Cyberinfrastructure for Emissions Data & Tools

Stefan Falke, CAPITA, Washington University
Gregory Stella, Alpine Geophysics, LLC
Terry Keating, EPA
Brooke Hemming, EPA

15th Annual Emission Inventory Conference
Reinventing Inventories: New Ideas in New Orleans

May 17, 2006
Project Overview

Objectives: advance the implementation of the Networked Environmental Information Systems for Global Emissions Inventories (NEISGEI), an EPA initiative to develop a web-based global air emissions inventory network to provide
  • access to distributed emission inventory data
  • tools for data processing and analysis
  • means for sharing data & tools
  • an environment for collaboration

Approach: Develop, test, and implement components of an air quality cyberinfrastructure using the latest advances in information technology to make multi-scale air emissions data and tools easier to find, use and integrate.
Cyberinfrastructure

Cyberinfrastructure - information sciences and technologies used to build new types of scientific and engineering knowledge environments with the goal of pursuing research and management more effectively and efficiently.

“Contemporary projects require effective federation of both distributed resources (data and facilities) and distributed, multidisciplinary expertise and cyberinfrastructure is a key to making this possible.”


(Atkins, 2004)
…is both a conceptual framework and implementation effort for the development of an integrated, distributed air emissions inventory

- Tie together data at multi-spatial and temporal scales
- Provide shared, online tools for processing and analysis
- Provide for the seamless merging, processing and analysis of Internet accessible air quality-relevant data
- Make use of existing resources – partner/link with related projects
- Build a broad-based air emissions user community: scientists, regulators, policy analysts and the public
- Create the network and toolkit via specific, task-oriented projects

An air emissions “cyberinfrastructure”
Conceptual Diagram of an Emissions Cyberinfrastructure

Data
- XML
- Emissions Inventories
- Portals
- Activity Data
- Emissions Factors
- Surrogates

Wrappers/Adapters/Standards

Data Catalogs
- Emissions Inventory Catalog
- Geospatial One-Stop

Mediators/Portals

Users & Projects
- Report Generation
- Data Analysis
- Comparison of Emissions Methods
- Model Development

Web Tools/Services
- GIS
- Spatial Allocation
- Estimation Methods
- Transport Models

Activity Data

Comparison of Emissions Methods

Emissions Inventories

Model Development

Report Generation

Emissions Factors

Data Analysis

Surrogates

GIS

Spatial Allocation

Estimation Methods

Transport Models

Users & Projects

Mediators/Portals

Wrappers/Adapters/Standards

Data Catalogs
A community resource providing access to, descriptions of, and dialogues about an array of content and services for exploring and sharing emissions data, tools and ideas.

Built using **LifeRay**, and open-source portal package

Beta version accessible through [http://www.neisgei.org](http://www.neisgei.org)
Java Portlet Specification (JSR-168), defines a standard application programming interface for J2EE (Java) based portal platforms. A portal developer can find collections of JSR-168 portlets on the web and simply embed them in their portal.

Web Services for Remote Portlets (WSRP) is an XML and web services specification that allows the remote sharing of portlets. WSRP allows portlets running on one portal to be displayed in another portal without requiring any additional programming by the portal developers.

To the end-user, it appears that the portlets are running locally within their portal, when they may actually reside in remotely-running portals.

The hope is that these standards (along with others) allow content exchange among emissions-related portals.
Federated data system - DataFed

The air quality community is supported by a **non-intrusive, incremental data integration** infrastructure based on Internet standards (web services) and a set of web-tools evolving through the federated data system, **DataFed**. (Husar et al., 2004)

http://datafed.net
DataFed Information Flow

Integrated Data System for Air Quality-Analysis

Source: Husar et al., 2004
Current info systems are project/program oriented and provide end-to-end solutions.

Part of the data resources can be shared for re-use through DataFed.

Through DataFed, the data are homogenized into multi-dimensional cubes.

Data processing and rendering can then be performed through standard web services.

Each project/program can be augmented by federated data and services.
Geospatial Web Standards

Standards for finding, accessing, portraying, and processing geospatial data are defined by the Open Geospatial Consortium (OGC).

- **Web Map Server (WMS)** for exchanging map images, but the
- **Web Feature Service (WFS)** retrieves discrete feature data
- **Web Coverage Service (WCS)** allows access to multidimensional data that represent coverages, such as grids.
- **Sensor Observation Service (SOS)** multidimensional access to measurement data

While these standards are based on the geospatial domain, many are designed to be extended to support non-geographic data “dimensions,” such as time and the many other dimension tables found in emissions inventories.

DataFed-OGC Description: [http://www.datafed.net/DataLinks/OGC/OGC.htm](http://www.datafed.net/DataLinks/OGC/OGC.htm)
Web Map Service (WMS)

<table>
<thead>
<tr>
<th>GetMap Request</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://webapps.datafed.net/ogc_EPA.wsfl">http://webapps.datafed.net/ogc_EPA.wsfl</a></td>
</tr>
<tr>
<td>?SERVICE=wms</td>
</tr>
<tr>
<td>&amp;REQUEST=GetMap</td>
</tr>
<tr>
<td>&amp;VERSION=1.1.1</td>
</tr>
<tr>
<td>&amp;SRS=EPSG:4326</td>
</tr>
<tr>
<td>&amp;STYLES=</td>
</tr>
<tr>
<td>&amp;LAYERS=NEI_EGU.SO2</td>
</tr>
<tr>
<td>&amp;BBOX=-127.00,24.00,-66.00,50.00</td>
</tr>
<tr>
<td>&amp;TIME=1999-01-01T00:00:00</td>
</tr>
<tr>
<td>&amp;FORMAT=image/png</td>
</tr>
<tr>
<td>&amp;EXCEPTIONS=application/vnd.ogc.se_inimage</td>
</tr>
<tr>
<td>&amp;TRANSPARENT=TRUE</td>
</tr>
<tr>
<td>&amp;BGCOLOR=0xFFFFFF</td>
</tr>
<tr>
<td>&amp;WIDTH=700</td>
</tr>
<tr>
<td>&amp;HEIGHT=350</td>
</tr>
</tbody>
</table>
Web Coverage Service (WCS)

GetCoverage Request

http://webapps.datafed.net/ogc_EPA.wsfl
?SERVICE=wcs
&REQUEST=GetCoverage
&VERSION=1.0.0
&CRS=EPSG:4326
&COVERAGE=EPA_CAMD_HOUR.SO2_MASS
&FORMAT=NetCDF-table
&BBOX=-82.4606,42.9258,-82.4606,42.9258,0,0
&TIME=2002-04-01T15:00:00Z/2002-04-30T15:00:00Z
&WIDTH=700
&HEIGHT=350
&DEPTH=99
Multidimensional Air Quality Data Storage and Delivery through OGC Protocols

Relational Data Model

Dim Tables
- Sites
  - Site_ID
  - Mon_Site
  - Universal_Site_ID
  - Network_Code
  - Site_Code
  - Site_Index
  - Site_Collection_ID
  - USES_Site_ID
  - Site_Name
  - EON_Code
  - Longitude
  - Latitude
  - Elevation
  - Address
  - State_Province_Code
  - Site_Code
  - Site_Num

- Times
  - Site_Time
  - Loc_Time

- Monitoring/Estimates

Fact Table
- Air_Obs
  - Air_Obs_ID
  - Sample_ID
  - Parameter_ID
  - Parameter_Name
  - Site_ID
  - Method_ID
  - Method_Name
  - Analysis_Type
  - Date
  - Time
  - Value
  - Uncertainty

Dim Tables
- Parameters
  - Parameter_ID
  - Parameter_Code
  - Parameter_Name
  - Parameter_Property_Code
  - Parameter_Unit_Code
  - Unit_ID
  - Unit_Code
  - Size_Code
  - Parameter_Detector
  - Parameter_Measurement
  - Chem_CAS_Num

- Methods
  - Method_ID
  - Method_Name
  - Instrument_ID
  - Method_Desc
  - Sampling_Freq_Code
  - Sampling_Duration_Code
  - Sampling_Height

Polutant/Method

- Sulfur dioxide
- Carbon monoxide
- Nitrogen dioxide

Data View Services
- WMS
- WCS
- SOS

Facilities
- Monit./Estim.
- Poll-Meth

SensorML
Dynamic Emissions Browsing with Web Applications

A web application supplements a hard copy report by providing dynamic map and table creation allowing a user to focus on a spatial region or a particular plant.

Fire Locations and Smoke “Value Chain”

OGC Web Mapping Service (WMS) access to NOAA HMS fire location and smoke plume data make them directly accessible through the DataFed framework and available to NEISGEI applications.
Visualizing Emissions Data

Visualizing emissions data with GoogleEarth and World Wind:
1) involves very little programming
2) provides integrated access to GIS and satellite imagery
3) makes available more advanced visualization tools, such as World Wind’s temporal animation facility