Air Emission Inventory Reporting through EPA's Central Data Exchange

Lee Tooly, Sally Dombrowski
U.S. Environmental Protection Agency, Office of Air and Radiation, Emission Factor and Inventory Group, Mail Code MD-14, Research Triangle Park, NC 27711
tooly.lee@epamail.epa.gov
dombrowski.sally@epamail.epa.gov

Jeffrey Wells, Chris Clark
U.S. Environmental Protection Agency, Office of Information Collection, Collection Services Division, Mail Code 2823, 1200 Pennsylvania Avenue, Washington, DC 20460
wells.jeffrey@epamail.epa.gov
clark.chris@epamail.epa.gov

Roy Chaudet
Logistics Management Institute, 2000 Corporate Ridge, McLean, VA 22102-7805
rchaudet@lmi.org

ABSTRACT

The States and U.S. Environmental Protection Agency (EPA) are implementing electronic reporting as part of the National Environmental Information Exchange Network (Network). The Network will improve collection, management, and sharing of environmental information among States, Tribes, localities, and EPA. An important component of this network is the central point through which data will flow into and out of EPA. This central point is the agency's Central Data Exchange (CDX).

EPA is conducting initial testing of the CDX prototype using a limited set of reports and trading partners. Air emissions inventory submissions (AEI) are among the first of such reports to go through CDX pilot testing. Three pilot tests for submitting AEI data were conducted:

1) *Web- NIF flat file test*—air emission inventory flat files in the National Emission Inventory (NEI) Input Format (NIF) by uploading via the NEI Submittal Web form,

2) *ADR-XML test*—Active Data Retrieval (ADR) of AEI eXtensible Markup Language (XML) files, and

3) *Web-XML test*—XML files by uploading via the NEI Submittal Web form.

These pilot tests achieved the following results:

- All six state/local participants successfully filled out the submittal form and uploaded their zipped NIF flat files and XML files through the CDX.

- The test demonstrated the successful retrieval of an XML file from the Pennsylvania Department of Environmental Protection (DEP) via Active Data Retrieval.

- EPA's Emission Factor and Inventory Group (EFIG) successfully downloaded all submitted files from the CDX staging server.

- The NIF flat files received by EFIG were the same as the flat files submitted by the participants.

- The NIF 2.0 files, translated from the submitted XML files by CDX, were successfully read (NIF reader) by EFIG.

- The tests successfully demonstrated the first use of an electronic version of the NEI Submittal Web form, whereby data from the form could be directly loaded into EFIG's tracking system.

- The tests successfully demonstrated the first use of secure socket layer (SSL)-secured transmissions through the CDX and the first CDX use of mutual authentication via SSL for the flat file transmissions.

Pilot test participants provided positive feedback, constructive comments, and recommendations for improving both the submittal Web form and the XML file structure as described in the document type definition (DTD). In addition, options for optimizing the applications and the CDX system infrastructure to process large XML files are currently being evaluated and tested for the production system. As a result of these pilot tests and system refinements currently underway, EFIG plans to allow submission of NIF flat file and XML production data through the CDX beginning June 2001.

INTRODUCTION

The CDX will serve as EPA's enterprise-wide portal to the agency's information network and as the EPA's node on the environmental data exchange Network with the States.[1,2,3] The CDX will accept electronic and paper compliance submissions from industry, States, EPA systems, and other stakeholders. Once submissions are received, the CDX will perform the following functions:

- Scan the files for viruses
- Authenticate the source of the submission
- Archive the inbound data
- Acknowledge receipt of the submission
- Log the transaction
- Transform the data into a format required by the receiving EPA system
- Distribute the data to the appropriate EPA system, and to other stakeholders, as needed.

Initial testing and production reporting through EPA's prototype CDX is being conducted using a limited set of reports and trading partners. Air emissions inventories are one of the first such reports being pilot tested through the prototype CDX. For air emission inventories, CDX will receive submissions from state and local air programs.

Two electronic reporting formats were tested in the air emission inventory CDX pilot tests: National Emission Inventory (NEI) Input Format (NIF) flat files[4] and XML files. The NIF is a standard data transfer transaction defined and implemented by EFIG. The data content of the NIF has also been constructed in XML using the Document Type Definition (DTD) and data dictionary. Transmission of these formats were tested in the following pilots:

- *Web-NIF flat file test.* Submitting NIF flat files with the NEI Submittal Web form.
- *ADR-XML test*. Active Data Retrieval (ADR) of XML files.
- *Web-XML test.* Submitting XML files with the NEI Submittal Web form.

This paper summarizes the results of the three pilot tests for electronic air emissions inventory (AEI) submissions by state and local air programs through the CDX. The results of these tests are being used to develop CDX capabilities for submitting production data to EPA beginning June 2001.

TEST OVERVIEWS

An overview of the three pilot tests conducted and the results are presented in this section.

Web-Flat File Test

The NIF flat file pilot test was conducted on November 29 and 30, 2000. NIF flat file submissions were prepared by the state and local programs and electronically submitted to EFIG through the CDX. The following state and local air programs participated in the NIF flat file pilot test:

- Air Division, Alabama Department of Environmental Management
- Emissions Reporting and Assessment Unit, Air Quality Division, Michigan Department of Environmental Quality (two participants)
- West Virginia Office of Air Quality
- Emissions Inventory Unit, Maricopa County Environmental Services.

Figure 1 provides an overview of the data flow for this pilot test. The submission process involved accessing the CDX Web site via a secure socket layer (SSL) session, completing the National Emissions Inventory (NEI) Submittal Web form, and uploading the AEI file. The NEI Submittal Web form (see Appendix A) was pre-populated with submitter information from an internal CDX database populated with registration information provided by EFIG. The procedural steps test participants used to make submissions are shown in Figure 2.

Figure 1. Summary Data Flow for the NIF Flat File Pilot Test

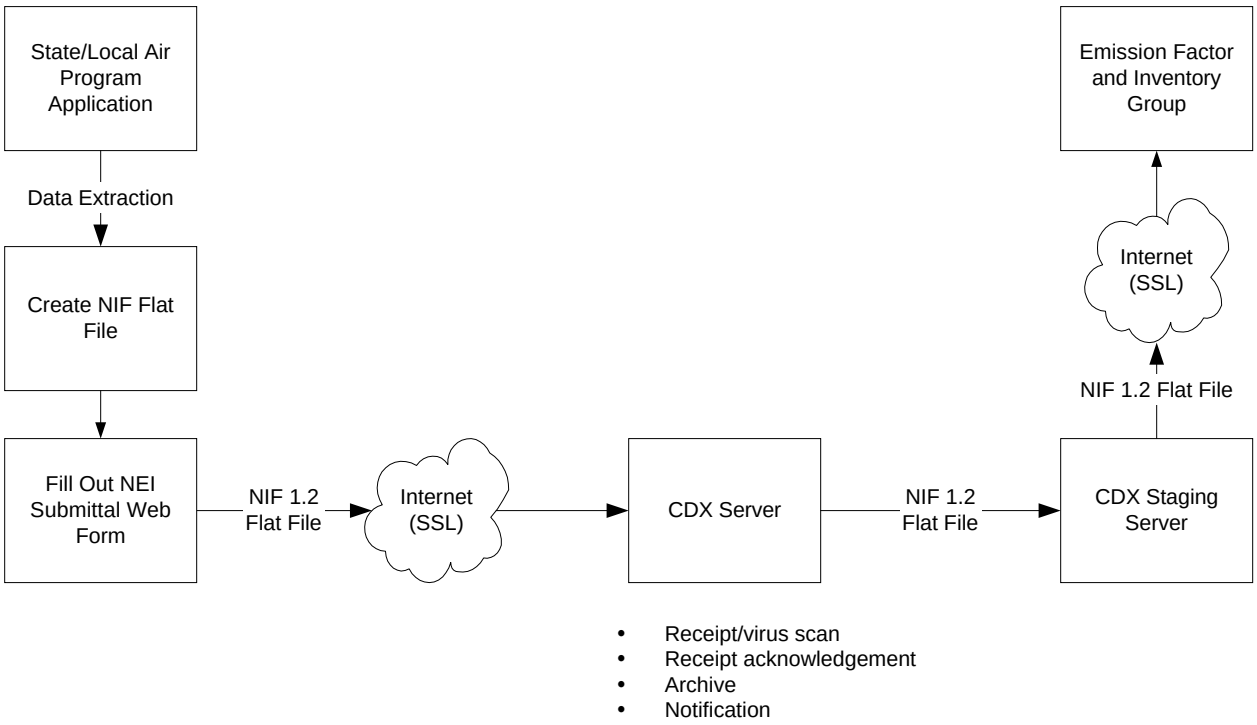
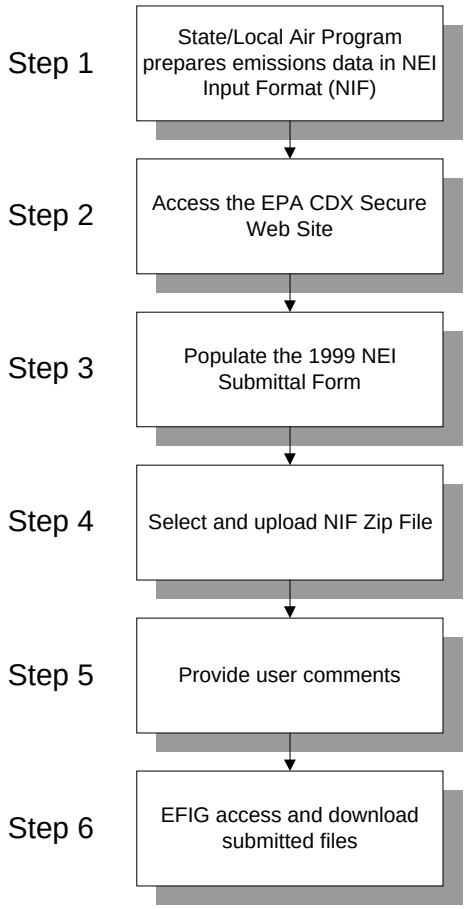


Figure 2. Test Submission Procedures



Once the participants submitted their files, the CDX received the submissions, scanned the files for viruses, logged the transactions, acknowledged receipt of the submissions, archived the submissions, notified EFIG that submissions had been processed, and distributed the files, as received, to EFIG. In turn, EFIG conducted a quality control (QC) review of the files received to ensure the files transmitted to the CDX were the same as those received by EFIG. Because only test data were transmitted, EFIG did not perform full quality assurance (QA) and data augmentation.

Summary of Results

The results of the NIF flat file test are as follows:

- All five participants successfully filled out the submittal Web form and uploaded their NIF flat files through the CDX. The upload time ranged from a few seconds to two minutes—approximately the same amount of time required for file transfer protocol (FTP) transmissions in previous years.

- EFIG successfully downloaded the submitted files from the CDX staging server. On the basis of EFIG's QC review, there appeared to be no corruption of files sent through the CDX.

- The test successfully demonstrated the first use of an electronic version of the NEI Submittal Web form, whereby data from the form could be directly loaded into EFIG's tracking system. The submittal form was pre-populated with registration information; it provided simple checks for complete and valid form information.

- The pilot test successfully demonstrated the first use of SSL-secured transmissions through the CDX and the first CDX use of mutual authentication via SSL. (An X.509 server certificate was used to authenticate the CDX server to the users. For mutual authentication, a Verisign Certificate was downloaded and issued to Steve Edick of the Michigan Department of Environmental Quality.)

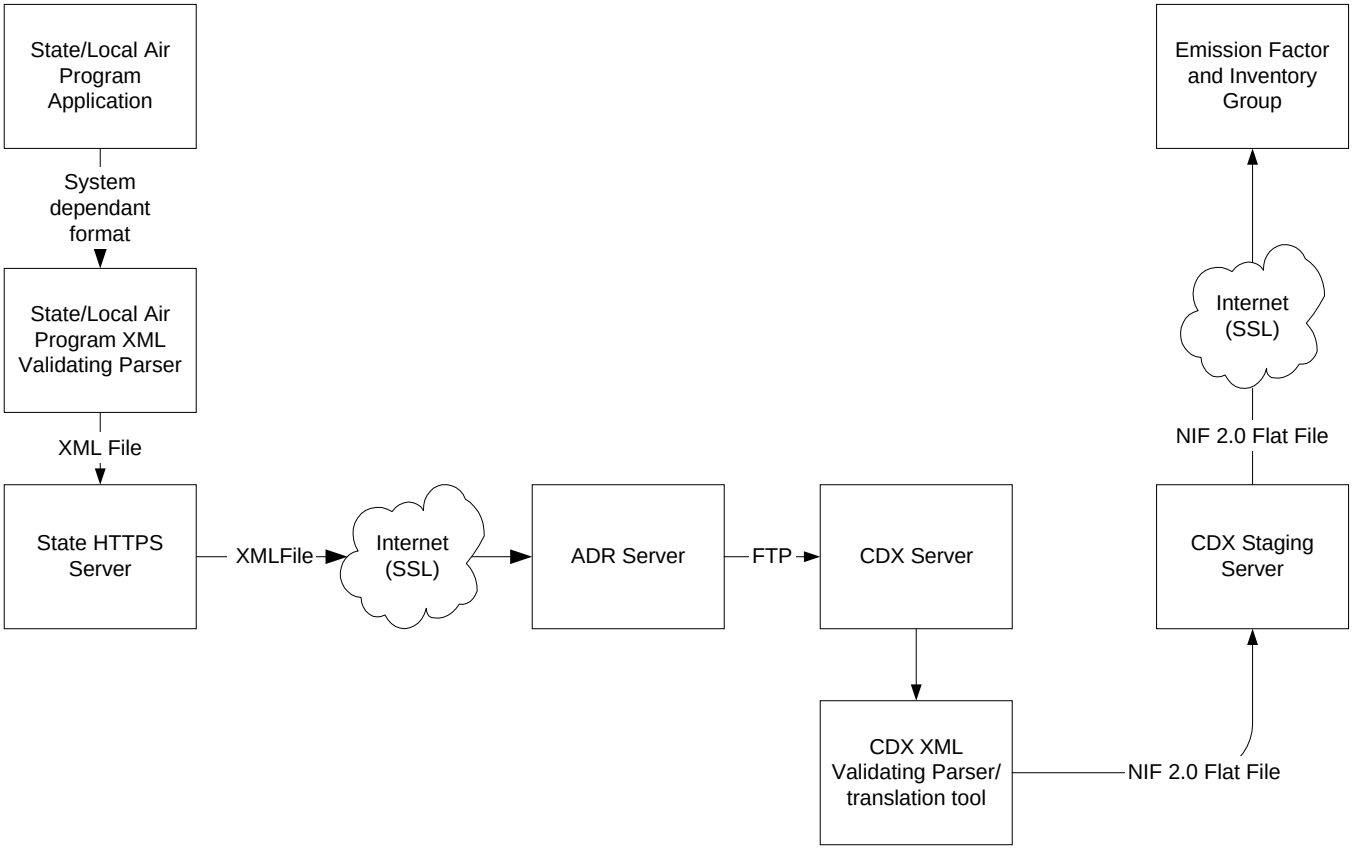
ADR-XML Test

Active data retrieval is the Central Data Exchange process that polls and regularly collects new data from remote servers across the Internet. The ADR-XML pilot test consisted of actively retrieving air emissions data in an XML file from the Pennsylvania DEP, processing through CDX, and distributing to EFIG.

The ADR-XML pilot test was conducted on February 22, 2001. A summary of the data flow for the pilot test is shown in Figure 3. The Pennsylvania DEP mapped air emissions data (nonroad source file for one county) to an XML file. The data structure and order for these XML files are defined by a DTD and data dictionary provided by EPA. PA DEP placed their XML file on the DEP server. EPA retrieved the data during an SSL session (using the ADR software, EC Knowledge), and sent the XML file to CDX via file transfer protocol (FTP).

Once CDX received the submitted data, the file was scanned for viruses, the transaction was logged, receipt of the submission was acknowledged, and the submission was archived. The submitted file was then parsed, validated, and translated into an NIF 2.0 flat file. The translated, outbound file was archived and the transaction was logged.

Figure 3. Summary Data Flow for the ADR-XML Pilot Test



When the file had been processed through CDX, EFIG was notified via e-mail. EFIG securely accessed the file during an SSL session and loaded the NIF 2.0 flat file into the NIF reader to verify the correct format. Because only test data were transmitted, EFIG did not perform full QA and data augmentation.

Summary of Results

The results of the ADR-XML pilot test are as follows:

- EPA successfully retrieved an air emission data XML file, processed it through the CDX, and distributed the file to EFIG for processing.
- The XML file was successfully validated and translated into an NIF 2.0 flat file during CDX processing.
- EFIG successfully downloaded the submitted files from the CDX staging server and loaded the test file into the NIF reader.
- The pilot test successfully demonstrated the first use of active data retrieval of an AEI XML file.

When loading the translated flat file, EFIG noted that “absolutely no problems were encountered.” EPA continued further testing of XML files in the Web-XML pilot test.

Web-XML Test

The Web-XML pilot test was conducted on March 13 and 14, 2001. Air emission inventory XML files were prepared by the following state air programs for testing:

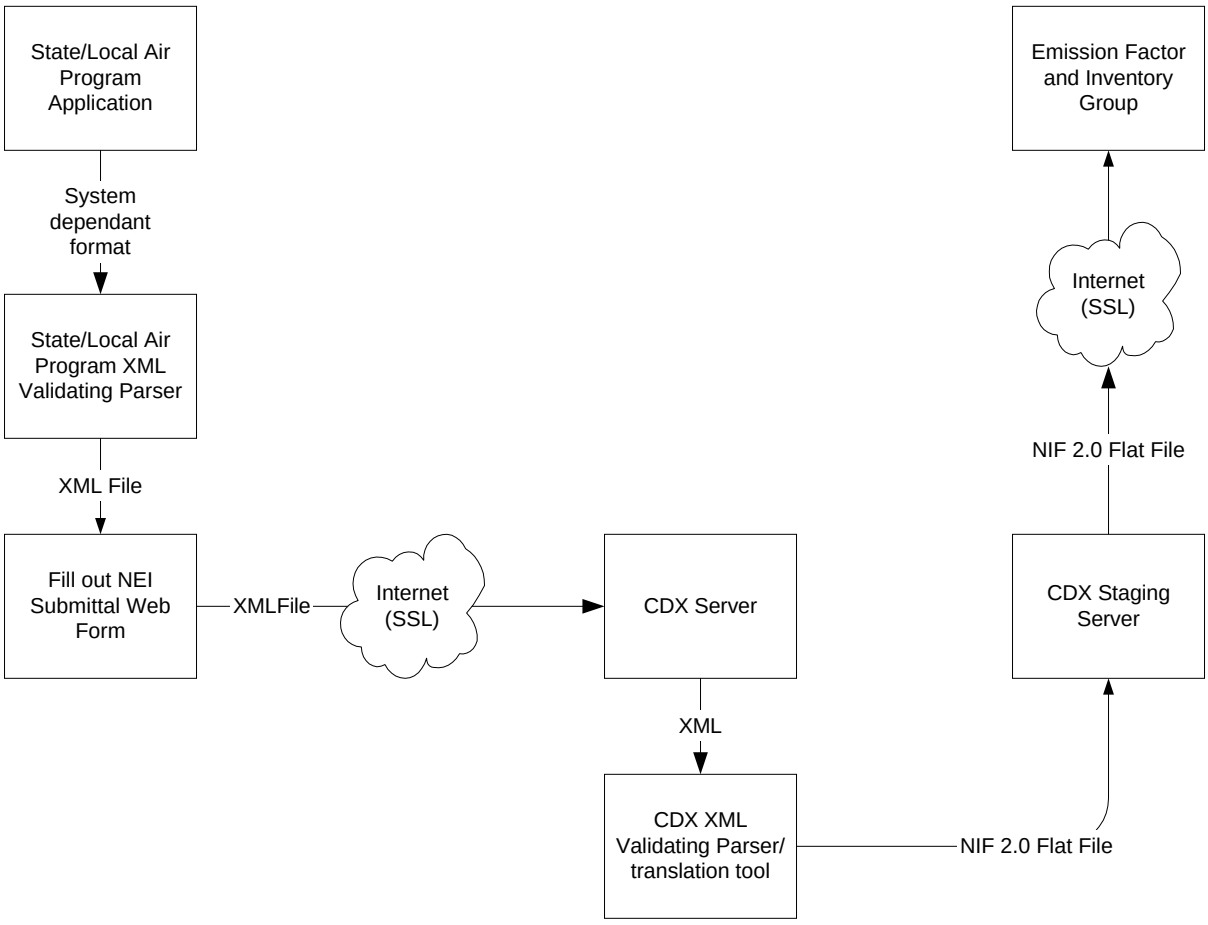
- Illinois EPA Bureau of Air (a 453 KB point source file and nonroad source file)
- Pennsylvania Department of Environmental Protection, Bureau of Air Quality (a 400 KB nonroad source file).

Figure 4 provides an overview of the data flow for this pilot test. The submission process involves the state air programs preparing their XML file, accessing the CDX website via SSL session, completing the NEI Submittal Web form, and uploading the AEI file.

The submitted XML files were received by CDX and scanned for viruses; the transaction was logged, receipt acknowledged, and the file was archived. The submitted files were parsed, validated, and translated into an NIF 2.0 flat file. The translated outbound file was archived and the transaction was logged.

EFIG was notified via e-mail when the files were processed through CDX and available on the staging server. EFIG securely accessed the files during a SSL session and loaded the NIF 2.0 flat file into the NIF reader to verify the correct format. Because only test data were transmitted, EFIG did not perform full quality assurance and data augmentation.

Figure 4. Summary Data Flow for the Web-XML Pilot Test



Summary of Results

The results of the Web-XML pilot test are as follows:

- All participants successfully filled out the submittal Web form and uploaded their XML files through the CDX.
- The XML files were successfully validated and translated into an NIF 2.0 flat file during CDX processing.
- EPA successfully downloaded the submitted files from the CDX staging server and loaded the test file into the NIF reader.
- The pilot test successfully demonstrated the use of SSL-secured transmissions through the CDX.

During this test a key concern became apparent, the file size for the emission inventory XML files can affect the processing performance of the current prototype CDX. We are currently looking at options for optimizing the XML processing application and the infrastructure for processing files as large as tens to hundreds of megabytes.

CONCLUSIONS AND RECOMMENDATIONS

Six state air programs and one local air program successfully submitted emission inventory data in NIF flat files and XML files through the CDX to EFIG via the Web. The Central Data Exchange, EPA's portal on the Network, processed the NIF flat file submissions and the XML submissions. In addition, EPA successfully retrieved an air emission XML file from the Pennsylvania DEP and moved it to CDX for processing. EPA's EFIG successfully downloaded the files from the CDX staging server and found that the NIF flat files received were the same as the files submitted. The XML files submitted were loaded into the NIF reader.

As a result of these pilot tests, EFIG plans to allow submission of NIF flat file and XML production data through the CDX via the NEI Submittal Web form or ADR in June 2001. Implementation of the following recommendations and application and infrastructure optimization will only enhance the submission process for production data.

Pilot participants provided positive feedback, constructive comments, and recommendations for improving the submittal Web form. Participant recommendations included the following: adding an upload progress indicator, clarifying the use of submittal number in Section II of the form, adding a label to the size cutoff in Section V of the form, and providing online help (hyperlinks).

Recommendations on how to improve the document type definition (DTD) for the XML file submissions were provided by David "Buzz" Asselmeier. His comments recommended revisions to the DTD that take the next step beyond initial "proof-of-concept" testing to a production environment:

- *Eliminate repeating key values*. Key values that are repeated in the DTD (e.g. transmittal record) could be presented once and called for any transformations/translations that are being performed.
- *Specify optional data tags.* Data tags do not have to be provided for optional fields if that data is not being provided. Although these tags were essential for the initial testing, they do not have to be provided for production submissions.

Mr. Asselmeier also provided a sample DTD for a point source file in order to illustrate his recommendations and demonstrate a reduction in XML file size. These recommendations will be implemented as part of the overall optimization of the application and system for processing XML files.

Options for the optimizing the applications and the CDX system infrastructure to process large XML files are currently being evaluated and tested. The test results will directly impact how production data in these large files will be submitted.

REFERENCES

1. Network Blueprint Team. *Blueprint for a National Environment Information Exchange Network.* Report to the State/EPA Information Management Workgroup. October 30, 2000.

2. U.S. Environmental Protection Agency. *Information Integration Network.* Office of Environmental Information, 21 December, 2000. <http://www.epa.gov/oei/iii.htm>.

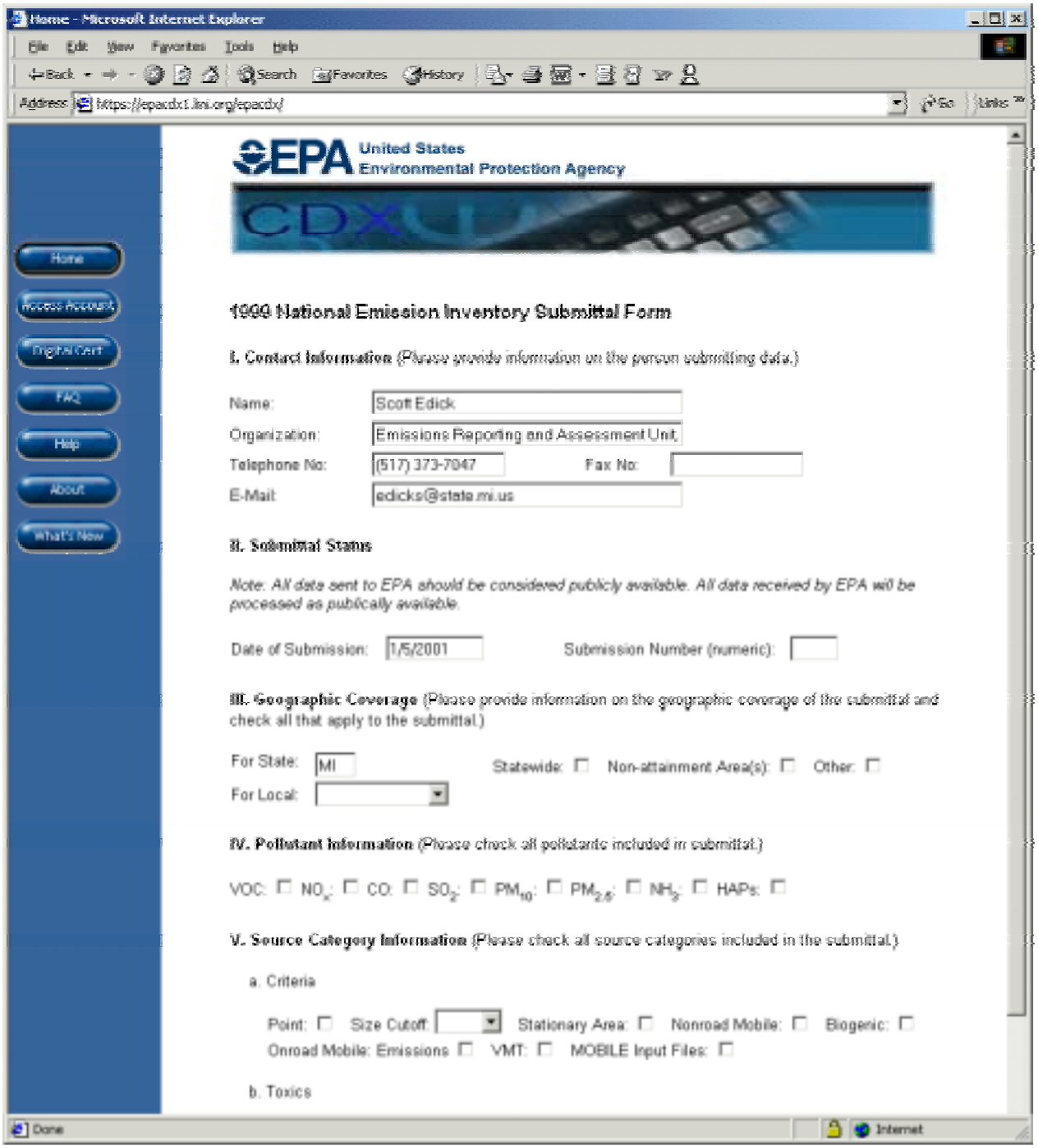
3. U.S. Environmental Protection Agency. Central Data Exchange. 19 December, 2000. <http://www.epa.gov/cdx>.

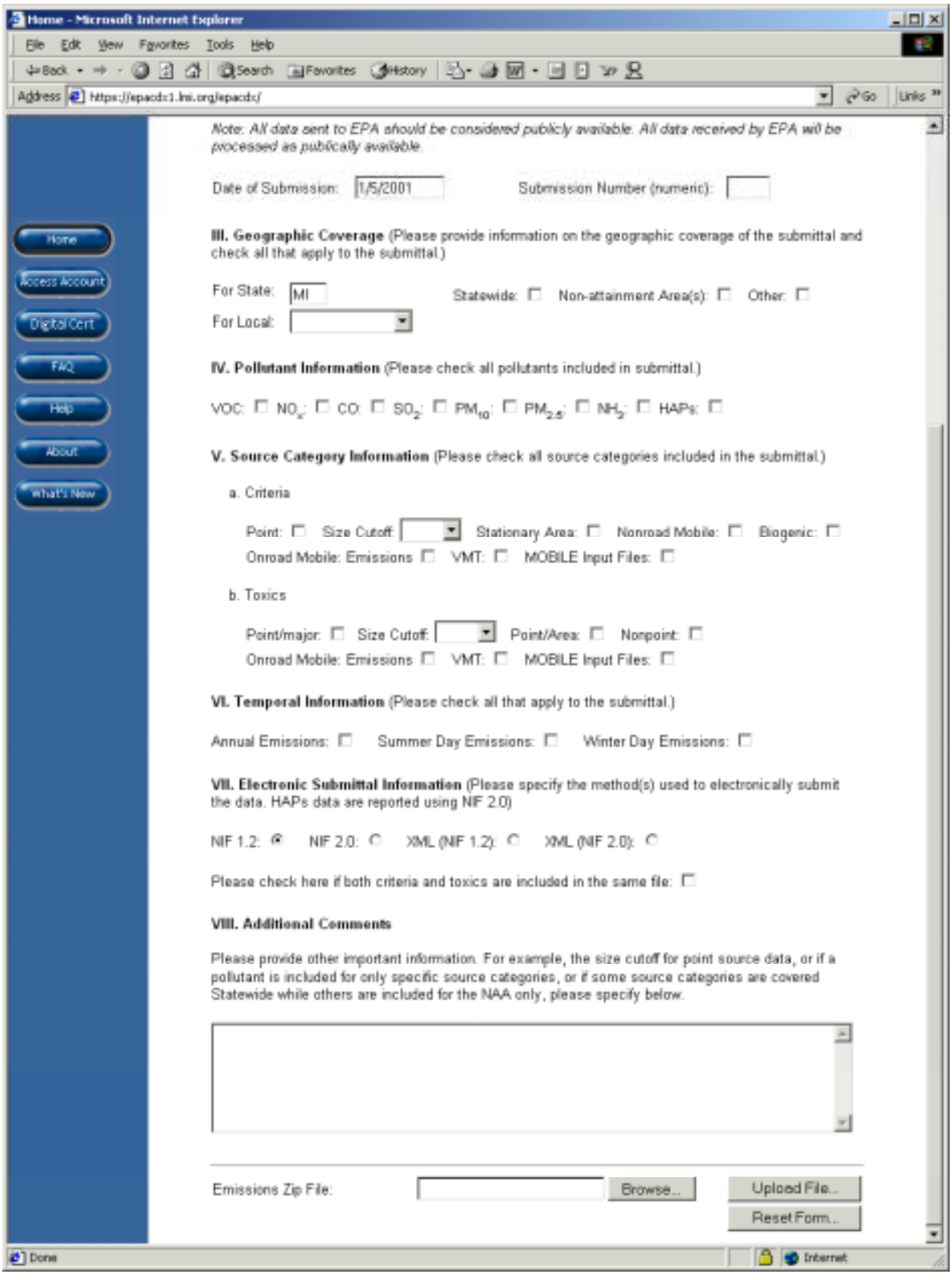
4. U.S. Environmental Protection Agency. *NIF 2.0 User Package*. Technology Transfer Network, 6 March, 2001. <http://www.epa.gov/ttn/chief/eidocs/index.html#net>.

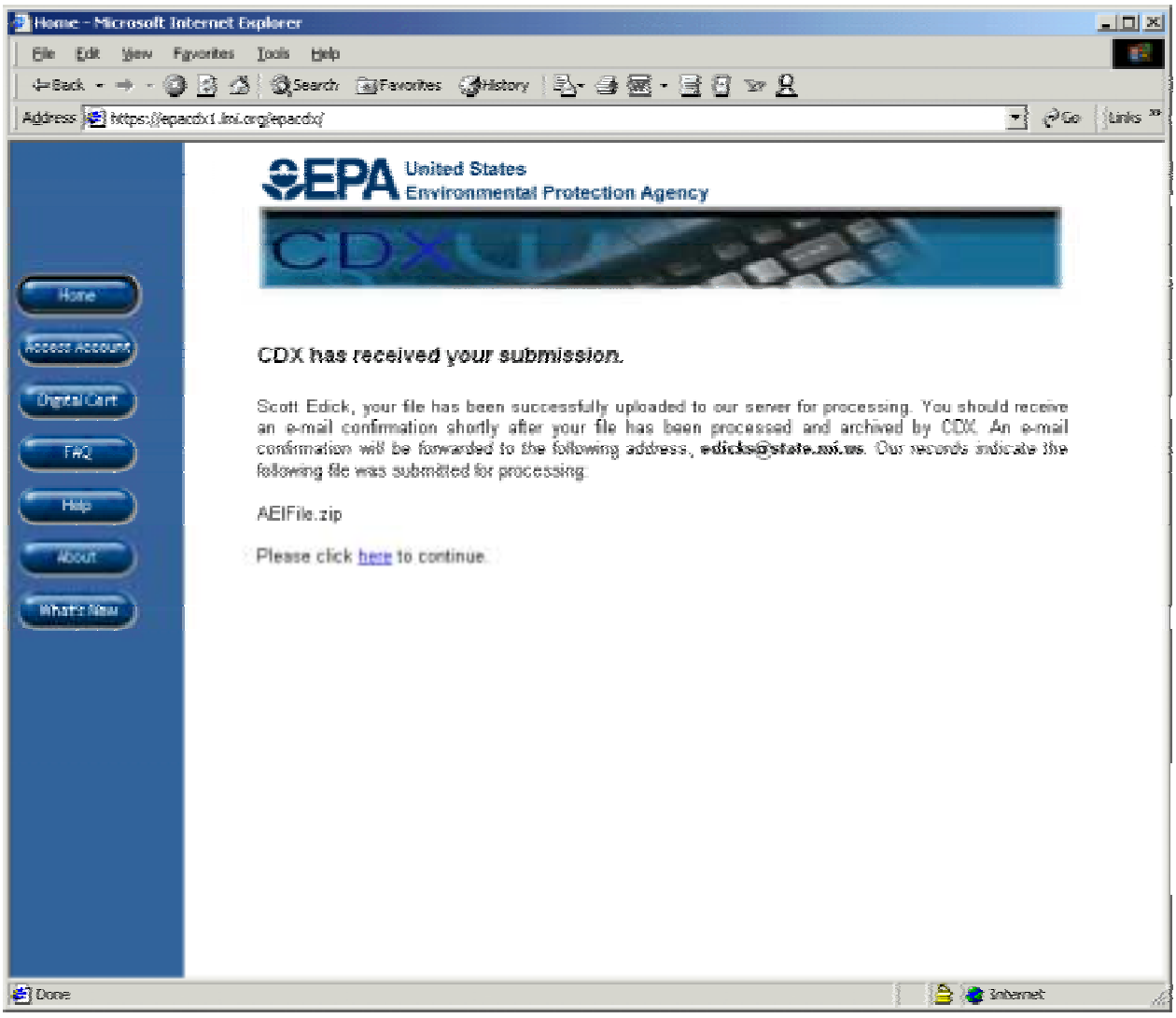
KEYWORDS

Electronic reporting
Central Data Exchange
Air emissions inventories
Extensible Markup Language
XML
NIF flat files
NIF 2.0
Secure socket layer
SSL
Document Type Definition
DTD
National Emission Inventory
NEI
NEI input format
National Environmental Information Exchange Network

APPENDIX A
NEI SUBMITTAL WEB FORM




Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address https://epacdc1.lmi.org/epacdc/
Go Links
Home
Access Account
Digital Cert
FAQ
Help
About
What's New
Note: All data sent to EPA should be considered publicly available. All data received by EPA will be processed as publically available.
Date of Submission: 1/5/2001
Submission Number (numeric):
III. Geographic Coverage (Please provide information on the geographic coverage of the submittal and check all that apply to the submittal.)
For State: MI
Statewide: Non-attainment Area(s): Other:
For Local:
IV. Pollutant Information (Please check all pollutants included in submittal.)
VOC: NOx: CO: SO2: PM10: PM2.5: NH3: HAPs:
V. Source Category Information (Please check all source categories included in the submittal.)
a. Criteria
Point: Size Cutoff: Stationary Area: Nonroad Mobile: Biogenic:
Onroad Mobile: Emissions VMT: MOBILE Input Files:
b. Toxics
Point/major: Size Cutoff: Point/Area: Nonpoint:
Onroad Mobile: Emissions VMT: MOBILE Input Files:
VI. Temporal Information (Please check all that apply to the submittal.)
Annual Emissions: Summer Day Emissions: Winter Day Emissions:
VII. Electronic Submittal Information (Please specify the method(s) used to electronically submit the data. HAPs data are reported using NIF 2.0)
NIF 1.2: NIF 2.0: XML (NIF 1.2): XML (NIF 2.0):
Please check here if both criteria and toxics are included in the same file:
VIII. Additional Comments
Please provide other important information. For example, the size cutoff for point source data, or if a pollutant is included for only specific source categories, or if some source categories are covered Statewide while others are included for the NAA only, please specify below.
Emissions Zip File:
Browse...
Upload File...
Reset Form...
Done
Internet


Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address https://epacdx1.lmi.org/epacdx/
Go Links
EPA United States Environmental Protection Agency
CDX
Home
Access Account
Digital Cart
FAQ
Help
About
What's New
CDX has received your submission.
Scott Edick, your file has been successfully uploaded to our server for processing. You should receive an e-mail confirmation shortly after your file has been processed and archived by CDX. An e-mail confirmation will be forwarded to the following address, edicks@state.mi.us. Our records indicate the following file was submitted for processing:
AEIFile.zip
Please click here to continue.
Done
Internet