Compilation and Design of a Functioning Distributed Database of North American Electric Generating Emissions

Stefan Falke
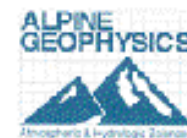
Center for Air Pollution Impact and Trend Analysis
Washington University in St. Louis

Gregory Stella

Alpine Geophysics, LLC

Terry Keating

US EPA – Office of Air & Radiation

Background

Air pollutant emission inventories for the US, Canada, and Mexico are compiled, stored and disseminated using different methods

The development of a single comprehensive and accurate emissions inventory is essential for the coordinated reporting, policy development, transport analyses, and socio-economic studies that create an environment for collaboration among international researchers, policy-makers, and the interested public

In support of this longer term goal, the Commission on Environmental Cooperation (CEC) and the US EPA have initiated a project to develop a prototype web tool for enabling uniform access to distributed emissions data from North American electricity generating power plants.

Objectives

• Recommend and demonstrate to the CEC approaches for the comparability of techniques and methodologies for data gathering and analysis, data management, and electronic data communications for promoting access to publicly available electric utility emissions

• Identify, collect, and review existing sources of electric generating utility (EGU) emissions and activity databases, and provide a summary of the state-of-science

• Build and demonstrate a relational database and web browser tool to query, retrieve, and explore emissions data from these distributed databases.

The project's focus is on criteria pollutants and toxics because of their availability and accessibility.

Available Online Emissions Data

Data Source	Time Coverage	Pollutants	Reporting Level
NEI (US)	1985-1999 (criteria) 1996-1999 (HAPs)	NOx, SO2, CO, PM, VOC, HAPs	Boiler
eGrid (US)	1996-2000	NOx, SO2, CO2, Mercury	Boiler & Generator
Clean Air Markets (US)	1980, 1985, 1988-1999	NOx, SO2, CO2	Generator
NPRI (Canada)	1994-2001	HAPs (Criteria starting in 2002)	Facility

These are publicly available, on-line accessible emissions data. Other data resources are available, but at this time only in hard copy form and therefore not usable in demonstrating distributed database concepts.

Database Fields Mapping

Emissions inventories are based on different underlying data models.

Gaining an in-depth understanding of these models and their formats and the similarities and differences among the inventories is a crucial step in developing an integrated inventory.

Each inventory uses a uniquely defined set of field names. However, many of these field names are similar to (or their content is similar to) fields in another country's inventory.

The relationships among the inventories will be captured by developing a "mapping" among fields in each inventory.

This mapping will provide a set of connections that can subsequently be applied to automated query and integration of data from multiple inventories.
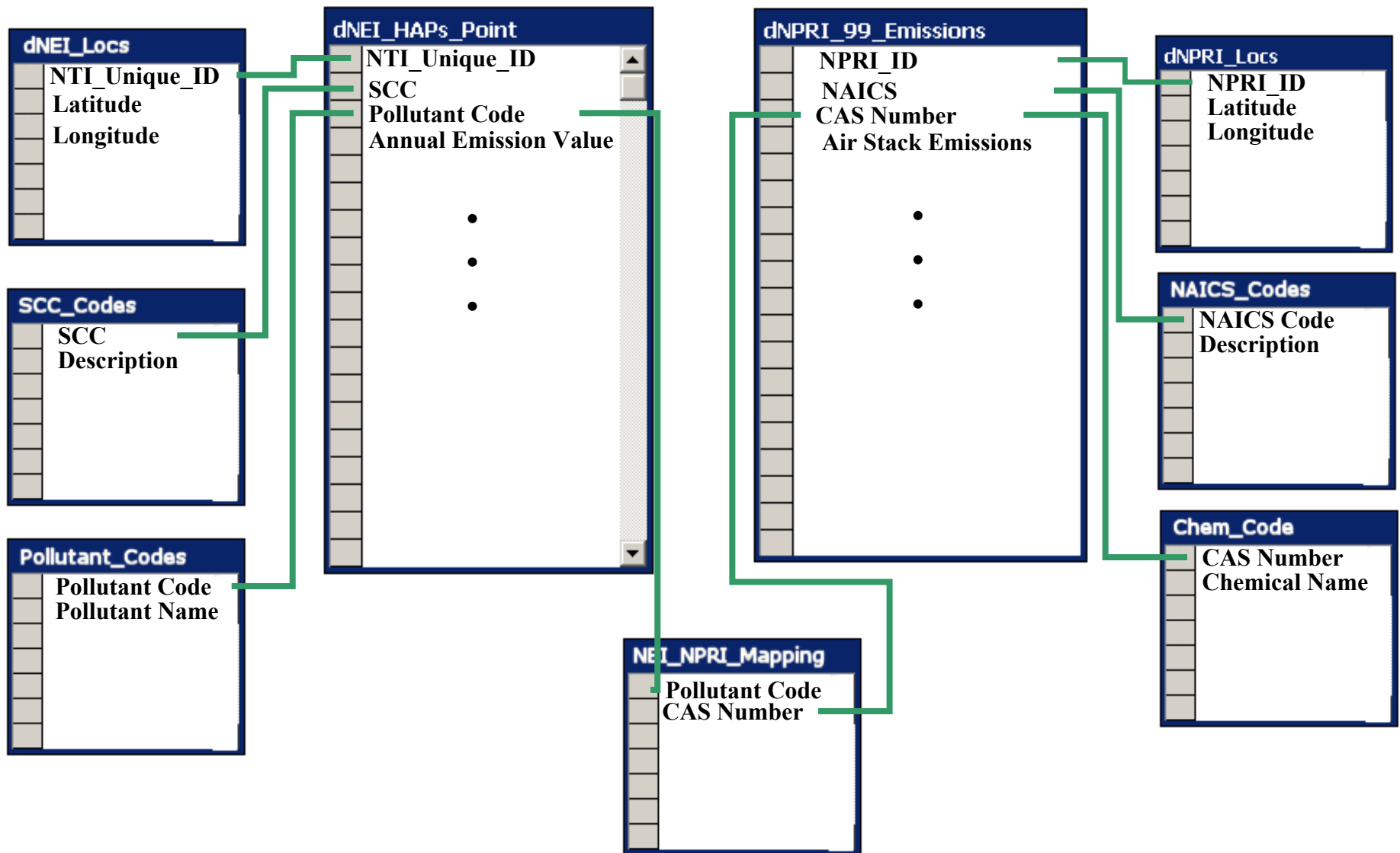
Database Fields Mapping
Total Facility Emissions Example

Data Element	NPRI (Canada)	NEI (US)
Reporting Year	Facility.ReportYear	tblPointTR.intInventoryYear
Industrial Classification	Facility.US_SIC	tblPointSI.strSICPrimary
Pollutant	Substanc.CAS_Number	tblPointEM.strPollutantCode
Emission Units	Substanc.Units	tblPointEM.strEmissionUnitNumerator
Emissions - Stack Release	Substanc.AirSta_V	tblPointEM.dblEmissionNumericValue + tblPointER.strEmissionReleasePointType ("02")
Emissions - Fugitive Release	Substanc.AirFug_V	tblPointEM.dblEmissionNumericValue + tblPointER.strEmissionReleasePointType ("01")

Data Key: Table.Field Name

Sometimes the mapping requires the combination of more than one fields in one data table to match a single field in another table.

Relational Database Structure Example

dNEI_Locs
- NTI_Unique_ID
- Latitude
- Longitude

dNEI_HAPs_Point
- NTI_Unique_ID
- SCC
- Pollutant Code
- Annual Emission Value

  ⋮

dNPRI_99_Emissions
- NPRI_ID
- NAICS
- CAS Number
- Air Stack Emissions

  ⋮

dNPRI_Locs
- NPRI_ID
- Latitude
- Longitude

SCC_Codes
- SCC
- Description

NAICS_Codes
- NAICS Code
- Description

Pollutant_Codes
- Pollutant Code
- Pollutant Name

Chem_Code
- CAS Number
- Chemical Name

NEI_NPRI_Mapping
- Pollutant Code
- CAS Number

NEI and NPRI databases are 'joined' using their related field mappings thereby allowing integrated database queries.

Integrated Emissions Database through Queries

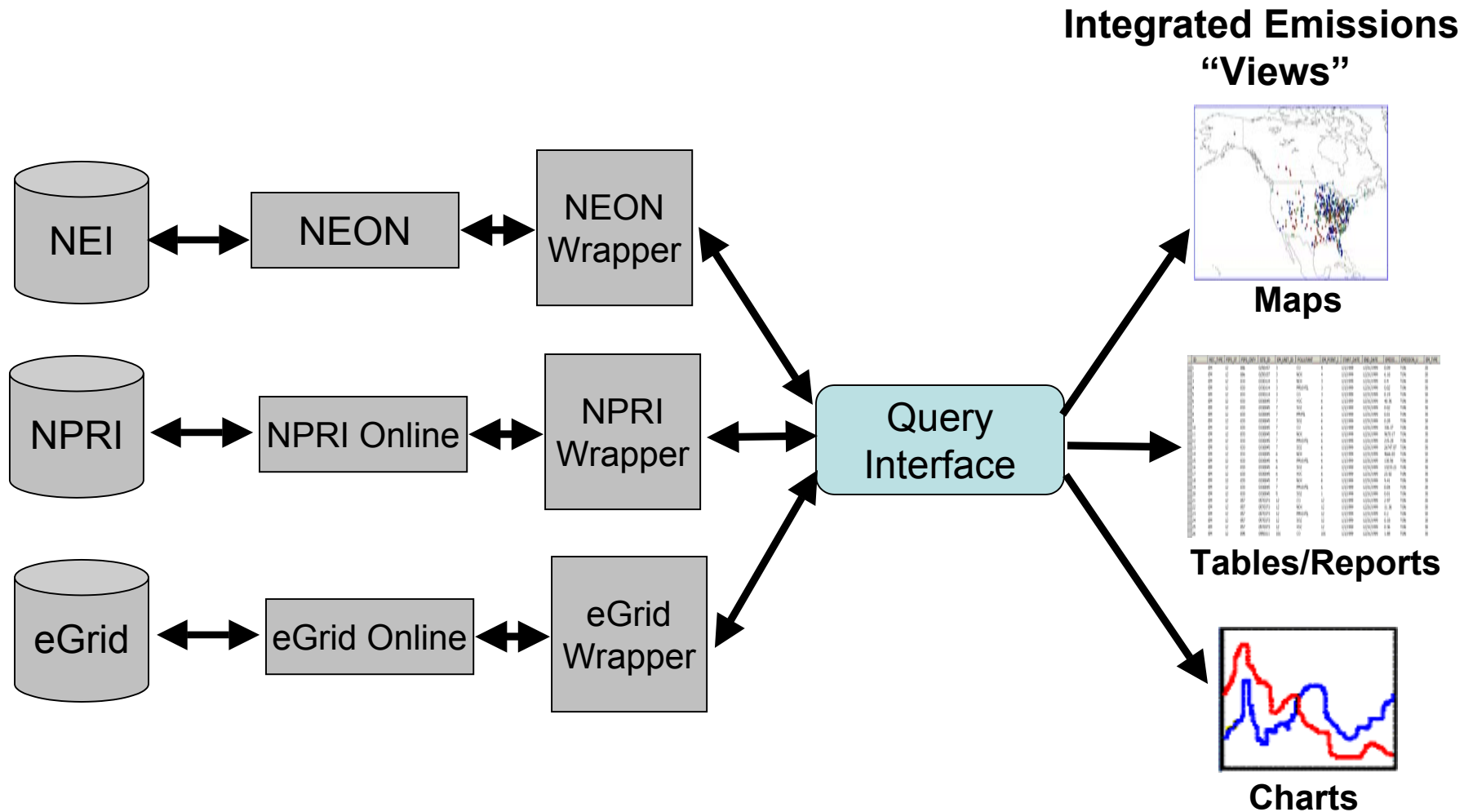1999 Annual Emissions of Hydrogen Fluoride
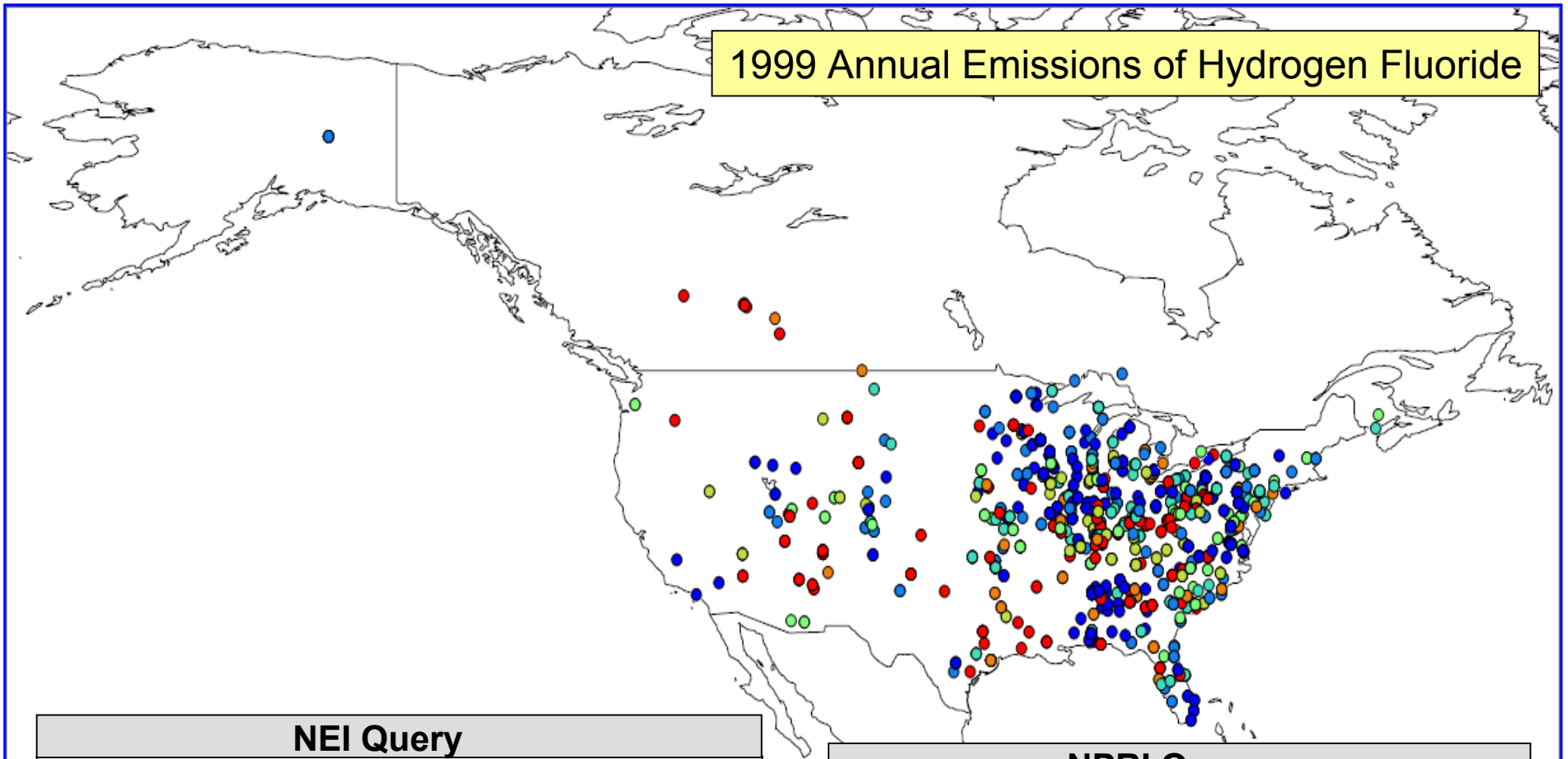
NEI Query

```
SELECT [NTI_Unique_ID] AS Loc_Code,
YCoordinate as Lat, XCoordinate as Lon,
[Emissions(TPY)] AS [VALUE]
FROM dNEI_HAPs_Point INNER JOIN dNEI_Locs
ON dNEI_HAPs_Point.NTI_Unique_ID =
dNPEI_Locs.NPRI_ID
WHERE (dNEI_HAPs_Point.[PollutantCode] =
'7664393') AND (dNEI_HAPs_Point.[SCC] LIKE
'101%')
```

NPRI Query

```
SELECT NPRI_ID AS Loc_Code, LATI_DEC AS
Lat, LONG_DEC AS Lon, AIRSTA_V AS [VALUE]
FROM dNPRI_99_Emissions INNER JOIN
dNPRI_Locs ON dNPRI_99_Emissions.NPRI_ID
= dNPRI_Locs.NPRI_ID
WHERE (dNPRI_99_Emissions.[CAS_NUMBER] =
'7664-39-3') AND
(dNPRI_99_Emissions.[NAICS_4] ='2211')
```

Distributed Emissions Network

Integrated Emissions "Views"

Maps

Tables/Reports

Charts

NEI

NEON

NEON Wrapper

NPRI

NPRI Online

NPRI Wrapper

eGrid

eGrid Online

eGrid Wrapper

Query Interface

Distributed databases are accessible through 'wrappers' that translate data into common formats which allow integrated analysis through the application of database field 'mappings'.

A network of distributed databases should not require changes in operation by the data provider once their data are internet accessible.

This project is further described on its website:

http://capita.wustl.edu/NAmEN

This purpose of this website is to not only document the project's progress but to also provide interactive testing and feedback from the broader community. You are invited to participate at any level.

You may also contact either Stefan Falke (stefan@me.wustl.edu) or Greg Stella (gms@alpinegeophysics.com) with questions or comments