields to NIF. Table 1 summarizes these changes.[5]

This standard will help users gauge the accuracy and reliability of a given set of coordinates. The primary responsibility for populating these fields lies with the data submitter as it is difficult if not impossible to discern the origin of a latitude/longitude without being the primary author of the data. Since this standard was not part of NIF 2.0, ERG will only be able to populate these fields whenever latitude/longitudes were obtained from the TeleAtlas Geocoding EZ Locator Service (http://geocode.com). We assign Geocoder latitude /longitudes whenever the existing coordinates are null, clearly incorrect, or plotted well outside the county boundaries.

We have flagged the geocoded coordinate pairs in the NEI with the explanatory codes listed in Table 2. We will populate the latitude/longitude data standards for these geocoded coordinates with the default values shown there.

2.3 Chemical Identification Data Standard

The Chemical Identification Data Standard provides for the use of common identifiers throughout the Agency for all chemical substances regulated or monitored by EPA environmental programs. This standard provides unique, unambiguous, chemically correct common names for all chemicals substances and groupings in EPA's system and will facilitate automated searches for chemical substances across EPA programs and their databases.[5]

The Chemical Identification standard consists of the following data elements:

- ***EPA Chemical Internal Tracking Number (CRS ID)*** - the unique record number assigned to all chemical substances and chemical groupings for internal tracking within EPA systems;
- ***Chemical Abstracts Service Registry Number*** - the unique number assigned by Chemical Abstracts Service (CAS) to a chemical substance;
- ***Chemical Substance Systematic Name (9^{th} Collective Index Name)*** - the name assigned to a chemical substance that describes it in terms of its molecular composition;
- ***EPA Chemical Identifier*** - The unique number assigned (in CRS) when CAS number not available; and
- ***EPA Chemical Registry Name*** - the name EPA has selected as the name to be commonly used by EPA in referring to a chemical substance.

The NEI has its own lookup table of pollutant codes for the 188 CAA-listed HAPs. Most of these codes are CAS (Chemical Abstract Service) numbers. To comply with the Chemical Identification Standard, these NEI pollutant codes must be mapped to the CRS IDs. OEI is currently reviewing the NEI HAP dictionary and will assign CRS IDs to all NEI pollutant codes. Since each CRS ID is associated with all of the other data elements listed above, once the CRS mapping is completed, we will be able to associate the NEI pollutant codes to all of the required data elements (CAS #, Chemical Substance Systemic Name etc.) via a lookup table.

2.4 Facility Identification Data Standard

The facility identification data standard consists of core data elements that properly identify the location, the affiliated organizations, individual business activities, and the environmental interest of a facility site.[5]

To implement this standard, EFIG must map its facilities to the FRS (Federal Registry System) ID maintained by OEI, increase the length of several fields, add one field, change the name of some fields, and replace address information with locational address data. The changes to the NIF sites table as required by the standard are summarized in Table 3.

Implementation of this standard for the NEI database will also require the following:

- Adding tribal code to ALL tables in all output files (not just the sites table).
- Deleting mailing addresses from the location address field and retaining only physical address information.

This standard removes the address type code from NIF. This means that we will have to review the current addresses, determine if they are physical, mailing or other address type, and then delete all non-physical addresses. This is fairly straightforward as long as the address type code is filled in, but in many cases it is blank. Therefore, we will need to review all addresses manually and delete those that appear to be mailing or other non-physical addresses. For example, if the submitter has not filled out the address type code, but has given a "P.O. Box" as the location, and has listed a state for the address that is different from the state FIPS code, we will assume this is a mailing address and delete it.

Mapping NTI facilities to FRS IDs will be the most challenging aspect of implementing this standard. Currently, the NEI for HAPs contains an "NTI Unique Facility ID" that is assigned to all co-located sites. Two years ago, ERG mapped the NTI Unique IDs in the 1996 database to facilities in the NEI for criteria pollutants database. This assignment was submitted to OEI, which then assigned FRS IDs to NEI facilities. Since that time, the 1996 NEI for HAPs has undergone at least one round of major revision. Not only have facilities been deleted from the 1996 NEI, but many others have been added. We have also revised some facility ID assignments where it appeared that two sites are not the same facility or were, in fact, co-located sites. Furthermore, although the universe of 1999 NEI facilities overlaps with that of the 1996 inventory, they are not completely coincident. We plan to work closely with OEI on the matching of NEI facilities to FRS facilities.

2.5 Contact Standards

The contact standards provide a consistent method of describing the contact person submitting data to the NEI. These standards include point of contact, address, and communication information. The main effect of this standard is the addition of fields to the Transmittals table that describe the contact's phone number and electronic address.[5]

3.0 CONCLUSIONS

A number of steps are being taken to comply with the IQG. This includes improving filing; documenting the purposes, uses and limitations of the inventory; devising a rating for each emission estimate; and tracking changes in the NEI over time at a detailed level. The overall goal is to increase users' understanding of the NEI, improve its quality, and ensure that qualified third parties could replicate the inventory.

The data standards are an agency wide mechanism to make data more consistent and searchable across programs. Additional fields will be introduced to the NEI describe the facility, contacts, pollutants and latitude/longitudes. The NAICS code will replace the outdated SIC code and geographic coordinates will be better described. The main responsibility will fall on state and local agencies to populate the NAICS and latitude/longitude standards. However, when state and local agencies are unable to fill in these required fields, EFIG will provide defaulted values. EFIG is also primarily responsible for the assignment of FRS and CRS IDs.

The establishment of EPA's information quality guidelines and consistent data standards have imposed new requirements on the NEI. These measures have mandated a thoughtful and purposeful look at the manner in which data is handled, transferred, archived, merged, revised and outputted to the public. These new measures while initially requiring additional effort to implement, will in the long run, make the inventory better able to serve the purposes for which it was intended. The NEI was developed in order to determine if the Maximum Achievable Control Technology (MACT) program and other Clean Air Act (CAA) programs are successful in reducing emissions and human health risk due to HAP emissions. If there is greater confidence in the quality and sources of the data, then there can be greater confidence in the modeling and risk assessments undertaken to assess the CAA programs.

REFERENCES

1. Office of Management and Budget (OMB), 2002. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies. http://www.whitehouse.gov/omb/fedreg/reproducible.html

2. U.S. Environmental Protection Agency (EPA), 2003. National Emissions Inventory Input Format. http://www.epa.gov/ttn/chief/nif/index.html

3. U.S. Environmental Protection Agency (EPA), 2003. Environmental Data Registry. http://oaspub.epa.gov/edr/epastd$.startup

4. U.S. Environmental Protection Agency (EPA), 2002. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency. http://www.epa.gov/oei/qualityguidelines/

5. U.S. Environmental Protection Agency (EPA), 2003 Environmental Data Registry. EPA Final Data Standards. http://oaspub.epa.gov/edr/epastd$.startup

6. U.S. Census Bureau. Development of NAICS. http://www.census.gov/epcd/www/naicsdev.htm

7. U.S. Census Bureau. North American Industry Classification System -- Revisions for 2002. http://www.census.gov/epcd/naics02/

KEYWORDS

National Emission Inventory (NEI)
Information Quality Guidelines
NEI Input Format (NIF)
EPA Data Standards

Table 1. Latitude/longitude standards.

Latitude/Longitude Standard	Change	Description	Comments
Latitude Measure	Rename field	Y Coordinate - The measure of the angular distance on a meridian north or south of the equator.	+78.123456 The number of decimal positions recorded is determined by the precision of the measurement.
Longitude Measure	Rename field	X Coordinate - The measure of the angular distance on a meridian east or west of the prime meridian.	-123.234561 The number of decimal positions recorded is determined by the precision of the measurement
Source Map Scale Number	Add field	The number that represents the proportional distance on the ground for one unit of measure on the map or photo.	Only used when a map has been used to determine latitude/longitude. e.g. 125,000
Horizontal Collection Method Code	Add field	Method used to determine the latitude and longitude coordinates for a point on the earth.	e.g., 001 = address-matching house number, 018 on interpolation-map, 028 = Global Positioning Method, with unspecified parameters.
Horizontal Accuracy Measure	Add field	The measure of the accuracy (in meters) of the latitude and longitude coordinates.	
Horizontal Reference Datum Code	Add field	The code that represents the reference datum used in determining latitude and longitude coordinates.	001 = North American Datum of 1927 002 = North American Datum of 1983 003 = World Geodetic System of 1984
Reference Point Code	Add field	The code that represents the place for which geographic coordinates were established.	e.g. 101 = Entrance point of a facility or station.; 105 = Point where substance is processed, treated, settled, or stored.; 106 = Point where a substance is released.
Coordinate Data Source Code	Add field	The code that represents the party responsible for providing the latitude and longitude coordinates	e.g. EPA Headquarters, a state agency, tribal organization, EPA regional office etc.

Table 2. Geocoder default flags and default values for latitude/longitude standard.

Code	Description	Source Map Scale	Horizontal Collection Method Code & Description	Horizontal Reference Datum	Horizontal Accuracy (meters)	Coordinate Data Source Code
Exact	Match is to within a unique intersection or within a single side of a single street block.	24000	002 - Determination method based on address matching-block face.	001 - North American Datum of 1927	12	080 or 084*
Near	Match is to a single street block but the correct placement within block is unknown.	24000	003 - Determination method based on address matching-street centerline.	001 - North American Datum of 1927	50	080 or 084*
Zipcode+2	Match to a 5-digit zip code, plus the first two digits of the 4-digit extension.	24000	038 - Determination method based the center of an area defined by the 5-digit ZIP code and its 2-digit geographic segment extension.	001 - North American Datum of 1927	100	080 or 084*
Zipcode5	Match to a 5-digit zip code.	24000	026 - Determination method based on zipcode-centroid.	001 - North American Datum of 1927	10000	080 or 084*
Zipcode3	Match is to a 3-digit zip code.	24000	021 - Determination method based on interpolation-other.	001 - North American Datum of 1927	1000	080 or 084*
SCF3	Match to multiple 3-digit zip codes based on postal service Sectional Center Facility (SCF).	24000	021 - Determination method based on interpolation-other.	001 - North American Datum of 1927	1000	080 or 084*
Ambig	Match is to multiple street segments.	24000	007 - Determination method based on address matching-other.	001 - North American Datum of 1927 001	20000	080 or 084*
Countycent	County centroid.	24000	021 - Determination method based on interpolation-other; 030 - based on a digital map source (TIGER).	001 - North American Datum of 1927	N/A	080 or 084*

* Coordinates are derived from USPS, Census Bureau Tiger server, or Eagle's TeleAtlas. These correspond to codes 080 (government agency) and 084 (contracting organization).

Table 3. Facility identification standard.

Facility ID Standard	Change	Description
Country Name	Increase field length from 20 to 40 characters.	The name that represents a primary geopolitical unit of the world.
County and State FIPS	Combine state and county FIPS codes into one field.	The code that represents the county or county equivalent and the state or state equivalent of the United States.
Facility Registry Identifier	Change name of field from Federal Facility ID to Facility Registry Identifier.	The identification assigned by the EPA Facility Registry System to uniquely identify a facility site.
Facility Site Name	Increase field length from 50 to 80 characters.	The public or commercial name of a facility site.
Locality Name	Increase field length from 30 to 60 characters.	The name of the city, town or village or other locality, when identifiable, within whose boundaries the facility site is located.
Location Address	Delete address type code. Replace Streetline 1, Street Line 2 and Street Line 3 with one field (location address).	The address that describes the physical (geographic) location of the front door or main entrance of a facility site, including urban-style street address or rural address.
Location Zip Code	Increase length from 10 to 14 characters.	The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service.
Organization Dun and Bradstreet Number	Decrease length from 10 to 9 characters.	The Data Universal Numbering System (DUNS) number assigned by Dun and Bradstreet to uniquely identify business establishments.
State Facility Identifier	Change name of field from Site ID to State Facility Identifier.	The unique identification number used by a state to identify a facility size.
Tribal Code	Add field for Tribal Code.	Code that represents American Indian Tribes and Alaskan Native entities.