Plans for Enhanced AQS Data Accessibility
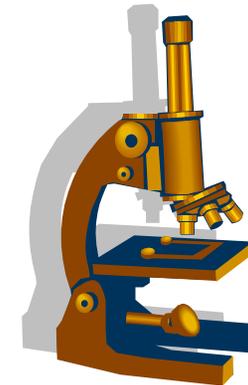
SAMWG Meeting

May 14, 2004

Nick Mangus

US EPA / OAQPS

The Vision for Enhanced Access to AQS Data

- Allow access to data
  - Timely
  - Complete
  - Universal (available to anyone)
- Features
  - Integrated with related data
    - Ambient data
    - Metadata
  - Fine and medium scale views

What Does This Mean

- **Access**
  - Extracting data
    - For use with own analytical tools
  - Looking at it with a Business Intelligence tool
- **Timely**
  - Depends on your needs
    - Submitters: need to know <u>immediately</u> if data that was submitted made it into the system
    - Rest of us: <u>daily</u> or <u>quarterly</u> depending on analytical needs
    - The data can only go out as fast as it comes in (see "complete…")

What Does This Mean (2)

- ## Complete
  - All of the raw and summary ambient data that is in AQS

- ## Available to anyone
  - Currently only registered AQS users have unfettered access to all data
  - We have web pages with summaries and extracts
    - Can be dated

Another Way to Say It

- **Present**
  - Registered users can get data out of AQS (~630) via reports
  - Anyone can query AirData web pages for annual summary data for criteria pollutants
  - Anyone can download data extracts from the TTN web pages

- **Future**
  - Anyone can run certain AQS reports
  - Anyone can extract the data of their choosing
  - We will provide limited visualization / analytical tools

How Do We Get There?

- Via a series of initiatives
  - The AQS Data Mart
  - AirQuest: The OAQPS Data Warehouse
  - AQS Web Services
  - Easier to access web pages (interim)
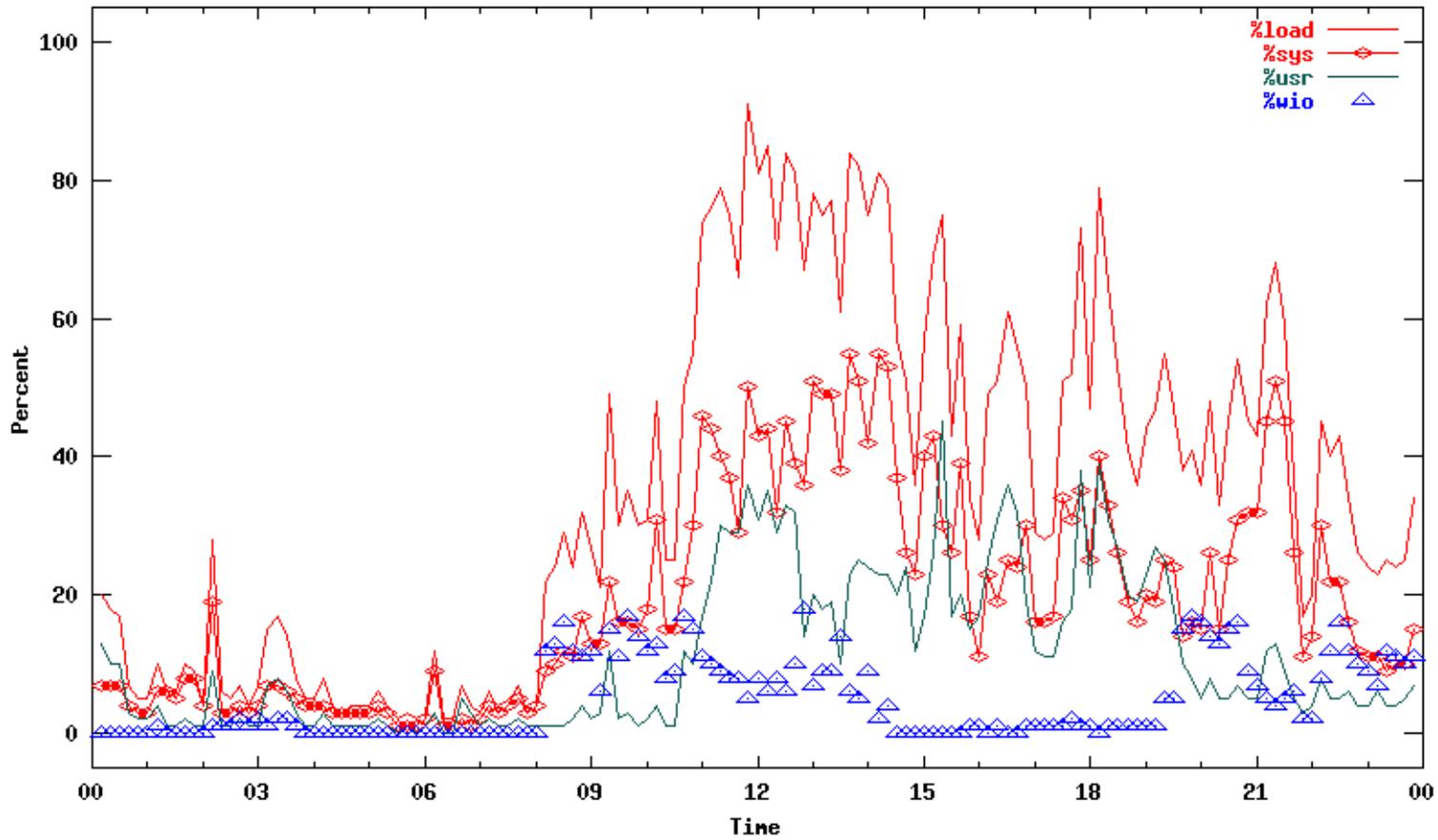- These are being developed in coordination with the EPA OEI

AQS Data Mart

- **What (this is <u>very draft</u>):**
  - A "copy" ("mirror"? "replica"?) of AQS raw and summary data
  - Frequently updated
    - We're looking at every night
    - 1980 – Present
  - Transformed into a structure that makes queries faster (dimensional data model)
  - Allows for data extractions -> files
  - Allows for use of exploratory (BI) tools
  - Available via the TBD EPA portal web page

AQS Data Mart (2)

- **Why:**
  - To make the "live" data readily available without increasing load on the production system
    - AQS is a regulatory system optimized for input
  - To be able to put AQS and AirNow data in same place
  - To link to common Agency definitions of Substance, etc.
  - Building block for EPA framework for multimedia data
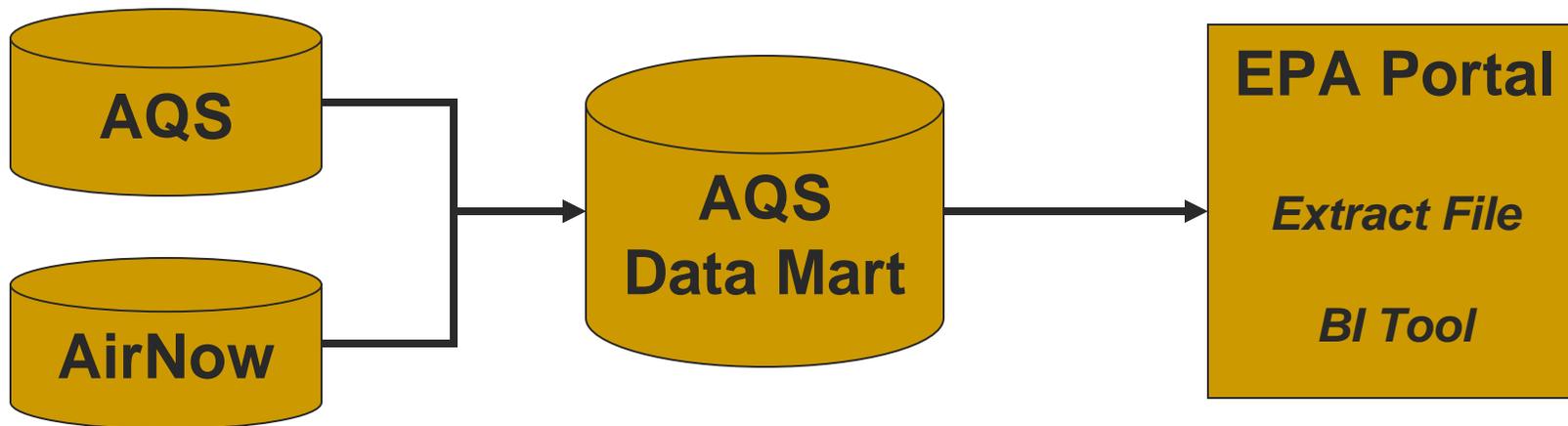
Typical (Random) Canyon Load

CPU usage for canyon.rtpnc.epa.gov on 10 May 2004

AQS Data Mart (3)

- **Intended users**
  - People who know the data
  - People who want to download a large chunk of data to analyze
    - States, RPOs, Etc.
    - Academics in health and exposure fields
  - People with no other access who need to drill into detailed (raw) data
  - Access will be via the EPA Portal
    - Registered users only (debate over what this means)
    - Priority will be given to partners

AQS Data Mart Concept

AQS

AirNow

AQS Data Mart

EPA Portal

Extract File

BI Tool

BI Functions

- **Query**
  - Tables
  - Charts
  - Plots
  - Filtering and Drilling

AQS Data Mart Status

- Currently in the design / procurement phase
  - Getting help from EMAD on making sure it's useful and useable
- Development is on the "fast track"
- Will start development this summer
- Prototype deployment target: End of FY
- Initial production deployment: End of CY
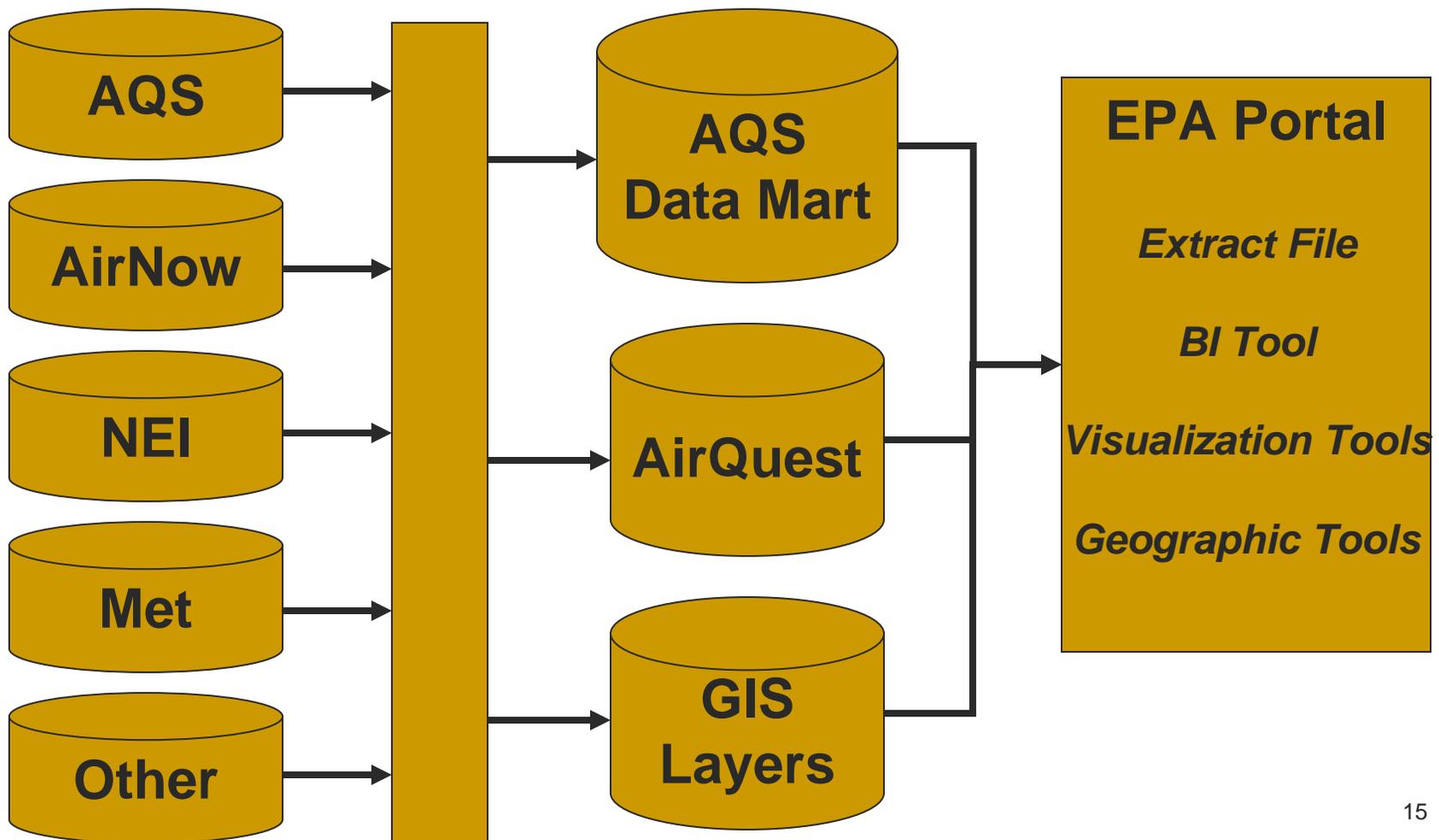- Full production deployment: Spring 2005

AirQuest

- What:
  - Another "data warehouse" with the following data
    - AQS Criteria Summaries
    - NEI Criteria
    - Acid Rain CEM data
    - Acid Rain fuel and control data
    - Mercury stack tests
    - Coal pile / delivery samples
    - Emission factors
    - Meteorology
    - Demographics
    - More Coming soon: NEI Toxics, NATA results, PM2.5 Speciation from AQS, AirNow, gridded model results
  - Updated quarterly

AirQuest (2)

- **Why:**
  - To integrate data so that queries <u>across</u> information are easy
    - Reduce the time spent collecting and collating data
    - Allows for discovery-type queries
- **Intended users:**
  - Decision makers
    - There are several tools ready to connect to the front including business intelligence and analytics (BIA) and geographic information systems (GIS)
  - Analysts
    - Other analytical tools can be connected
    - Direct SQL connections also available (maybe only inside the EPA firewall)

AQS Data Mart + AirQuest Concept

15

AirQuest Status

- Initial prototype deployed
  - Currently only internal OAQPS access
- Training for users / testing: June-July
- Make fixes based on user feedback: summer
- Migrate to central environment: summer
- Roll-out to partners soon thereafter
- Access via EPA Portal and 2-3 applications done or in development
  - Business Objects, Air Explorer, GIS Queries
- Enhancements: continuous

AQS Web Services

- What:
  - The ability to use the web services framework to run certain AQS reports
  - Web services analogy: extending HTML services one level down – to the application
  - Service descriptions are published so everyone knows what they provide and how to use
  - Request: an XML document containing the report and the selection criteria (e.g., the query)
  - Response: an XML document containing the data from AQS

Web Services (2)

- **Why:**
  - To prove the concept of the technology working with AQS and OEI
  - To allow anyone to request the most popular AQS reports (that are low in resource consumption)
    - Quick Look
    - Data Completeness
    - Air Quality Index Summary
    - Maximum Values

Web Services Status

- EPA infrastructure is in place
- AQS work just beginning
  - Have to develop some new XML schemas
  - Have to develop the web service descriptions
  - The services will be the existing programs that run the reports
- Should not be a huge task
- Available late this summer
- Via the EPA node on the Exchange Network

In the Interim

- Web Pages
  - Air Data – Annual Summary
    - TTN – Related Links
  - AQS Query – more detailed + ref.
    - Refresh coming next month
    - TTN – Related Links
  - Detailed Data Download Page
    - TTN – Requesting AQS Data
  - Data Request Form
    - TTN – Requesting AQS Data

U.S. Environmental Protection Agency

Technology Transfer Network
Air Quality System

Recent Additions | Contact Us | Print Version   Search: [____] **GO**

EPA Home > TTN > AQS > Requesting AQS Data > Download Detailed AQS Data

Download Detailed AQS Data

The table below provides links to files of detail data (a.k.a., raw data) extracted from the AQS database in input transaction format (RD transactions). These files following the naming convention of "RD_501_ppppp_yyyy.ZIP", where "ppppp" = AQS parameter code and "yyyy" = year of data. All files are zipped to conserve disk space. Refer to our "Tools" page if you need assistance with file types.

A file containing 2000-2003 site and monitor records as of 3/5/04 in input transaction format 📄 is also available as a text file. Site records begin with the letter "A"; monitor records begin with "M".

Menu (left sidebar):
- System Overview
- AQS Web Application
- Manuals and Guides
- Precision and Accuracy Data
- Frequently Asked Questions
- Memos
- User Registration
- Training
- AQS Conference
- **Requesting AQS Data**
- Technical Forum
- Contacts
- **Related Links**

Raw Data retrieved from AQS

Year of data, zipped file size, (unzipped file sizes), Date of extraction											
	Updates for these years may come at any time				Updates for these years are no longer expected						
Pollutant (parameter code) File name (substitute year for "yyyy")	2003 most are incomplete until at least 7/1/04	2002	2001	2000	1999	1998	1997	1996	1995	1994
Carbon Monoxide (42101) RD_501_42401_yyyy.ZIP	42101_2003 13711KB (241645KB) 3/20/04	42101_2002 14585KB (261659KB) 3/16/04	42101_2001 14581KB (269358KB) 3/19/04	42101_2000 15383KB (283253KB) 3/19/04	42101_1999 9812KB (173441KB) 3/17/04	42101_1998 10031KB (176125KB) 3/17/04	42101_1997 10136KB (177404KB) 3/17/04	42101_1996 10102KB (176122KB) 3/17/04	42101_1995 10171KB (175888KB) 3/17/04	42101_1994 10774KB (185853KB) 3/17/04
Lead - PB - daily (12128) RD_501_12128_yyyy.ZIP	12128_2003 47KB (637KB) 3/13/04	12128_2002 65KB (872KB) 3/13/04	12128_2001 70KB (926KB) 3/13/04	12128_2000 66KB (919KB) 3/13/04	12128_1999 72KB (996KB) 3/13/04	12128_1998 85KB (1190KB) 3/13/04	12128_1997 104KB (1484KB) 3/13/04	12128_1996 120KB (1700KB) 3/13/04	12128_1995 126KB (1774KB) 3/13/04	12128_1994 136KB (1895KB) 3/13/04
Nitrogen Dioxide	42602_2003 13706KB	42602_2002 13051KB	42602_2001 13785KB	42602_2000 13712KB	42602_1999 13374KB	42602_1998 13838KB	42602_1997 13706KB	42602_1996 12473KB	42602_1995 12540KB	42602_1994 12767KB

Summary

- **AQS Data Mart**
  - Narrow, but deep
  - Available late this year or early next year
- **AirQuest**
  - Wide, but not so deep
  - Available late this year
- **Web Pages**
  - In the interim
  - Already in place – will better organize
- **Web Services**
  - Pilot a new way to communicate with AQS
  - Available late this year

Longer Term

- There are obvious synergies to the projects
- The two data warehouses should become transparent to the user
  - A data request is filled, source is irrelevant to the user
- Generic tools hit all data
  - Business Intelligence
  - Web services
- Custom tools enhanced?
  - Specific analytical tools added as applications or web services

Hearing From You

- If you have specific data needs or functionality requests, please <u>email</u> me
  - mangus.nick@epa.gov
- We're in a lifecycle
  - You can influence the current one
  - You can influence future ones
- Suggestions on descriptive data that would make the data more useable
  - "Field Guide to AQS Data"