An Introduction to the Air Quality Data Mart

2007 AQS Conference

Topics

- Background
- Contents of the Air Quality Data Mart
- Accessing the Air Quality Data Mart
- Air Quality Field Guide

Background-1
(What is a Data Warehouse?)

- It is a database
- Optimized for retrieval and analysis
- Enterprise Content – Includes data for all aspects of an organization
- Data Integration Platform – designed to allow integrated analysis of data from different sources.
- Typically Historical Data (static data)

Background-2
What is a Data Mart?

- Segment or subset of a Data Warehouse (e.g. Air Quality data only, instead of all EPA data)
- Targeted to a specific purpose or domain
- Consistent with organization and representations of higher-level warehouse (e.g. mapping AQS sites to “Facilities”)

Air Quality Data Mart
What is it?

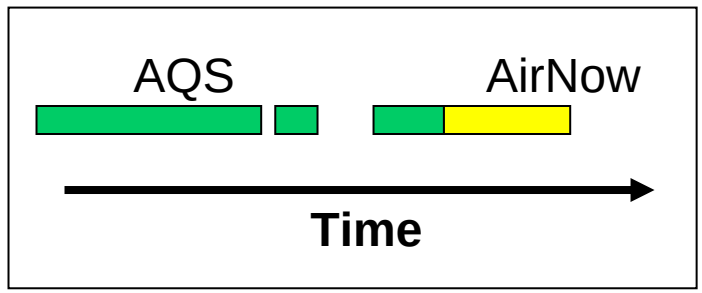
- A single place to get all of the Air Quality data EPA has
  - Still only contains data in AQS
    - Yes IMPROVE because loaded in to AQS
    - No AirNow and CASTNET yet
- Separate from AQS so that each system can operate without impacting the other
  - Updated every work night
- Target users:
  - Scientific and Research community
  - High volume data consumers
  - Application builders (e.g. Air quality modelers)
  - Exchange network participants

Air Quality Data Mart Contents

- Raw air quality measurement data
- Summary Data
- Site and Monitor Data

Contents: Raw Data

- AQS
  - Production: 1980 to Present
    - All parameters
    - All hours
    - Regardless of qualifiers, method, etc…
    - Includes NAAQS averages (with “Event” data included)
    - Available in local time or GMT
    - Note: no precision and accuracy data
- AirNow
  - Production: Opt-in program for the submitting Agencies
    - By Agency, by parameter, maybe by site
    - POC matching
    - Only after last AQS value
    - Purge if over 1 year old
  - Implement by end of 2007



Air Quality Summary Data

- **Daily Summaries**
  - **From AQS**
    - **Observation Count**
    - **Observation percent**
    - **Arithmetic Mean**
    - **Maximum and hour of Max**
    - **AQI**
    - **Values > Standards**
  - **Additional summaries with AirNow Values**
    - **AQI**
    - **Observation Count**
    - **Arithmetic Mean**
    - **Max value**
- **Annual Summaries**
  - **From AQS**
    - **Obs count and percent**
    - **Valid day**
    - **Required count**
    - **Null obs count**
    - **Exceptional values count**
    - **Min value**
    - **Arith Mean and Std Dev**
    - **Geo Mean and Std Dev**
    - **1st – 4th Max**
    - **Percentiles (99, 98, 95, 90, 75, 50)**
    - **Values > Standards**
    - **Half MDL obs count**
    - **Days > alert level**
    - **Est days > Standards**
    - **Missing days assumed < Standards**
  - **Additional AirNow values**
    - **Obs count (total, AQS, AirNow)**
    - **Arith Mean and Std Dev**
    - **1st – 4th Max**
    - **Percentiles (99, 98, 95, 90, 75, 50)**

Site and Monitor Descriptions

- Basically all the info from AQS
  - Geopolitical
    - AQS Site ID
    - Latitude, longitude, altitude…
    - Tribe, State, county, UAR, AQCR, Class I area…
    - Obstructions
  - Administrative
    - Agency roles
    - Operational dates
  - Monitoring protocol
    - Parameter
    - Duration
    - Sample Method
    - Analysis Method
    - Schedule
    - Frequency

How to access the Data Mart

- Start with http://www.epa.gov/ttn/airs/aqsdatamart/
- No single “standard” access method
- Concept: Access mechanism separate from data storage
- Toolkit provides for building interfaces (Web Services)
- Two “example” interfaces provided
  - Stand-alone “Direct” interface
  - CDX Node Client interface
- Other user interfaces planned and coming on-line (example Air Compare)

Access Toolkit

- Web Services Application Programming Interface (API)
- Interface Facilities
  - Authenticate: Identify who you are based on user-id and password
  - GetList: Provides information about retrievable data
    - What data elements can be retrieved
    - What are the valid values for query parameters (e.g. states, counties, substances, and etc.)
  - GetMonitorData: Provides definitions for sites & monitors
  - GetRawData: Provides measurement data in AQS XML Submission schema
  - GetValues: Provides measurement or summary data in “GeoCoded” format (i.e. by latitude and longitude)
  - GetSitesByThreshold: Provides list of sites with measurement or summary value larger than user provided threshold.
  - GetStatus: Obtain the status of a previously submitted query.
  - Download: Retrieve the results of a previously submitted query.

Access Toolkit 2

- Strategy: General purpose interface
  - Building blocks for specialized interface
  - Simple queries that can be composed to perform complex retrievals
  - API for toolkit can be machine generated by industry standard tools.

DEMO

See Data Mart Interface Quick Start Guide

XML Notes

- Good news: Outputs are XML; Bad News: Outputs are XML.
- What is XML:
  - XML documents are structured, self-describing files
  - Intended to be both machine processable and human readable
  - Use “tags” to label each piece of content (like HTML)
  - Use a “Schema” to define the meaning and structure of content.

XML Style Sheets

- XML “style sheets” associate a display format with an XML document (completely analogous to HTML style sheets)
- ALL Air Quality Data Mart output documents have style sheets.
- User-specific style sheets can be developed (with XSLT)
- Style sheets can be used to generate HTML, PDF, and structured text (e.g. CSV files)

Documentation

- Air Quality Data Mart documentation can be found at http://www.epa.gov/ttn/airs/aqsdatamart/
- Of particular interest is the “Field Guide”
  - This document provides introductory descriptions of AQS data and objects and the relationships between them.
- Demo

Links

- OAR Air Pollution Data Sources web page: http://www.epa.gov/air/airpolldata.html
- AQS: http://www.epa.gov/ttn/airs/airsaqs/index.htm
- AirNow: http://airnow.gov/
- AirNow Tech: http://www.airnowtech.org/
- AirData: http://www.epa.gov/air/data/
- AirExplorer: http://www.epa.gov/airexplorer/
- AirCompare: Will link from AirNow when deployed
- AQS Data Page: http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm
- NATA: http://www.epa.gov/ttn/atw/natamain/
- Air Trends: http://www.epa.gov/air/airtrends/
- AQS Data Mart: http://www.epa.gov/ttn/airs/aqsdatamart/index.htm
- CASTNET: http://www.epa.gov/castnet/