

***EPA's National Emission Inventory Criteria Data
System – Functions for Streamlining Data
Processing and QA***

Steve Boone, E.H. Pechan & Associates, Inc.

NATION EMISSION INVENTORY (NEI)

- Annual Compilation
- Emission Trends
- Modeling & Regulatory Analysis
- National Air Pollutant Trends Update

NEI (CRITERIA)

- Oracle Database –
Data Warehouse
- NIF Submittals
 - Access
 - ASCII
 - XML  

SECTORS

- Stationary Point
- Area/Non-point
- Onroad Mobile
- Nonroad

TOOLS

- Back end database
 - Oracle

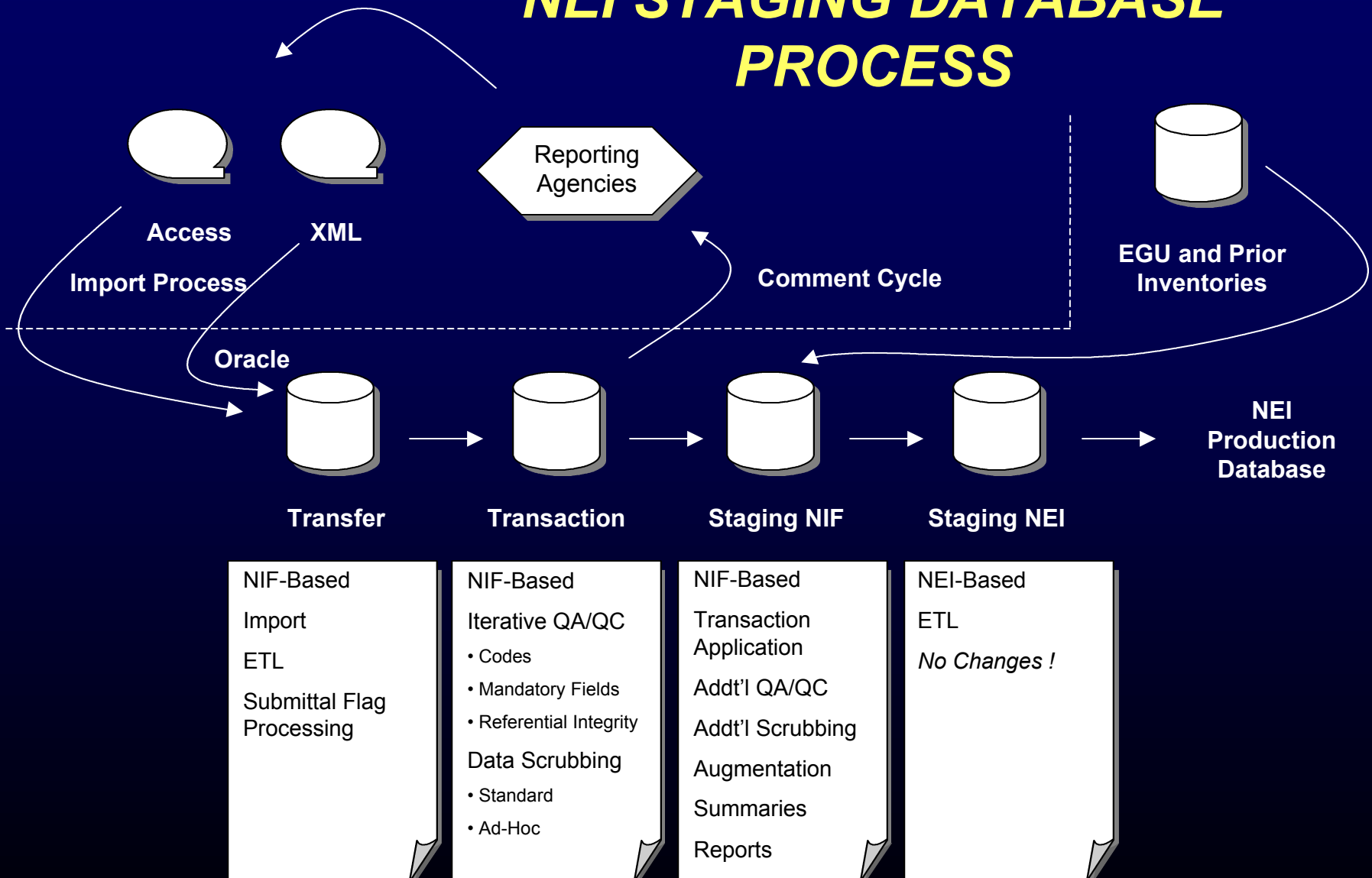
TOOLS

- Front end applications
 - Microsoft Access
 - Import
 - Export
 - Some Reports
 - Java
 - XML Transfer

TOOLS

- Front end applications (*continued*)
 - Oracle Tools
 - Enterprise Manager
 - Database administration
 - SQL*Plus
 - Script Execution
 - SQL*Loader
 - Import
 - SQL*Reports

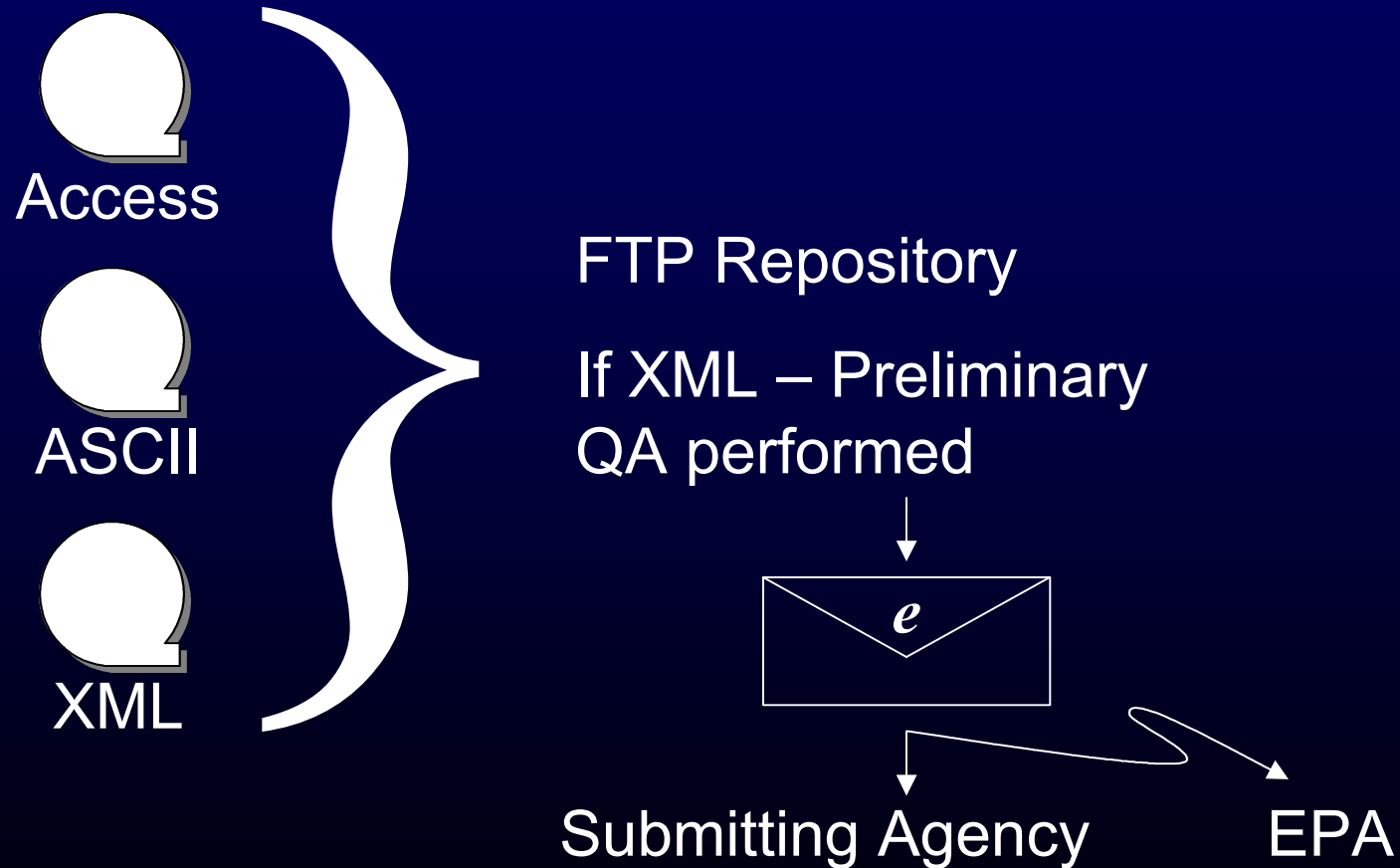
NEI STAGING DATABASE PROCESS



STREAMLINING

- Central Data Exchange (CDX) 
- Extensible Markup Language (XML)
- Quality Assurance/Quality Control (QA/QC) 
- Scrubbing 
- Audit/Logging/Security 
- Versioning 
- Efficiency 

CDX



XML

- JAVA
- Simple API for XML (SAX) Parser
- XML Software Descriptor (XSD) - Schema

```
<xsd:element name = "SiteSubmissionGroup" type = "SiteSubmissionGroupType"/>
    <xsd:annotation>
        <xsd:documentation>Site record - schema only</xsd:documentation>
    </xsd:annotation>

<xsd:complexType name = "SiteSubmissionGroupType">
    <xsd:sequence>
        <xsd:element ref = "RecordTypeCode"/>
        <xsd:element ref = "SiteKeyFieldsGroup"/>
        <xsd:element ref = "TransactionSubmittalCode" minOccurs = "0"/>
        <xsd:element ref = "SiteDetails"/>
    </xsd:sequence>
</xsd:complexType>

<xsd:complexType name = "SiteKeyFieldsGroupType">
    <xsd:sequence>
        <xsd:element ref = "CountyStateFIPSCode"/>
        <xsd:element ref = "StateFacilityIdentifier"/>
        <xsd:element ref = "TribalCode" minOccurs = "0"/>
    </xsd:sequence>
</xsd:complexType>
```

QA/QC

- Enhanced QA/QC each year
- Mandatory Items
- Check codes for validity
- Cross field checks
 - e.g., Release Point Diameter, Gas Flow Rate and Release Point Exit Velocity; County Code and Zip Code
- Information validity
 - e.g., SCC, pollutant, controls, efficiency

SCRUBBING

- Data validation triggers on production database will not allow certain invalid data to be inserted
- Change what can be determined as incorrect and use known correct data
- Change invalid or out of range values to acceptable values
- Change nulls in mandatory fields to unknown values
 - e.g., SIC to 9999

AUDIT/LOGGING/SECURITY

- Records all changes, keeps old records
- Enable reversing or undoing the process quickly and taking another
- Facilitates recovery from hardware or database failure
- Self-documents the process and facilitates making the data defensible
- Oracle imposed security matrix

VERSIONING

- Tracks multiple releases
- Self documents the differences between releases
- Facilitates retrieval of records updated between releases

EFFICIENCY

- Database tuning
- *9i*
- Multiple physical database drives
- Code tightening
- Reducing multiple pass processes

EPA's NEI Data System – Functions for Streamlining Data

- End of presentation