

The Environmental Protection Agency Constructing the NEI XML Schema

June 10, 2004

E.H. Pechan & Associates, Inc.

Steven Riley Boone and Donna McKenzie

steve.boone@pechan.com; donna.mckenzie@pechan.com

3622 Lyckan Parkway, Suite 2002

Durham, NC 27707

Background

- EPA/Emission Factors Inventory Group (EFIG) Project
- NEI (National Emission Inventory)
- NIF (NEI Input Format)
 - Access
 - ASCII
 - <http://www.epa.gov/ttn/chief/nif/index.html>
- NEI XML format
 - [http://oaspub.epa.gov/emg/xmlsearch\\$.startup](http://oaspub.epa.gov/emg/xmlsearch$.startup)

XML

- XML
 - eXtensible Markup Language
 - A subset of SGML (Standard Generalized Markup Language)
 - Format for structured documents and data
- World Wide Web Consortium (W3C)
 - XML language standards
 - <http://www.w3.org>

XML Definitions

- XML Schema
 - Defines for XML Documents
 - Structure
 - Elements
 - Data Types
- XML Instance Document
 - Data
 - XML format
- XML Validation
 - Conformance of instance document to a schema

NEI XML Schema Construction

- NEI XML Schema Development
- Piloting the NEI XML Schema
- Creation of NEI XML Instance Documents
- Submission of NEI XML Instance Documents
- NEI XML Translator Development

NEI XML Schema Development

- NEI XML Schemas
 - Developed to be compatible with NIF V 3.0 content
 - Mapped each schema element to NIF 3.0 element
 - Applied data standards of the Environmental Data Registry (EDR)
 - <http://www.epa.gov/edr/>
 - Reviewed by TRG (Technical Resource Group) for conformance
 - <http://www.exchangenetwork.net/common/default.asp>

NEI XML Schema Development

- NEI XML Schemas
 - Reviewed by TRG (Technical Resource Group) for data and structural conformance
 - W3C Standards
 - XML Design Rules and Convention (DRC)
 - Core Reference Model (CRM)
 - Review of NEI XML 2.0

NEI XML Schema Development

- Revisions to NEI XML Schema
 - Referential Integrity
 - Key/keyref
 - Development of “Common” reference schema for common elements
 - “Flattening” of Schema

NEI XML Schema Example

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema targetNamespace="http://www.epa.gov/exchangenetwork"
  xmlns:nei="http://www.epa.gov/exchangenetwork"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified" attributeFormDefault="unqualified"
  version="3.0">
  <xsd:include schemaLocation="EN_NEI_Common_v3_0.xsd"/>
  <!-- Start of Schema Header -->
  <xsd:complexType name="TransmittalSubmissionGroupType">
    <xsd:sequence>
      <xsd:element ref="nei:TransmittalRecordTypeCode"/>
      <xsd:element ref="nei:CountyStateFIPSCode"/>
      .....
      <xsd:element ref="nei:TribalCode"/>
    </xsd:sequence>
    <xsd:attribute name="schemaVersion" type="xsd:decimal"
  use="required"/>
  </xsd:complexType>
```

NEI XML Schema Development

- EN_NEI_AreaNonroad_v3_0.xsd
- EN_NEI_Biogenic_v3_0.xsd
- EN_NEI_Onroad_v3_0.xsd
- EN_NEI_Point_v3_0.xsd
- EN_NEI_Common_v3_0.xsd

Piloting the NEI XML Schema

- Submitting Agencies
 - A subset of states
 - Iterative testing
 - Schema revisions

Creation of XML Instance Documents

- Develop XML instance documents according the NEI XML Schema
- Validated in the field during file generation
 - XMLSpy or other tools
 - Differences in validation tools
- Validated at EPA upon submittal

NEI XML Instance Document

```
<?xml version="1.0" encoding="UTF-8"?>
<BiogenicSubmissionGroup schemaVersion="3.14"
  xmlns="http://www.epa.gov/exchangenetwork"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.epa.gov/exchangenetwork
D:\XMLTRA~1\EN_NEI_Biogenic_v3_0.xsd">
  <TransmittalSubmissionGroup schemaVersion="3.14">
    <TransmittalRecordTypeCode>TR</TransmittalRecordTypeCode>
    <CountyStateFIPSCode>CSFC5</CountyStateFIPSCode>
    .....
    <TribalCode>TC3</TribalCode>
  </TransmittalSubmissionGroup>
```

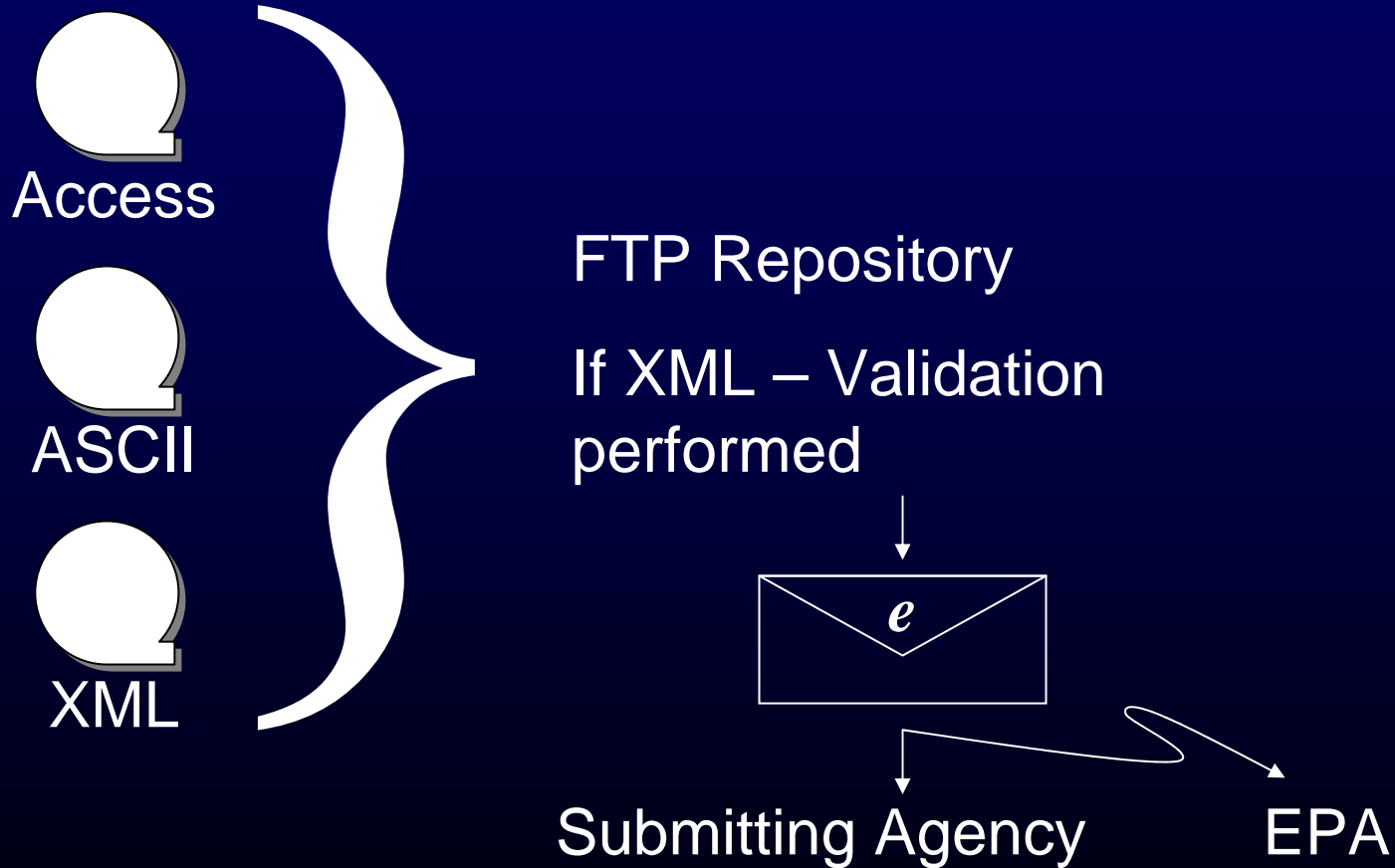
XML Data Validation

- NEI XML Schema Validation
 - Referential Integrity
 - Duplication
 - Data Structure and Format
 - Tag Naming Conventions
 - Pattern Matching
 - Data Types

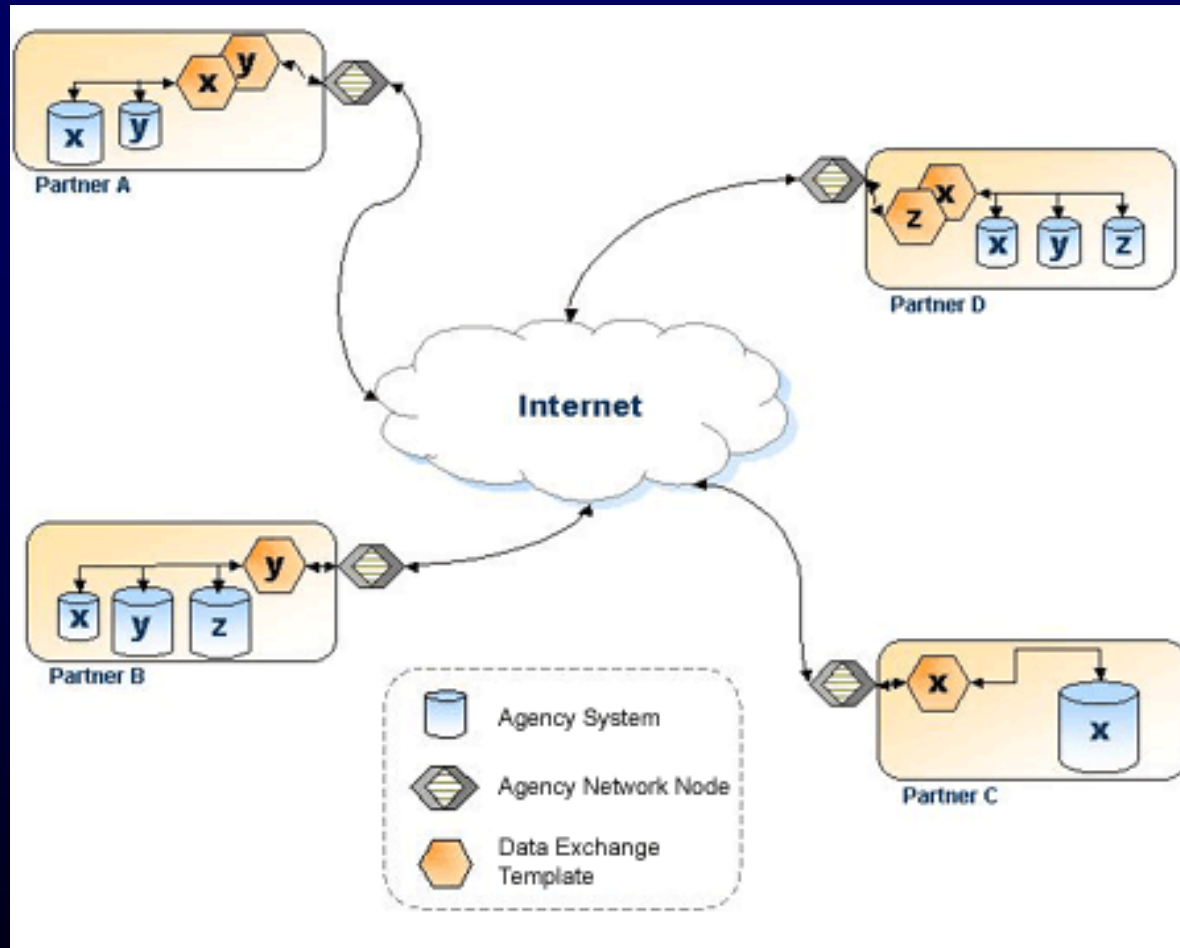
Submission of NEI XML Documents

- Emission Inventory Data Submission using the NEI XML format
 - Submission to Emission Factors Inventory Group (EFIG)
 - Direct Submit through CDX (Central Data Exchange)
 - http://cdx.epa.gov/epa_home.asp
 - ExchangeNetwork (Node to Node)
 - <http://www.exchangenetwork.net/common/default.asp>
 - <http://www.epa.gov/Networkg/library/netiplan0202.pdf>

Submission of NEI XML Documents



Exchange Network



NEI XML Translator

- NEI XML Translator Utility is used to convert NEI XML document to NIF 3.0 ASCII format
- ASCII format can be used as is or imported into a database

NEI XML Translator

- Technical specifications
 - Visual Basic 6.0
 - Microsoft XML
 - XML Stylesheets (XSLT)

NEI XML Translator

- NEI XML files can be large (100 MB and upwards)
- XML Parsing (reading the document)
 - DOM (Document Object Model)
 - Entire document is read in – dynamic access, but memory intensive
 - SAX (Simple API for XML)
 - Event driven, easier on memory, increases application logic needs
 - StaX (Streaming API for XML)
 - New, better memory efficiency than DOM, easier document access than SAX

NEI XML Translator

- NEI XML Files are read, not changed
- Combination of DOM and SAX
- SAX is used to select out a subdocument (for example, Emission Unit Submission Group)
- DOM reads the entire subdocument
- Reduces memory needs to portions of the entire file

NEI XML Translator

- Transform XML to ASCII
 - XML Stylesheets
 - Element Matching
 - Template
 - Application serves as engine for transformation
 - Applied to parsed subdocuments
 - A single Point NEI XML file – results in 8 text files in NIF ASCII 3.0 format

NEI XML Translator

NEI XML Translator

Action

XML File to Translate: C:\Program Files\NEI XML Translator\TestAppBiogenic.xml

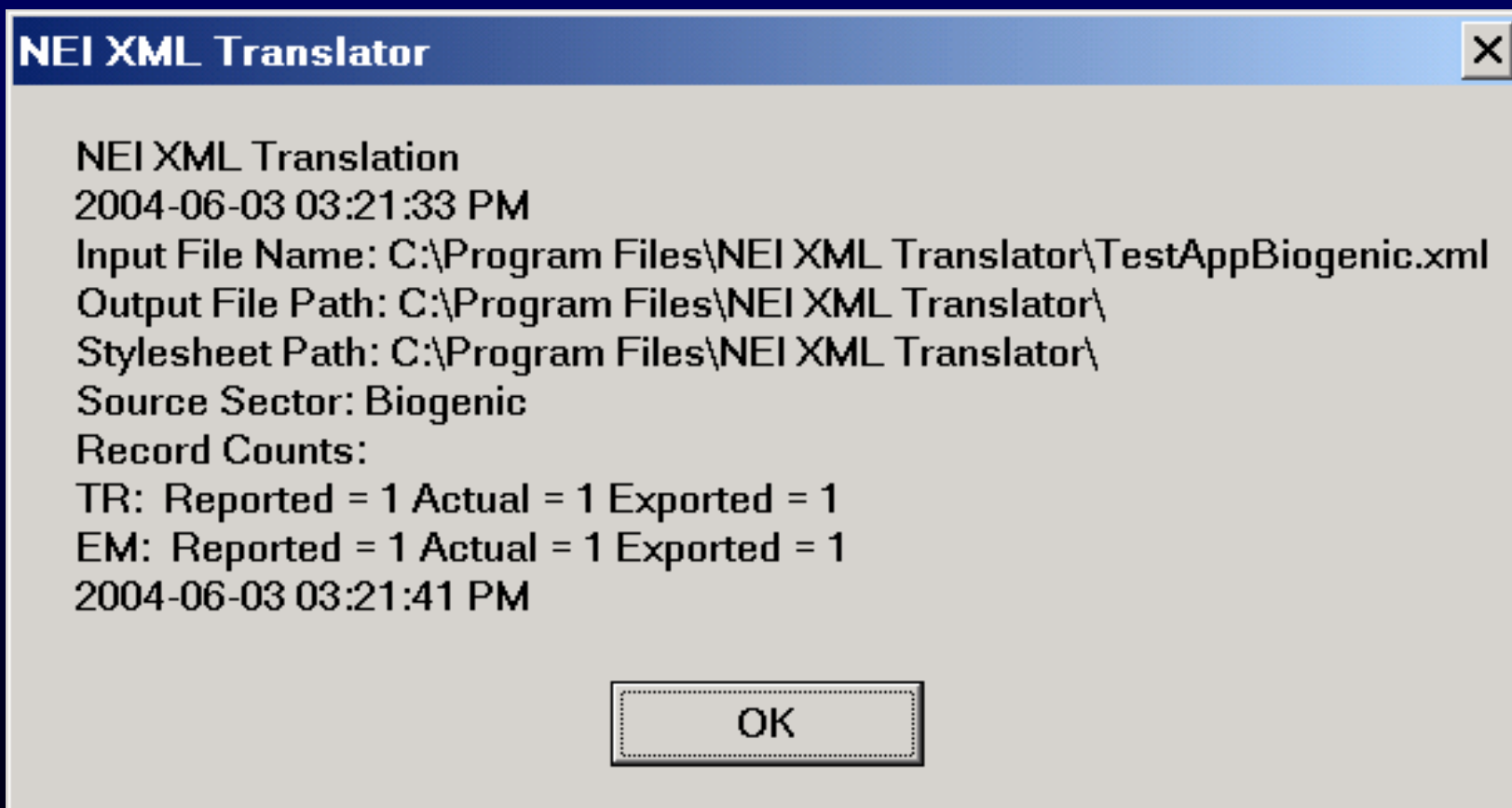
XML Stylesheet Path: C:\Program Files\NEI XML Translator\

Text File Output Path: C:\Program Files\NEI XML Translator\

Inventory Source: Biogenic

Translate

NEI XML Translator



Summary

- Schema Development
 - Element Mapping
 - Standards Conformance
 - Common Reference Document
- Schema Validation
 - Testing
 - Multiple Environments
- Translation Utility Development