

**DATABASE AND ANALYTICAL TOOL DEVELOPMENT
FOR THE MANAGEMENT OF DATA
DERIVED FROM US DOE (NETL) FUNDED
FINE PARTICULATE (PM_{2.5}) RESEARCH**

**SEMI-ANNUAL
TECHNICAL PROGRESS REPORT**



Submitted to:

U. S. Department of Energy
National Energy Technology Laboratory
Pittsburgh, PA 15236

DOE AWARD NUMBER: DE-FC26-02NT41476

DATABASE AND ANALYTICAL TOOL DEVELOPMENT FOR THE MANAGEMENT OF DATA DERIVED FROM US DOE (NETL) FUNDED FINE PARTICULATE (PM_{2.5}) RESEARCH

Semi-Annual Technical Progress Report

REPORT PERIOD START DATE: February 12, 2004

REPORT PERIOD END DATE: August 11, 2004

PRINCIPAL AUTHORS: Robinson P. Khosah, Ph.D.
Advanced Technology Systems, Inc.
Charles G. Crawford
Advanced Technology Systems, Inc.

DATE REPORT ISSUED: September 2004

DOE AWARD NUMBER: DE-FC26-02NT41476

CONTRACTOR: **Advanced Technology Systems, Inc.**
639 Alpha Drive – RIDC Park
Pittsburgh, PA 15238-2819

SUBCONTRACTORS: **Ohio University**
E317 Grover
Ohio University
Athens, OH 45701
Texas A&M University – Kingsville
MSC213
Texas A&M University - Kingsville
Kingsville, TX 78363

ATS PROJECT NO: 01-050-S&T-P

ABSTRACT

Advanced Technology Systems, Inc. (*ATS*) was contracted by the U. S. Department of Energy's National Energy Technology Laboratory (DOE-NETL) to develop a state-of-the-art, scalable and robust web-accessible database application to manage the extensive data sets resulting from the DOE-NETL-sponsored ambient air monitoring programs in the upper Ohio River valley region. The data management system was designed to include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions. The project is being conducted in two phases. Phase One includes the following tasks: (1) data inventory/benchmarking, including the establishment of an external stakeholder group; (2) development of a data management system; (3) population of the database; (4) development of a web-based data retrieval system, and (5) establishment of an internal quality assurance/quality control system on data management. Phase Two, which is currently underway, involves the development of a platform for on-line data analysis. Phase Two includes the following tasks: (1) development of a sponsor and stakeholder/user website with extensive online analytical tools; (2) development of a public website; (3) incorporation of an extensive online help system into each website; and (4) incorporation of a graphical representation (mapping) system into each website. The project is now into its twenty-fourth month of development activities.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

TABLE OF CONTENTS

ABSTRACT	i
LIST OF GRAPHIC MATERIALS	iii
EXECUTIVE SUMMARY	iii
I. INTRODUCTION	1
A. Project Goals and Objectives	1
B. Project Phase Development	2
II. EXPERIMENTAL	3
A. Phase One Tasks	3
Task 1.1 – Data Benchmarking/Inventory for Database Applications	3
Task 1.2 - Develop Data Management System Architecture	3
Task 1.3 - Population of Database	3
Task 1.4 - Develop Web-Based Retrieval System	5
Task 1.5 - Develop QA/QC System	9
B. Phase Two Tasks	9
Task 2.1 - Develop Stakeholder Website	9
Task 2.2 - Develop Public Website	10
Task 2.3 - Develop Online Help Feature	12
Task 2.4 - Provide Graphical Representation Data	12
Task 2.5 - Performance Testing	12
III. RESULTS AND DISCUSSION	18
IV. CONCLUSION	19
V. REFERENCES	21
VI. BIBLIOGRAPHY	22
VII. LIST OF TERMS, ACRONYMS AND ABBREVIATIONS	23

LIST OF GRAPHICAL MATERIALS

Figure 1 - Data Import Operation.....	4
Figure 2 - Static Download Controls	5
Figure 3 - Query Process Map	6
Figure 4 - Site Query Selection Controls	7
Figure 5 - Site GIS Map Selection Controls.....	7
Figure 6 - Selected Site Information	8
Figure 7 - Data Type Selection Controls.....	8
Figure 8 - Screenshot of Stakeholder Website	10
Figure 9 - Screenshot of Public Website	11
Figure 10 - Advanced Calculator	13
Figure 11 - Graph of Calculator Generated Data.....	14
Figure 12 - Site Comparison Tool.....	15
Figure 13 - Site Comparison Tool Options	15
Figure 14 - Site Comparison Tool (Box and Whisker Plot)	16
Figure 15 - Box and Whisker Plot.....	16
Figure 16 - Time Series Plot.....	17
Figure 17 - Frequency Histogram.....	18

EXECUTIVE SUMMARY

Advanced Technology Systems, Inc. (*ATS*) was contracted by the U. S. Department of Energy's National Energy Technology Laboratory (DOE-NETL) in August, 2002, to develop a state-of-the-art, scalable and robust web-accessible database application to manage the extensive data sets resulting from ambient air monitoring programs in the upper Ohio River valley region that have been sponsored by DOE-NETL.

Research projects sponsored by DOE-NETL collected large amounts of data on PM_{2.5} and other air pollutants at ambient monitoring sites in the upper Ohio River valley region between 1999 and 2003. Extensive monitoring sites have been operated by DOE-NETL and its contractors in Pittsburgh, PA (two (2) urban sites), Holbrook, PA (rural site), South Park, PA (suburban site), and Steubenville, OH. Less-extensive monitoring sites have been operated in six other locations in PA, OH and WV. The main objectives of the current effort are to gather the data from all these monitoring sites into a common database, and to develop analytical tools that will make the data easily accessible to researchers and the public via the Internet.

In addition to the data collected by DOE-NETL and its contractors, the database will include, to the greatest extent possible, ambient air data collected by other agencies in the upper Ohio River valley region, such as the U.S. EPA, Pennsylvania Department of Environmental Protection (PA-DEP), West Virginia Division of Environmental Protection (WV-DEP), Ohio EPA, and the Allegheny County Health Department (ACHD). Although emphasis is being placed on the upper Ohio River valley region, the database may also include data collected at other DOE-NETL sponsored sites outside the region, such as sites operated by the Tennessee Valley Authority in the Great Smokey Mountains and by the Southern Research Institute in North Birmingham, AL. The database and analytical tool development effort is also being coordinated, to the extent possible, with a similar effort by U.S. EPA to develop a relational database for data collected at its "PM Supersites". This coordination will ensure that the database and analytical tools produced under the DOE-NETL effort will be readily accessible to a wide variety of stakeholders.

The data management system will include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

The project is being conducted in two phases. The entire project has been divided into ten primary tasks and those have been segmented into two primary phases. The project is now into its twenty-fourth month of development tasks and Phase Two began in August 2003. Phase One consisted of design and specification tasks related to designing, implementing and populating the primary database that will house the collected data. Phase Two consists of tasks involving the design, implementation and testing of both website interfaces along with any analytical tools and features integrated into the project's websites.

I. INTRODUCTION

Advanced Technology Systems, Inc. (*ATS*) was contracted by the U. S. Department of Energy's National Energy Technology Laboratory (DOE-NETL) in August, 2002, to develop a state-of-the-art, scalable and robust web-accessible database application to manage the extensive data sets resulting from ambient air monitoring programs in the upper Ohio River valley region that have been sponsored by DOE-NETL.

Research projects sponsored by DOE-NETL collected large amounts of data on PM_{2.5} and other air pollutants at ambient monitoring sites in the upper Ohio River valley region between 1999 and 2003. Extensive monitoring sites have been operated by DOE-NETL and its contractors in Pittsburgh, PA (two (2) urban sites), Holbrook, PA (rural site), South Park, PA (suburban site), and Steubenville, OH. Less-extensive monitoring sites have been operated in six other locations in PA, OH and WV. The main objectives of the current effort are to gather the data from all these monitoring sites into a common database, and to develop analytical tools that will make the data easily accessible to researchers and the public via the Internet.

In addition to the data collected by DOE-NETL and its contractors, the database will include, to the greatest extent possible, ambient air data collected by other agencies in the upper Ohio River valley region, such as the U.S. EPA, Pennsylvania Department of Environmental Protection (PA-DEP), West Virginia Division of Environmental Protection (WV-DEP), Ohio EPA, and the Allegheny County Health Department (ACHD). Although emphasis is being placed on the upper Ohio River valley region, the database may also include data collected at other DOE-NETL sponsored sites outside the region, such as sites operated by the Tennessee Valley Authority in the Great Smokey Mountains and by the Southern Research Institute in North Birmingham, AL. The database and analytical tool development effort is also being coordinated, to the extent possible, with a similar effort by U.S. EPA to develop a relational database for data collected at its "PM Supersites". This coordination will ensure that the database and analytical tools produced under the DOE-NETL effort will be readily accessible to a wide variety of stakeholders.

The data management system will include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

The project is being conducted in two phases. The project is now into its twenty-fourth month of development activities.

A. Project Goals and Objectives

The main objective of this database development effort is to gather the data, acquired from all these monitoring sites, into a common database, and to develop analytical tools that will make the data easily accessible to researchers and the public via the Internet.

The proposed data management system will include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

B. Project Phase Development

The project is being conducted in two phases. The first phase includes data inventory, benchmarking and database population tasks, as well as the development of data management architecture, a web-based retrieval system and an internal QA/QC system. A progress summary for Phase One is shown in the following table:

Task #	Description	Planned completed %	Actual completed %
1.1	Data Inventory/Benchmarking for Database Applications	100%	100%
1.2	Develop Data Management System Architecture	100%	98%
1.3	Population of Database	100%	90%
1.4	Develop Web-based Retrieval System	100%	95%
1.5	Develop Internal QA/QC System	N/A	N/A

N/A – on-going activity.

The second project development phase is currently underway. The tasks involved in this phase include the development of a stakeholder-specific website, a publicly accessible website and an online help feature. This phase also includes the development of special analysis tools to provide a graphical representation of the data and, of course, a series of performance tests designed to provide the best possible data management solution.

A progress summary for Phase Two is shown in the following table:

Task #	Description	Planned completed %	Actual completed %
2.1	Develop Stakeholder Website	100%	65%
2.2	Develop Public Website	100%	35%
2.3	Develop Online Help Feature	100%	5%
2.4	Provide Graphical Representation of Data	100%	35%
2.5	Performance Test	100%	5%

II. EXPERIMENTAL

A. Phase One Tasks

Task 1.1 – Data Benchmarking/Inventory for Database Applications

Any database application development effort requires some knowledge of the types and number of data contained in the resulting database. In addition to knowing this information, it is also wise to investigate or benchmark existing applications and development efforts that are similar in design or nature as the application being developed. Therefore, *ATS* proposed to conduct benchmarking investigations of existing projects, activities and applications prior to embarking on this project, as well as evaluate and quantify the data destined for usage with this application.

Several items were described in detail within the first, second and third Semi-Annual Technical Reports for this project. Those included the CARB Data Management Project in California and the EPA Supersite Database Development Project.

Task 1.2 - Develop Data Management System Architecture

The first semi-annual technical report contains detailed information regarding the design of the data management system architecture. To summarize, a system has been developed using MS SQL Server 2000 Enterprise Edition, MS Windows 2000 Advanced Server and external hardware, to provide the data management system architecture for this project.

A series of database objects and scripts have been constructed to ease all software development tasks and to accommodate expansion of the system to accommodate more users and data. The second semi-annual technical report contains detailed descriptions of *stored procedures*, or static queries that are stored within the database structure.

Additional changes to the Data Management System Architecture have been made to accommodate additional features such as our Geographic Information Systems (GIS) site selection tools and to improve the overall performance of the data structure.

Task 1.3 - Population of Database

A series of conversion scripts and data processing utilities have been created to reformat the supplied data files into the application-specific format, but to accommodate a large amount of data received from Desert Research Institute (DRI), a special application has been developed to *map* data from these files directly into the database. This special application, named *PM Data Imports Utility*, combines dynamic processing capabilities with static data to automate the data population process.

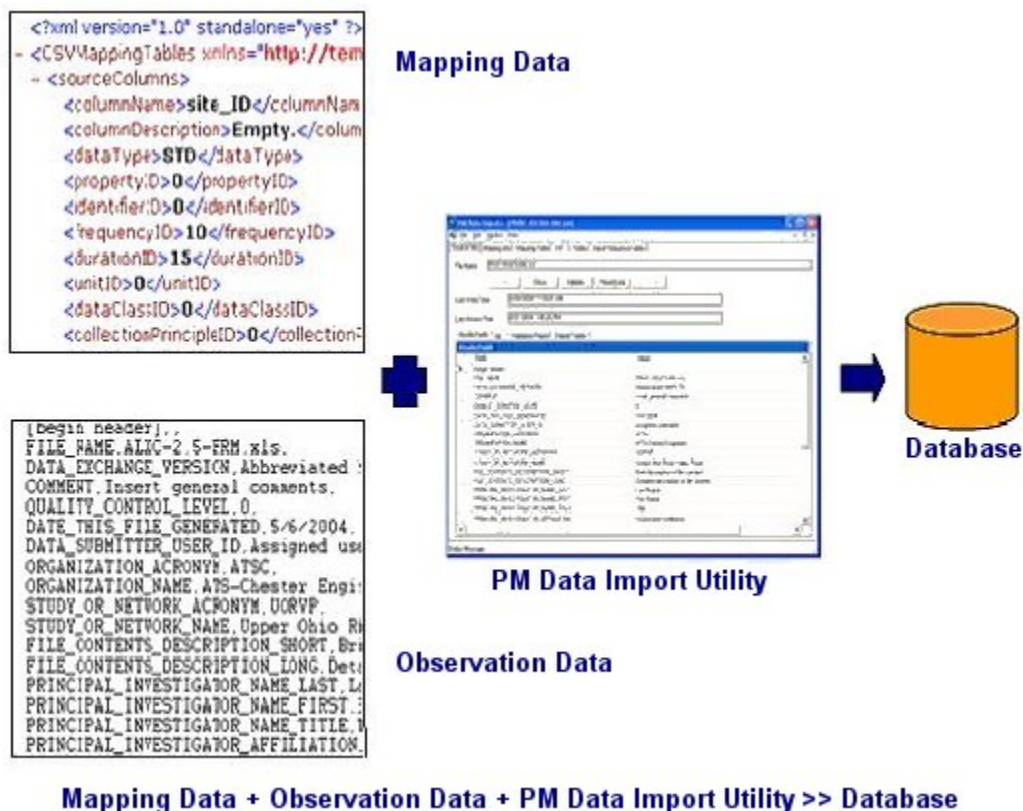


Figure 1-Data Import Operation

The project team *pre-processed*, or separated the data files by species and filter IDs (most DRI-processed data utilized a dual-filter collection process) and then utilized this application to import the data into the database. This process involved the mapping of DRI parameters to NARSTO parameters and sometimes involved creating new parameter records for items such as ‘Strong Acidity’, which equates an acidity level equivalent to levels of H₂SO₄, or Sulfuric Acid.

Other datasets, such as the continuous TEOM data and data obtained from the Environmental Protection Agency (EPA) for data collection sites in Pennsylvania, Ohio, West Virginia and Kentucky, were much easier to import into the system. For these datasets, the project team utilized Data Transformation Services (DTS) scripts to automate the population of the database and accommodate the inclusion of additional datasets in the same format.

There are three primary datasets not present in the database at this time. They include data collected by the NETL Supersite, Pittsburgh AQS, and SCAMP sites. The project team has not yet received the data from NETL and SCAMP and the Pittsburgh AQS data is still undergoing QA/QC checks after data preprocessing operations.

Task 1.4 - Develop Web-Based Retrieval System

Once the data is transferred to the database, users specify which data is to be retrieved through the Query Builder Interface. This interface provides resulting datasets in a tabular format and saves the query parameters for retrieval by the analysis tools. Static datasets are also provided via HTTP protocol and users complete a criteria selection process to download the original data files.

STATIC DATA DOWNLOADS

This process is similar to, but more limited than, the Query Builder Interface. Users select the type of data desired from each collection site's parameter list controls, and then click on the "Download Files" button above the selection list controls, to retrieve their data files.

Site	Select Parameters (Use Ctrl-click for selecting multiple)
Holbrook	Summary Mass Data SFS Mass Data SFS Speciation(Detailed) TEOM 2.5 Data
Lawrenceville	Summary Mass Data SFS Mass Data SFS Speciation(Detailed) TEOM 2.5 Data
Schenley Park	Surface TEOM GAS Continuous (SO4,NO3) Meteorological Denuded Organic Data
Morgantown	Summary Mass Data SFS Mass Data

Figure 2-Static Download Controls

DYNAMIC (AD HOC) QUERY SYSTEMS

Overview

The primary purpose of this project is to provide dynamic capabilities to this data retrieval system. Researchers need access to an ad hoc query system to build their own datasets, or to merge them with other datasets from multiple sources. The design of the query interface is the most critical aspect to consider when designing a web-based data retrieval system. It should provide the user an efficient means of deriving the output required from the database, without needing to understand the inner structure of the database or requiring the technical knowledge for writing Structured Query Language (SQL) queries and stored procedures.

Due to the complexity of the data model and the variety of available data types and parameters, a streamlined query process map was needed to develop an efficient query system interface. Figure 3 shows this query process map and details the data categories used to filter the data choices based on previously selected criteria.

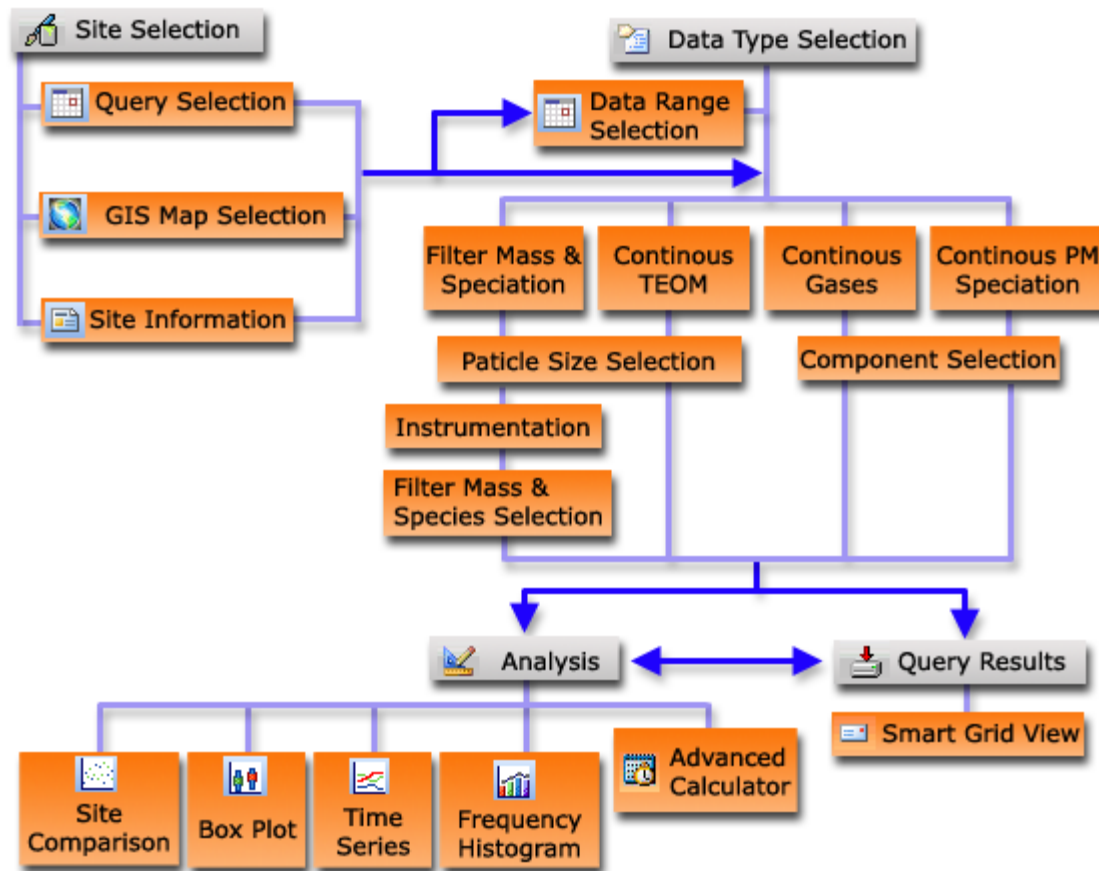


Figure 3-Query Process Map

The NETL and EPA air quality data comprised different types of pollutants, sampling periods, intervals and durations, instrumentations, and sampling methods. This kind of data composition might cause complexity, longer processing time, and frequent errors. Thus, the data query system was established based on hierarchical stages categorizing similar parameters and properties and selection comparability while users go back and forth for multiple selections.

Site Selection

As shown Figure 4, there are four site query selection controls/categories which enable users to quickly filter the available site list to a desired geographic or network focus area. Since all sites have different types of pollutants, instruments and sampling methods, site selection should be a starting point to explore the data through the query builder interface. This enables users to begin the query process with the desired site(s) and then continue with the selection of available data types for the selected sites.

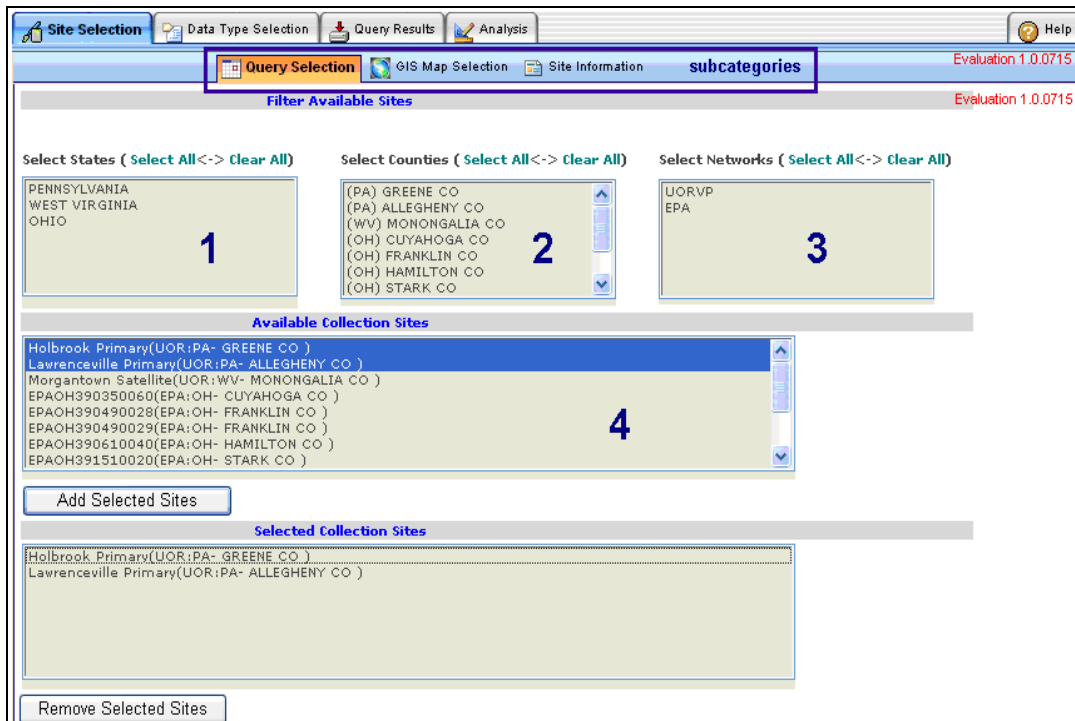


Figure 4-Site Query Selection Controls

There are three subcategories in the site selection, which are text format site selection, GIS based selection, and site-specific information of the selected sites to provide users not only text format site selection but also a Graphical User Interface (GUI) for site selection and data type available at the selected sites.

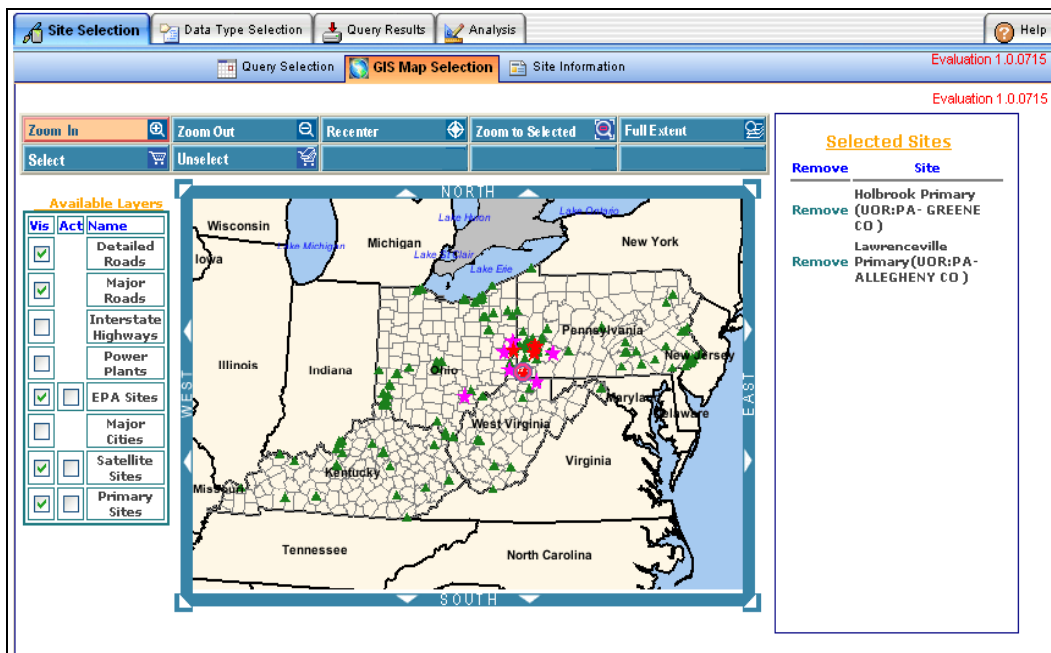


Figure 5-Site GIS Map Selection Controls

Selected Sites

Remove	SITE Name	Longitude	Latitude
Remove	Holbrook Primary(UOR:PA- GREENE CO)	-80.28501	39.81606
Site Name: Holbrook Primary Network: UOR State: PENNSYLVANIA Site Latitude: 39.81606 Start Date: 2/1/1999 12:00:00 AM County: GREENE CO Site Longitude: -80.28501 End Date: 9/30/2000 12:00:00 AM Address: Holbrook Primary			
Remove	Lawrenceville Primary(UOR:PA- ALLEGHENY CO)	-79.96217	40.43772
Site Name: Lawrenceville Primary Network: UOR State: PENNSYLVANIA Site Latitude: 40.43772 Start Date: 2/1/1999 12:00:00 AM County: ALLEGHENY CO Site Longitude: -79.96217 End Date: 9/30/2000 12:00:00 AM Address: Lawrenceville Primary			

Figure 6-Selected Site Information

Data Type Selection

The project team has worked diligently to ensure researchers have the best possible data retrieval tools at their disposal when retrieving data from the online application. Building a query follows a logical sequence and has parameters identified using standard terminology and classifications instead of data structure-specific terms and categories.

Data Type Selection

Reset Data Type Selection

Date Range
(Largest Available Date Range is Automatically Chosen by the System)

Select Start Date: 6/17/1999 Select End Date: 6/17/2002
Submit Date Change

☒ Filter Mass & Speciation ☐ Continuous TEOM ☐ Continuous Gases ☐ Continuous PM Speciation

Select Particle Size
☒ PM 2.5 ☐ PM 10 ☐ TSP ☐ All

Instrumentation: (Select All <-> Clear All)
Sequential Filter Sampler (SFS)
Federal Reference Method (FRM)
Sequential Gas Sampler (SGS)

Select Filter Mass And Species:
☐ Filter Mass

IONIC: (Select All <-> Clear All)
☐ EC ☐ OC
Chloride
Potassium
Sodium

Select All Species
Trace Elements: (Select All <-> Clear All)
Silver
Aluminum
Arsenic
Gold
Barium
Bromine
Calcium
Cadmium
Chlorine
Cobalt
Chromium
Copper

Others: (Select All <-> Clear All)
Volume Flow
Strong Acidity

Selected Parameters

Site	Parameter Name	Size	Frequency	Duration	Instrument	Start Date	End Date	Count
Holbrook Primary	Strong Acidity	2.5µ	Variable interval	24 hour	FRM	2/17/1999	8/5/2001	60
Lawrenceville Primary	Strong Acidity	2.5µ	Variable interval	24 hour	FRM	2/17/1999	8/5/2001	55

Figure 7-Data Type Selection Controls

Task 1.5 - Develop QA/QC System

The Quality Assurance/Quality Control (QA/QC) standards and processes established for this application provide for multiple layers of quality control. It is important to remember that the standards and processes mentioned in this document do not examine the quality of the data submitted, but rather ensure that the data entering the database is the same data provided by the submitting authority. The first, second and third semi-annual technical reports detail the automation techniques used to verify data integrity during the database population process.

B. Phase Two Tasks

Task 2.1 - Develop Stakeholder Website

As proposed by *ATS*, each stakeholder will have access to the entire data analysis package while the general public will have access to selected features through the public website described in Task 2.2. The stakeholder website will provide the ability to view and develop graphical representation of the digital data online for reports and for data analysis. The data analysis package will be an interactive tool that will be embedded in the data warehouse and repository. The querying of the data permits user-defined access and review of the data. Built-in online analytical tools for advanced data analysis will provide the following options:

- Dynamic/interactive charting capabilities – online graphing of the data in user-defined formats
- Trend analysis – time series of pollutant data – by species, monitor and region
- Back trajectory analysis
- Online point source modeling capabilities
- Multi dimensional plotting capabilities (three dimensions in space (x, y, z), and time)
- Statistical analysis of pollutant profiles and distributions
- Meteorological evaluations (influence on air pollutant concentrations)

The Stakeholder Website is being developed using Microsoft Visual Studio .NET, in conjunction with Microsoft Internet Information Services (IIS), Microsoft SQL Server 2000 and the .NET framework (a packaged addition for MS Windows 2000 or XP). The project team decided to utilize the .NET framework early in the planning stages of this project because of the extensive tool sets available for this platform and the tight integration of XML Web Services into the product. XML Web Services allow remote users to retrieve datasets locally, combine multiple data sources into a single dataset and exchange data with other datasets that may, or may not, be directly related to the PM_{2.5} data. This is crucial to be able to efficiently serve large quantities of data to researchers desiring to query for non-obvious associations between Particulate Matter data and other data sources, such as health-related statistical data sources.

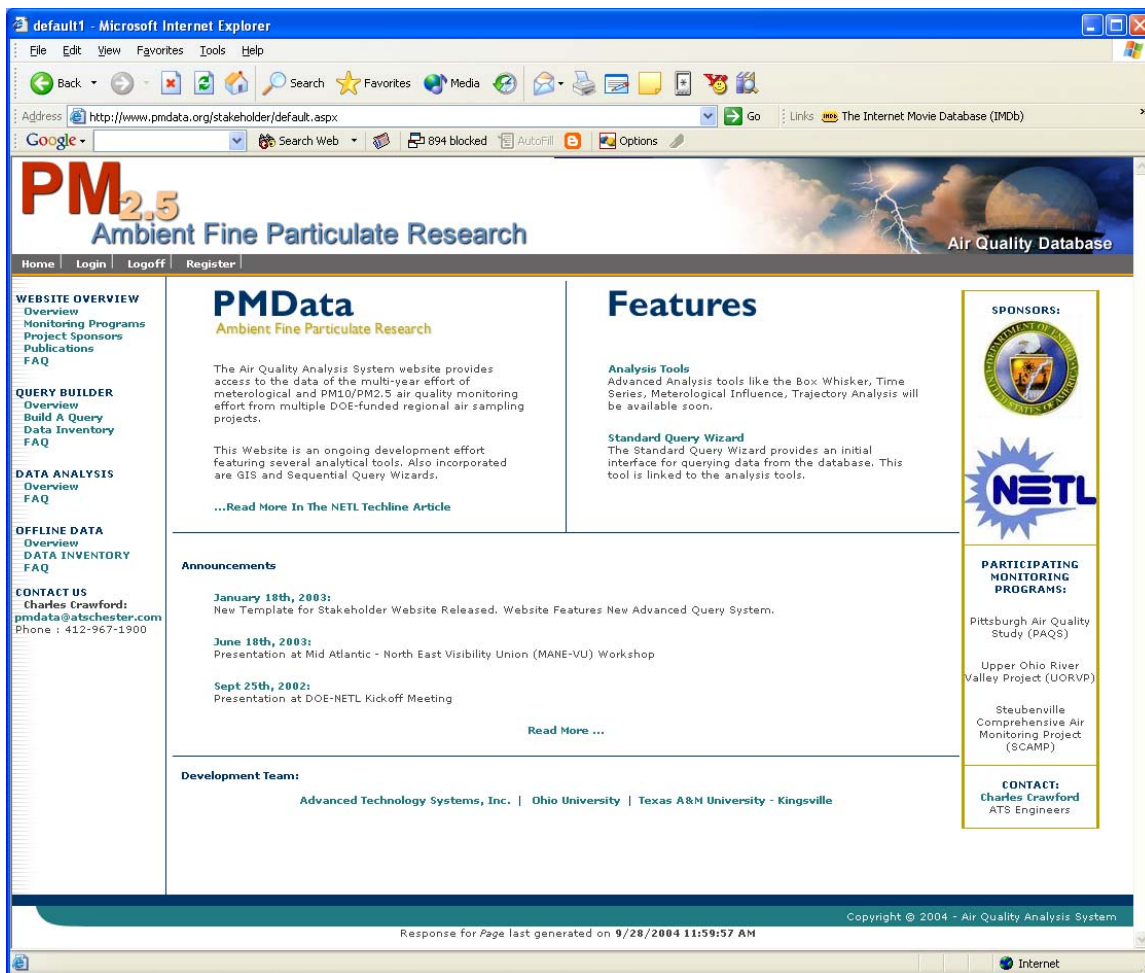


Figure 8-Screenshot of Stakeholder Website

As part of the development process, the project team will engage stakeholders to provide critical feedback so the development team can create *meaningful* and *useful* analysis tools. The stakeholder involvement will include an interactive web-cast and demonstration with key personnel from our project team and stakeholders.

A series of web casts for beta-testing of this development work have been conducted with a select group of stakeholders who have provided critical feedback on the types of analytical tools and interfaces which would be most valuable to the research and government/industry communities. The web casts have involved online demonstrations of the website and the existing analysis tools. This group of stakeholders will be used as a focus group to set priorities and metrics for the refining of this development process.

Another beta-testing web cast with an expanded stakeholder group is being planned for October 2004.

Task 2.2 - Develop Public Website

ATS proposed to construct a separate website connected to the data archive for public outreach, providing the citizens of the upper Ohio River valley and at-large, along with legislative and regulatory authorities, a resource and an educational tool highlighting the extensive monitoring programs undertaken by NETL. Publicly accessible sections of the

database application's web space will be available to everyone without log in. This portal will be different from the one for stakeholders, sponsors and developers, and will require registered users to submit a username and password combination before access to the restricted website is granted.



Figure 9-Screenshot of Public Website

This interactive web-based application will be the backbone of the public outreach system. The web delivery system will be designed as an information/decision support center and an educational tool. The system will provide clear and concise data summaries from the monitoring programs and will include easy-to-understand graphical representation of the data including spatial and temporal mapping of the data accompanied by the online help as described in Task 2.3. To insure that the website will deliver information in a clear and concise manner, the deliverables of this task will be reviewed continuously by environmental and community representatives from the region prior to launching.

To date, we have developed a preliminary version of the public website which was used by DOE-NETL to advertise data availability and included a data retrieval tool to download the original data files associated with this project.

Task 2.3 - Develop Online Help Feature

ATS proposed to construct an online help feature, in conjunction with the web-based application, and it will be developed to support both the sponsor/stakeholder and the public sections website. The online help and instruction component of the application will be an interactive system that will give depth, understanding and context to the environmental data presented. The online help will assist the user at any level of scientific background (novice to professional) in the interpretation of the data. The online help will provide assistance on the following general topics:

- Definitions that will provide clear explanations of the terminology used in evaluating air pollutants
- Explanation of the Federal and State Regulations pertaining to criteria pollutants
- Background information on atmospheric chemistry, transport and emissions of air pollutants
- Effects of meteorology on air pollution episodes
- Significance of the data as it relates to public health
- Information on community-based efforts that can impact ambient air pollution levels
- Navigation of the website itself

Task 2.4 - Provide Graphical Representation of Data

The graphing and analysis tools for this project have been developed with ChartFX for .NET graphing and charting controls using the C# .NET Web Forms environment. ChartFX for .NET graphing and charting controls generate all the graphs on the server and generate downloadable image files from a cached dataset on the MS SQL server. The major changes to the analysis tools during the previous six months include the addition of selection options and the utilization of session variables to improve data query performance.

To date, one calculation tool and five analysis tools, which are time series plot, box-and-whisker plot, frequency histogram, site comparison by scatter plot, and site comparison by box-and-whisker plot, have been developed and published on a web site for stakeholders to evaluate. During the previous period one calculation tool and three analysis tools, which are frequency histogram, site comparison by scatter plot, and site comparison by box-and-whisker plot, have been incorporated into the overall toolset through the development and evaluation process.

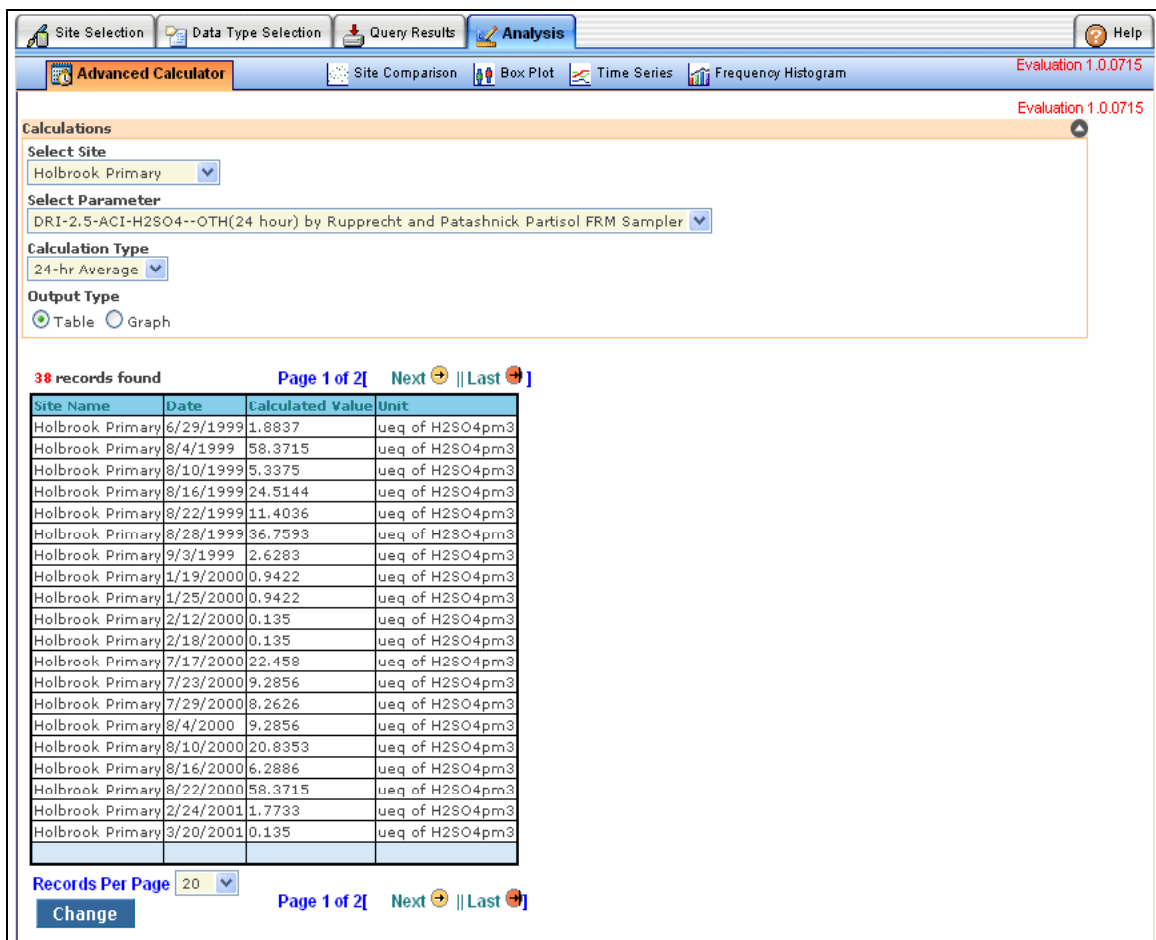


Figure 10-Advanced Calculator

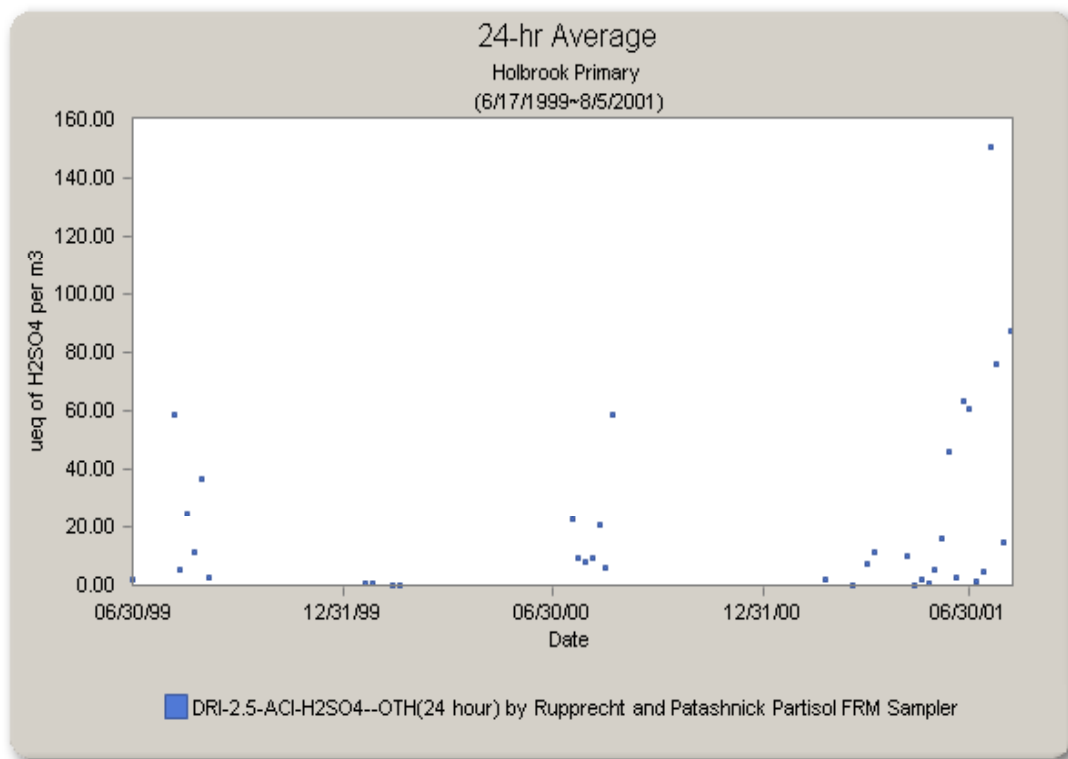


Figure 11-Graph of Calculator Generated Data

The calculation tool calculates daily maximum values of 1-hour average and/or 24-hour average values dynamically for the different types of dataset stored in various durations including 5-minute, 30-minute, 1-hour, 6-hour, and 24-hour. Additional calculation options including daily maximum values of 8-hour average, monthly average, and yearly average are under development.

Time series plot shows a trend of selected dataset and has the ability to show multiple sites trends on one graph. Box-and-whisker plots display a statistical summary of five variables, such as the most extreme values in the data set (the maximum and minimum values), the lower and upper quartiles (25%~75%), and median, by different time range including day of week, weekdays/weekends, and month. Frequency histogram replacing pie chart represents a frequency distribution by a bar chart whose heights represent observed frequencies. Site comparison by scatter plot shows geographical correlations between two sites for a certain parameter or between two parameters for a certain site. Site comparison by box-and-whisker plot shows an overview of statistical summary for several sites in one graph. Currently pie chart for chemical speciation, back trajectory and cluster analysis tools are under development. The project team is actively seeking input from stakeholders regarding desired analysis tools. All new tools requested from stakeholders have an evaluation and development process to incorporate these tools into the overall toolset.

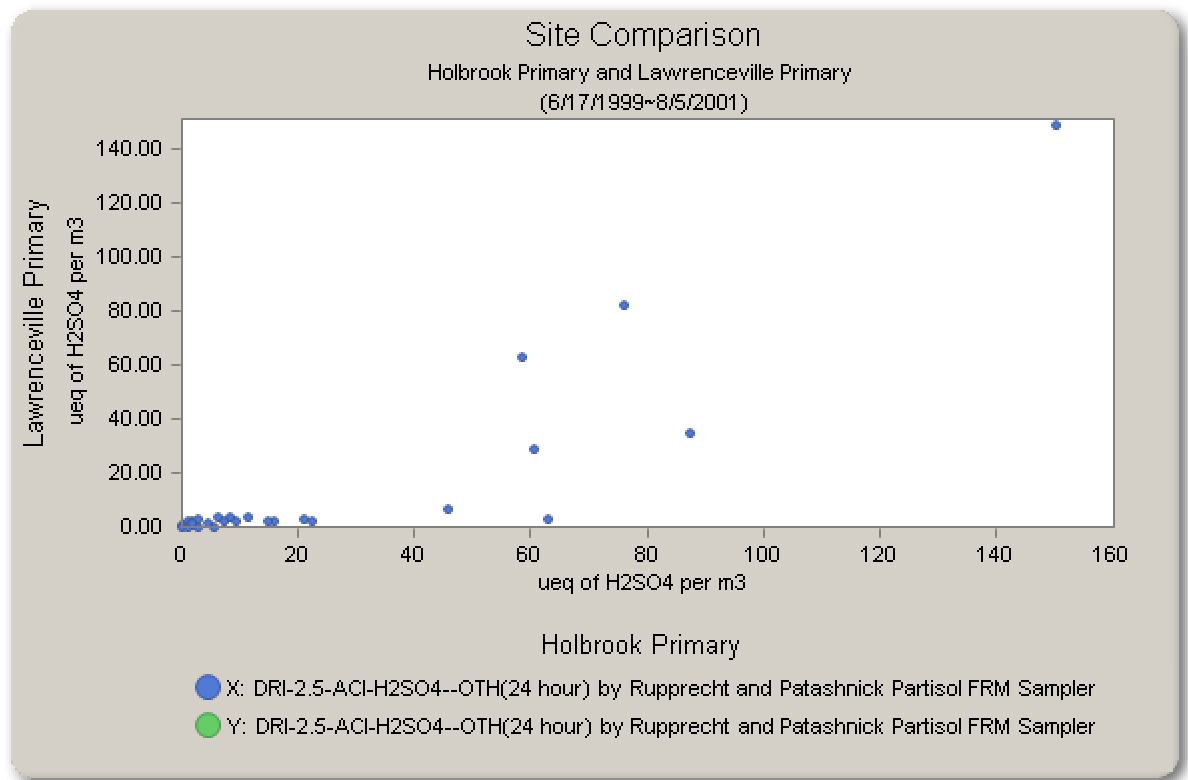


Figure 12-Site Comparison Tool

Options

X-axis
Site: Holbrook Primary
Pollutant: DRI-2.5-ACI-H₂SO₄--OTH(24 hour) by Rupprecht and Patashnick Partisol FRM Sampler

Y-axis
Site: Lawrenceville Primary
Pollutant: DRI-2.5-ACI-H₂SO₄--OTH(24 hour) by Rupprecht and Patashnick Partisol FRM Sampler

From: 6/17/1999 **To:** 6/17/2002 **Priority:** Site

Graph Options
☐ Grid Line (X-axis) ☐ Grid Line (Y-axis) ☒ Legend
 Max:
 Min:

Figure 13-Site Comparison Tool Options

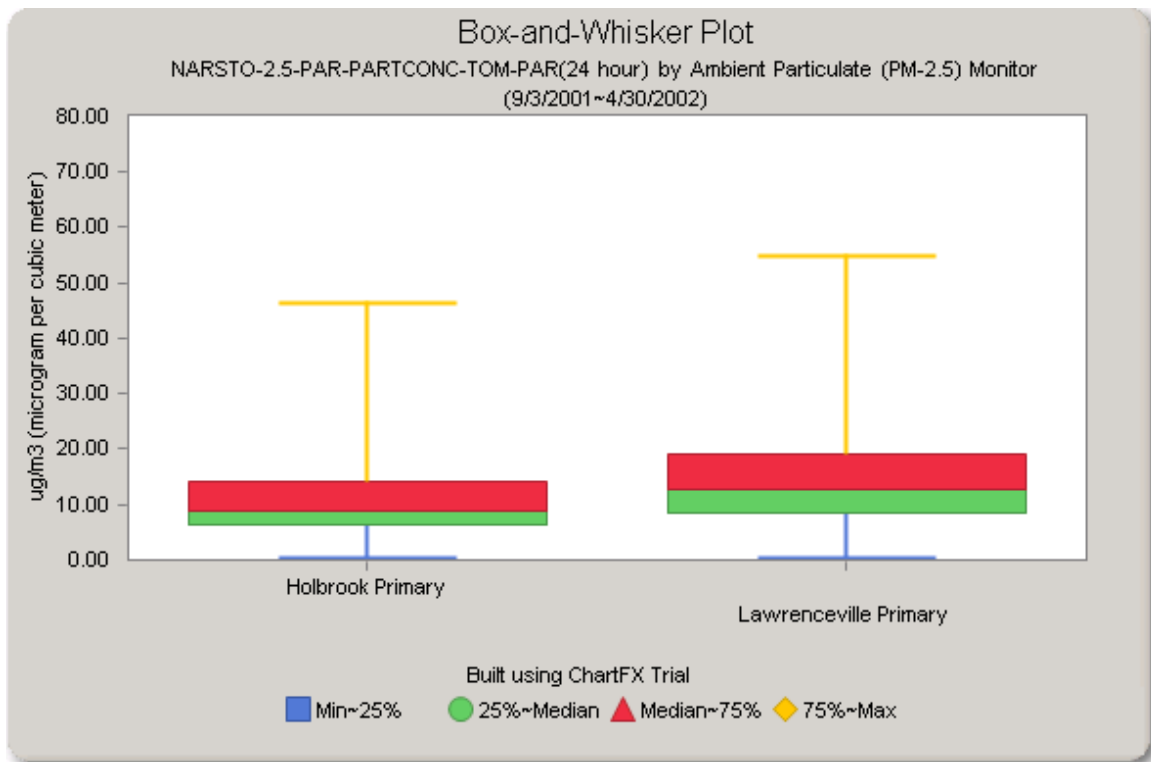


Figure 14-Site Comparison Tool (Box and Whisker Plot)

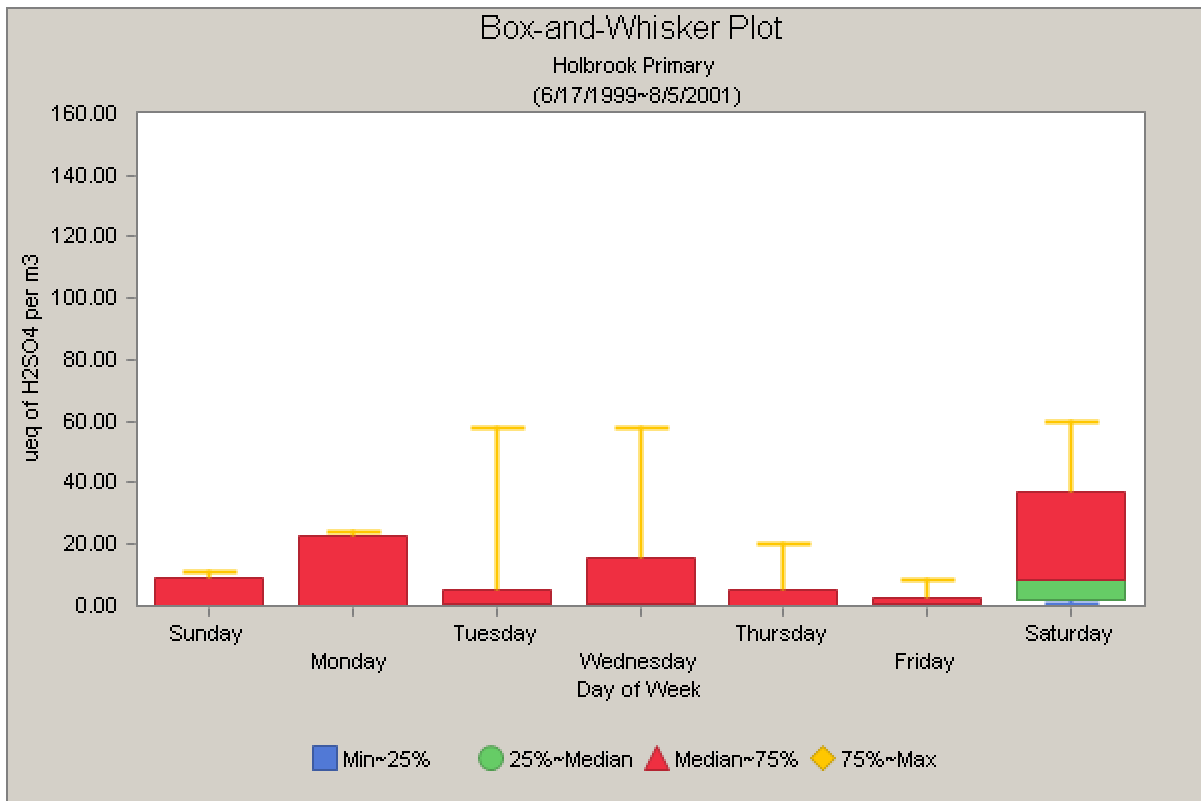


Figure 15-Box and Whisker Plot

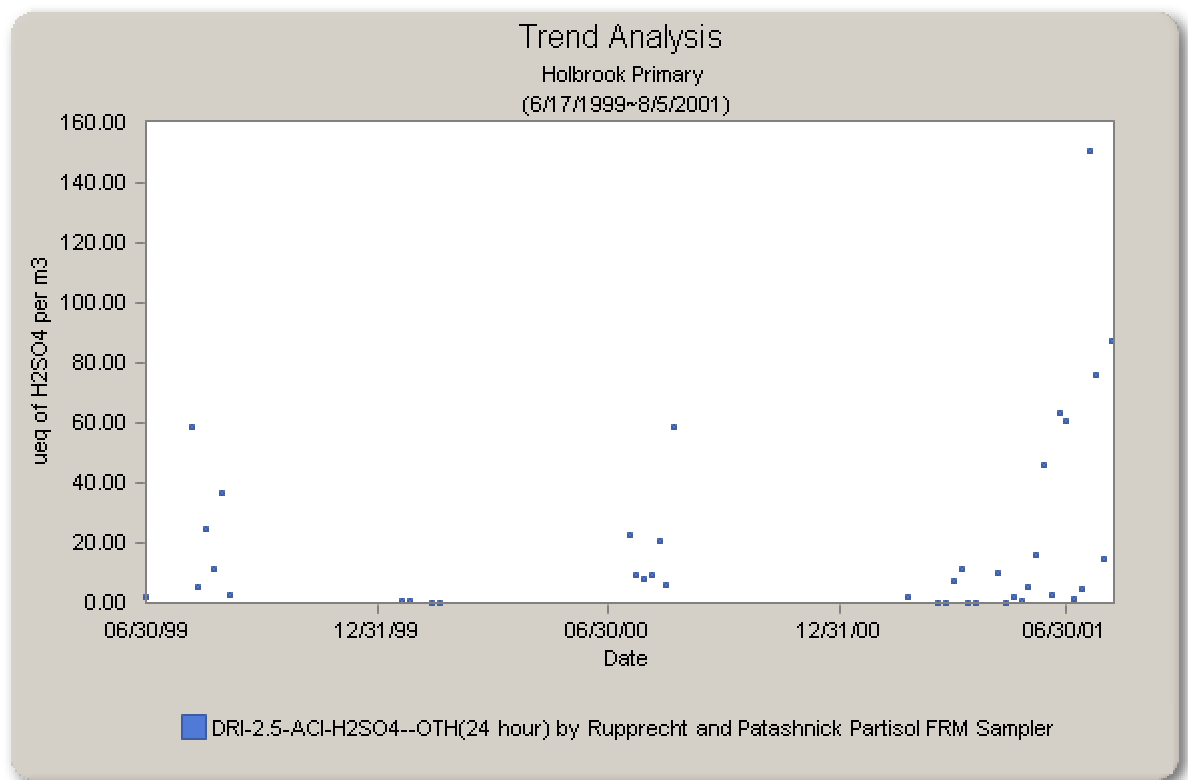


Figure 16-Time Series Plot

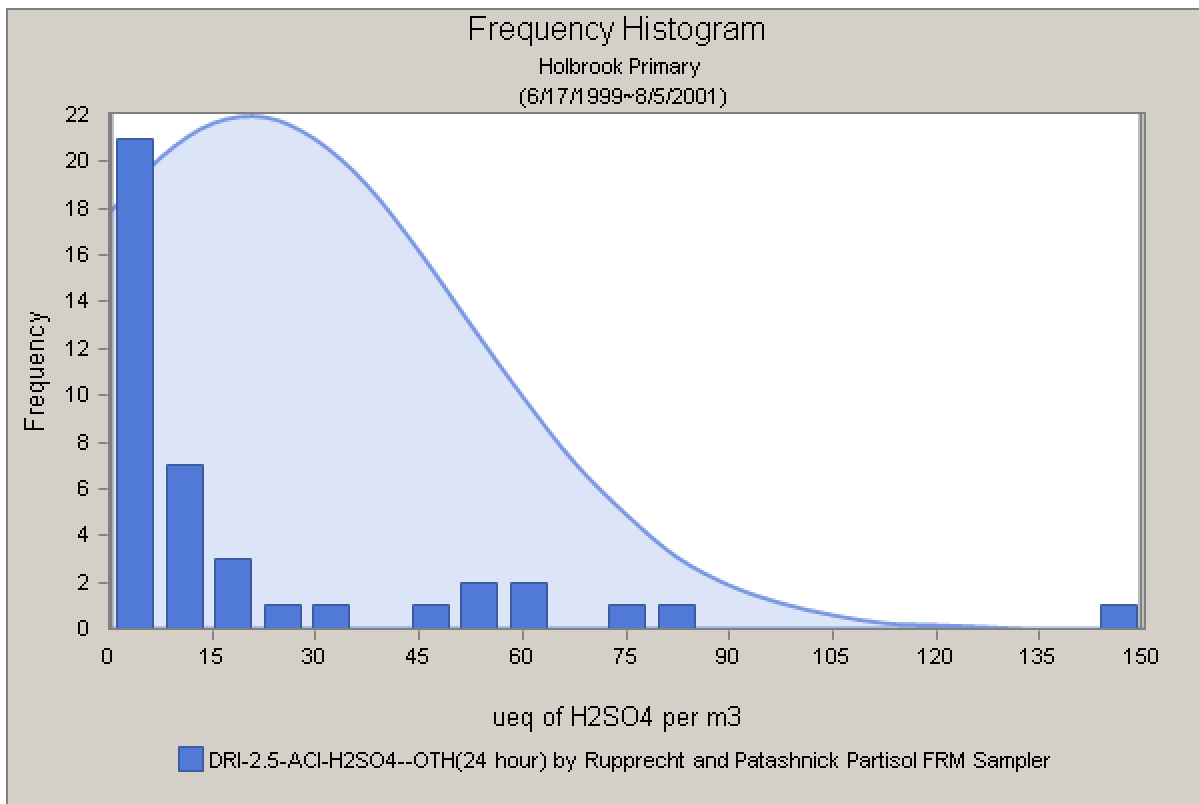


Figure 17-Frequency Histogram

Task 2.5 - Performance Testing

ATS, in coordination with all members of the external stakeholder group, will conduct a detailed testing program for the resulting application to verify the functionality and proper execution of all portions of this application. This testing program will provide for interactive user feedback, discussion forums and periodic email notifications and announcements. This testing program will help insure that the intended objectives of this project are met or exceeded. This effort may require revisiting and reworking the original designs, and consequently, will be an ongoing exercise in Phase Two of this project. A section of the website will be dedicated for posting of feedback from beta-testers of this application.

III. RESULTS AND DISCUSSION

The first phase includes data inventory, benchmarking and database population tasks, as well as the development of data management architecture, a web-based retrieval system and an internal QA/QC system. A progress summary for Phase One is shown in the following table:

Task #	Description	Planned completed %	Actual completed %
1.1	Data Inventory/Benchmarking for Database Applications	100%	100%
1.2	Develop Data Management System Architecture	100%	98%
1.3	Population of Database	100%	90%
1.4	Develop Web-based Retrieval System	100%	95%
1.5	Develop Internal QA/QC System	N/A	N/A

N/A – on-going activity.

The second project development phase, currently in progress, is slated for completion by August 2005. The tasks involved in this phase include the development of a stakeholder-specific website, a publicly accessible website and an online help feature. This phase also includes the development of special analysis tools to provide a graphical representation of the data and, of course, a series of performance tests designed to provide the best possible data management solution. A progress summary for Phase Two is shown in the following table:

Task #	Description	Planned completed %	Actual completed %
2.1	Develop Stakeholder Website	100%	65%
2.2	Develop Public Website	100%	35%
2.3	Develop Online Help Feature	100%	5%
2.4	Provide Graphical Representation of Data	100%	35%
2.5	Performance Test	100%	5%

IV. CONCLUSION

The development efforts have so far proceeded as expected. Some target milestones have not been met primarily due to delays in acquiring input data from third party sources. This is especially so with the data inventory task, where data reformatting issues have also been encountered. Delays have also been an inevitable consequence of the philosophy of the DOE COR and the project team to proceed very carefully and deliberately with the development of the stakeholder website. It is believed that the ultimate success of this project will require a high degree of stakeholder confidence and subsequent participation in the website development process, and that such participation will be greatly enhanced if stakeholders are presented with a relatively “polished” product at the outset. Therefore, the extra programming effort has been dedicated toward developing and refining a limited set of fully-functional graphic and analytical routines (e.g., time series analysis and box plots) before fully pursuing a potentially expensive program of stakeholder engagement. The DOE COR has also been actively involved in an on-going evaluation and β -testing of the developing website and analytical tools, proving critical feedback that has been instrumental in modifications that have made the application more user-friendly and the navigation much more dynamic.

The hurdles encountered, however, have not been insurmountable, as other parts of the project have proceeded as proposed. Since the level of effort associated with the slowed tasks is still the same, the costs to complete this task will not be impacted either negatively or positively. Therefore, a request for a twelve-month, no-cost, time extension to August of 2005 was submitted to DOE-NETL and approved in March 2004.

V. REFERENCES

1. **O'Brien, G., Hughes, V., Modeling Support, California Air Resources Board**
[Designing a Data Management System for the Central California Air Quality Studies](http://www.arb.ca.gov/airways/Documents/DOCS/AWMACCAQS_Mar2301.pdf)
http://www.arb.ca.gov/airways/Documents/DOCS/AWMACCAQS_Mar2301.pdf
2. **Professors Spyros Pandis, Cliff Davidson, and Allen Robinson, Carnegie Mellon University**
[EPA Supersite - Carnegie Mellon University](http://homer.cheme.cmu.edu)
<http://homer.cheme.cmu.edu>
3. **Robinson P. Khosah, Ph.D., Advanced Technology Systems, Inc.**
[Semi-Annual Technical Progress Report](http://www.netl.doe.gov/coalpower/environment/air_q/docs/40456r06.pdf)
http://www.netl.doe.gov/coalpower/environment/air_q/docs/40456r06.pdf
4. **Karen Magliano, California Air Resources Board; Philip Roth, Envair; Charles Blanchard, Envair; Steven Reynolds, Envair; Steve Ziman, Chevron; Rob DeMandel, Bay Area Air Quality Management District**
[California Regional Pm₁₀/Pm_{2.5} Air Quality Study: Objectives and Associated Data Analysis and Modeling Approaches](http://www.arb.ca.gov/airways/Documents/reports/objectiv.doc)
<http://www.arb.ca.gov/airways/Documents/reports/objectiv.doc>
5. **Robinson P. Khosah, Ph.D., Charles G. Crawford, Advanced Technology Systems, Inc.**
[Semi-Annual Technical Progress Report](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R01.pdf)
http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R01.pdf
6. **Robinson P. Khosah, Ph.D., Charles G. Crawford, Advanced Technology Systems, Inc.**
[Semi-Annual Technical Progress Report](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R02.pdf)
http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R02.pdf

VI. BIBLIOGRAPHY

Not applicable.

VII. LIST OF TERMS, ACRONYMS AND ABBREVIATIONS

Term	Definition
Admin Level	Security level indicating the degree of access a specific user possesses to administration utilities and data.
ACHD	Allegheny County Health Department
APM	Automated Population Module
Application Account	An application account (Windows 2000). This account is created and edited using Windows 2000.
Application Administrator	An individual responsible for managing application performance, user access and newsletter/announcement notification services.
Browse Level	Security level indicating the areas of the application and datasets that a specific user can 'browse' through.
Cached Data	Data retained at the server level to serve frequently polled data. These datasets are cached, or stored, at the server and reduce query loads on the database server, thus increasing overall efficiency and application response time.
CARB	California Air Resources Board
CSV	Comma Separated Value
Data Administrator	An individual responsible for managing the database housing the PM data, as well as managing all assigned data submission accounts.
Data Submitter	Individual user who has permission to submit data for inclusion in the PM database.
DOE-NETL	US Department of Energy's National Energy Technology Laboratory
Foreign Key	A non-negative whole number used to reference a data row in a related table.
FTP	File Transfer Protocol.
GIS	Geographic Information Systems
GMT Offset	Number of hours that, when added to the local time values, provides GMT Time values; e.g. 11:00AM local time, with a GMT offset value of -5 means that the GMT time value for this local time value would be 6:00AM GMT.

HTTP	Hypertext Transfer Protocol
Media	Filter used to collect speciation samples.
Metaflag	Localized flagging system specific to a particular submitting authority.
Method	Descriptive text that describes how data was collected.
NARSTO	An acronym for "North American Research Strategy for Tropospheric Ozone." A tri-national, public-private partnership for dealing with multiple features of tropospheric pollution, including ozone and suspended particulate matter.
NARSTO Metaflag	Standardized flagging system (NARSTO). Each metaflag is mapped to a NARSTO metaflag to provide meaningful results when querying across datasets originating from multiple submitting authorities.
NOAA	National Oceanic and Atmospheric Administration
PA-DEP	Pennsylvania Department of Environmental Protection
Parameter	A concatenated descriptive definition of what the observation value represents. Components of a valid parameter include: parameter property, parameter identifier, collection principle, and parameter source.
Parameter Identifier	Descriptive text that identifies a chemical property of a parameter.
Parameter Property	Descriptive text that identifies a physical property of a parameter.
Primary Key	Unique non-negative whole number used to reference each row in a database table. This is used to identify relationships between related items in related tables.
Parameter Source	Originating organization for parameter codes and descriptions.
QA/QC	Quality Assurance / Quality Control
QC Status	Quality control status code.
Read Level	Security level indicating the areas of the application and datasets to which a specific user has read access.
Sample Duration	Text describing the sample duration that is used to collect a specific sample. This usually applies only to filter data; sample duration of H12 indicates that the sample in question was taken over a 12-hour period.
Sample Frequency	Text describing the sample frequency, or interval, between regular readings; e.g. M15 indicates that a sample is taken every 15 minutes.

SQL	Structured Query Language
Subscriber	Individual user who has elected to receive email notification from pmdata.org.
System Account	A Windows 2000 Server account used to administer the network and/or application servers.
Systems Administrator	An individual responsible for managing the hardware and operating system(s) of the hosting computers and networks. This person ensures that the application and database is available to users and works to correct any connectivity issues that may occur.
User Account	Application account established for each user that contains each user's contact data and security profile.
US EPA	US Environmental Protection Agency
VCard	Virtual address card. This is similar to a rolodex entry, containing an address, city, state and zip code. A VCard can link to multiple entities sharing the same physical address. Entries also contain a location's county and country.
Write Level	Security level indicating the areas of the application and datasets to which a specific user may enter new records or modify existing records.
WV-DEP	West Virginia Department of Environmental Protection
XML	Extensible Markup Language