

**DATABASE AND ANALYTICAL TOOL DEVELOPMENT  
FOR THE MANAGEMENT OF DATA  
DERIVED FROM US DOE (NETL) FUNDED  
FINE PARTICULATE (PM<sub>2.5</sub>) 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<sub>2.5</sub>) RESEARCH**

## **Semi-Annual Technical Progress Report**

**REPORT PERIOD START DATE:** August 12, 2004

**REPORT PERIOD END DATE:** February 11, 2005

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

**DATE REPORT ISSUED:** February 2005

**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 thirtieth 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

LIST OF GRAPHICAL MATERIALS .....	ii
LIST OF TABLES .....	ii
EXECUTIVE SUMMARY .....	i
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 .....	4
Task 1.5 - Develop QA/QC System .....	8
B. Phase Two Tasks .....	8
Task 2.1 - Develop Stakeholder Website .....	8
Task 2.2 - Develop Public Website .....	10
Task 2.3 - Develop Online Help Feature .....	11
Task 2.4 - Provide Graphical Representation of Data .....	11
Task 2.5 - Performance Testing .....	18
III. RESULTS AND DISCUSSION .....	18
IV. CONCLUSION .....	19
V. REFERENCES .....	20
VI. BIBLIOGRAPHY .....	21
VII. LIST OF TERMS, ACRONYMS AND ABBREVIATIONS .....	22

## LIST OF GRAPHICAL MATERIALS

Figure 1 - Query Process Map .....	5
Figure 2 - New Tab Strip Control .....	6
Figure 3 - New Toolbar Control .....	6
Figure 4 - Database Record Inventory Control - grouped by parameter .....	7
Figure 5 - Database Record Inventory Control - grouped by site .....	7
Figure 6 - NARSTO QA/QC Flags Filter Control .....	8
Figure 7 - Screenshot of Stakeholder Website .....	9
Figure 8 - Screenshot of Public Website .....	10
Figure 9 - Data Calculation Tool .....	12
Figure 10 - Time Series Plots .....	12
Figure 11 - Box-Whisker Plots .....	13
Figure 12 - Frequency Histograms .....	13
Figure 13 - Site Comparison Graphs .....	14
Figure 14 - Speciation Filter Composition Tool with Timeline Control .....	15
Figure 15 – Sample Trajectory Map .....	16
Figure 16 - Sample Cluster Analysis Map .....	16
Figure 17 - Graph Options Panel .....	17
Figure 18 - Toolbar for .NET Enabled Charts .....	17

## LIST OF TABLES

Table 1 - Phase One Progress Summary .....	2
Table 2 - Phase Two Progress Summary .....	2
Table 3 - Phase One Progress Summary .....	18
Table 4 - Phase Two Progress Summary .....	19

## 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<sub>2.5</sub> 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 thirtieth 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<sub>2.5</sub> 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 thirtieth 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 table 1.

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

**Table 1 - Phase One Progress Summary**

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 table 2.

<b>Task #</b>	<b>Description</b>	<b>Planned completed %</b>	<b>Actual completed %</b>
2.1	Develop Stakeholder Website	100%	85%
2.2	Develop Public Website	100%	65%
2.3	Develop Online Help Feature	100%	15%
2.4	Provide Graphical Representation of Data	100%	75%
2.5	Performance Test	100%	20%

**Table 2 - Phase Two Progress Summary**



## 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.

The Data Management System Architecture design is complete and has been implemented for this project. Additions and changes to stored procedures are made as required, but these changes do not significantly impact the overall design and function of this architecture.

#### 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.

The project team *pre-processed*, or separated the data files by species and filter ID (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 H<sub>2</sub>SO<sub>4</sub>, or Sulfuric Acid.

Other datasets were much easier to import into the system, 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. 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 data from NETL and SCAMP has not yet been received by the project team and the Pittsburgh AQS data is still undergoing QA/QC checks after data preprocessing operations.

The project’s software developers are currently updating code to accommodate the automatic uploading of NARSTO formatted data files so that all of the Pittsburgh AQS data can be imported directly into the database, along with other NARSTO formatted files as required.

### **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. Previous Semi-Annual Technical Reports have detailed both the static data downloads controls and the dynamic query systems. In the past six months, the format of the static data download controls have remained the same, while the dynamic query systems have improved dramatically.

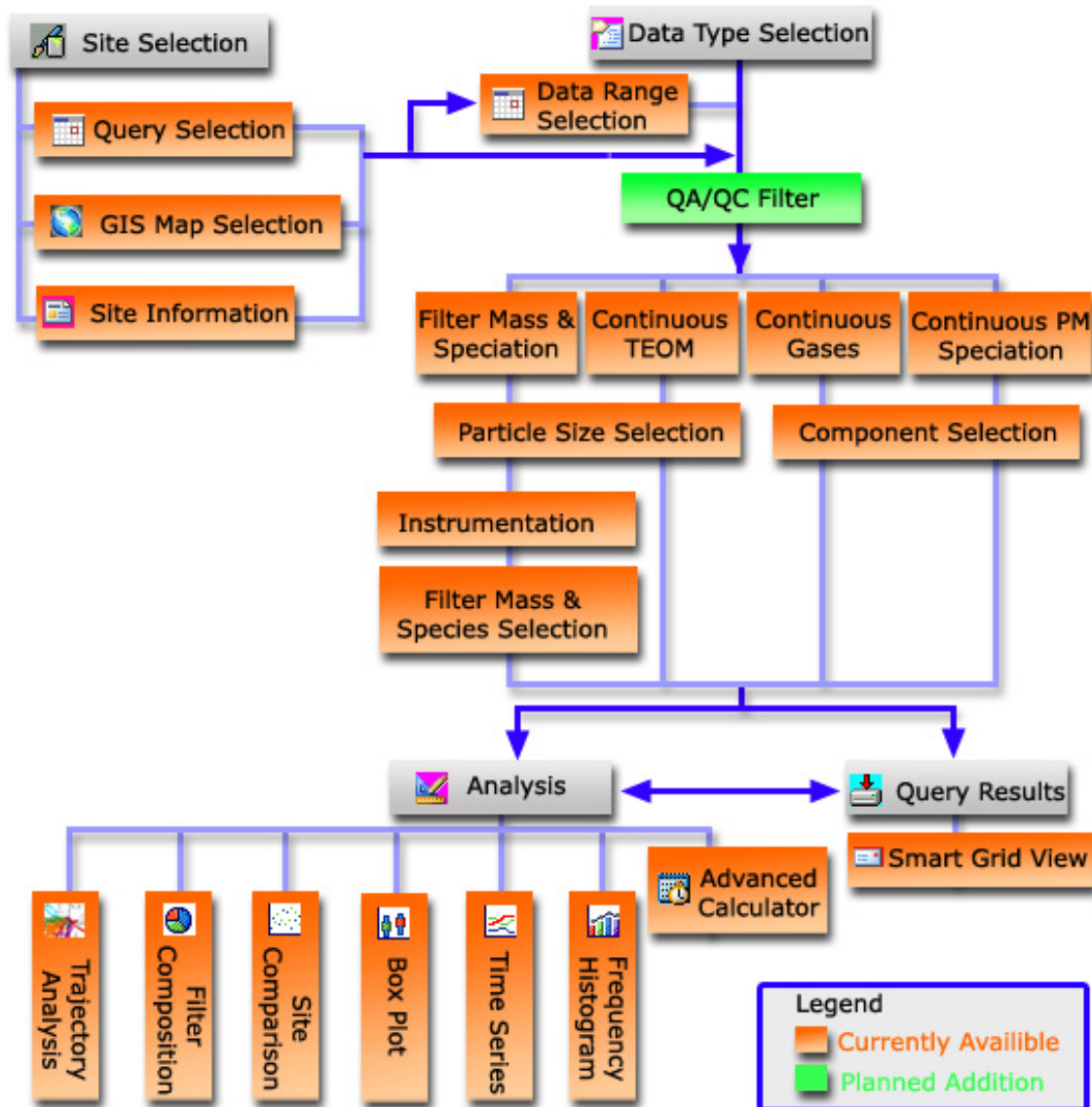
### **DYNAMIC (AD HOC) QUERY SYSTEMS**

#### ***Overview Revisited***

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. In the previous report, a process map was provided which outlined the

entire query builder system. In the past six months some analysis tools have been added which require adjustment to the query process map provided in that report. Figure 3 shows the new query process map and details the data categories used to filter the data choices based on previously selected criteria.



**Figure 1 - 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.

## *Interface Changes*

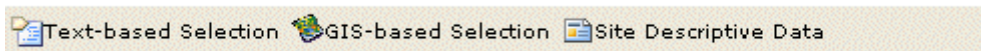
The previous Semi-Annual Technical Progress Report detailed several interface features of our query building engine. These interface features allow users to build a custom query based on filtering criteria the user selects from a series of controls. These controls are separated into logical groups. These groups are arranged and accessed through a series of tabs which switch the user to different parts of the query builder toolset.

Since this report was issued, however, several of these controls have been either replaced, or re-designed to accommodate a larger set of capabilities in the data retrieval process. One example of such interface alteration is the use of a different tab strip control. Figure 2 shows the new tab strip control in use at this time. The new tab strip control looks cleaner, has better response times, and eliminates any scrolling issues encountered with the previous control.



**Figure 2 - New Tab Strip Control**

The new toolbar control is also an addition to the controls which comprise the query builder toolset. Figure 3 shows the new toolbar control, which provides a clickable control to access specific sections of the query builder tools. This addition allows the programmers to group controls together based on functional divisions used for the tab headings and subdivisions or categories used in the toolbar. This control provides the same functionality provided by the previous tab strip control toolbar, which was more restrictive in appearance and location.



**Figure 3 - New Toolbar Control**

## *New Controls*

The Database Record Inventory controls are new to the query builder. These controls are nested data grid controls which group data by data collection site or by parameter. These controls allow users to obtain a birds-eye view of the data in the database. Users can quickly see how many data records are in the database and which parameters are available in this collection of data records. Alternatively, users can also choose a parameter and quickly identify how many records exist for that parameter at a given data collection site. Figure 4 shows the database record inventory data grouped by parameter, and Figure 5 shows the database record inventory data grouped by data collection site.

## Database Record Inventory

☐ Data Inventory by Site

☒ Data Inventory by Parameter

< >		
Parameter		Record Count
<b>PM2.5 Particle-Concentration [TEOM]</b>		<b>164650</b>
Parameter ID	Collection Site Name	Record Count
133	Holbrook Primary	88903
133	Lawrenceville Primary	75747
PM2.5 Element-TL [Filter+denuder]		143309
PM10 Particle-Concentration [TEOM]		90253
PM2.5 Particle-Mass [TEOM]		56794
PM10 Particle-Mass [TEOM]		21779
PM2.5 Flow-Vol [Not assigned]		4538
PM2.5 Element-AG [Filter+denuder]		3129
PM2.5 Filter-Mass [Filter+denuder]		2711
PM10 Flow-Vol [Other]		2368
PM2.5 Element-AL [Filter+denuder]		1701
PM2.5 Element-AS [Filter+denuder]		1701
PM2.5 Ionic-CHL [Filter+denuder]		1638

Page 1 - of - 10

Figure 4 - Database Record Inventory Control - grouped by parameter

## Database Record Inventory

☒ Data Inventory by Site

☐ Data Inventory by Parameter

< >		
Data Collection Site		Record Count
Lawrenceville Primary		258641
<b>Holbrook Primary</b>		<b>134066</b>
Site ID	Description	Record Count
0	PM2.5 Particle-Concentration [TEOM]	88903
0	PM2.5 Particle-Mass [TEOM]	32122
0	PM2.5 Flow-Vol [Not assigned]	1028
0	PM2.5 Element-AG [Filter+denuder]	740
0	PM10 Flow-Vol [Other]	638
0	PM2.5 Filter-Mass [Filter+denuder]	634
EPAPA420958000		23181
EPAPA420050001		19036
EPAPA420030064		17990
EPAOH390490029		15834
EPAOH390490028		15665
EPAOH390610040		14516
EPAPA420010001		13991
EPAPA420030008		13706
EPAOH391510020		4588
EPAOH390350060		4235

Page 1 - of - 2

Figure 5 - Database Record Inventory Control - grouped by site

### Coming Soon

Future versions of the query builder will include a new control used to filter various levels of quality control flags. Given that the obtained data, formatted according to the NARSTO specifications, contains the broadest set of quality control flags, this filter allows users to select the types of data to retrieve based on the standard NARSTO flag definitions. Some datasets contain more detailed meta-data in locally-defined quality

control flags, but each locally-defined flag corresponds to a suitable and corresponding NARSTO flag. Figure 6 shows the Quality Control Flag Filter control.

Metaflag	NARSTO Metaflag Description
NF	Not flagged.
V0	Valid value.
V1	Valid value but comprised wholly or partially of below detection limit data.
V2	Valid estimated value.
V3	Valid interpolated value.
V4	Valid value despite failing to meet some QC or statistical criteria.
V5	Valid value but qualified because of possible contamination (e.g.; pollution source; laboratory contamination source).
V6	Valid value but qualified due to non-standard sampling conditions (e.g.; instrument malfunction; sample handling).
V7	Valid value but set equal to the detection limit (DL) because the measured value was below the DL.
M1	Missing value because no value is available.
M2	Missing value because invalidated by data originator.
H1	Historical data that have not been assessed or validated.

Figure 6 - NARSTO QA/QC Flags Filter Control

## 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 have been provided with 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
- Statistical analysis of pollutant profiles and distributions
- Back trajectory analysis
- Speciation filter composition analysis

Development efforts are also underway to test and potentially add the following options to this list:

- Online point source modeling capabilities
- Multi dimensional plotting capabilities (three dimensions in space (x, y, z), and time)
- 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<sub>2.5</sub> data. Figure 7 is a screenshot of the stakeholder website.

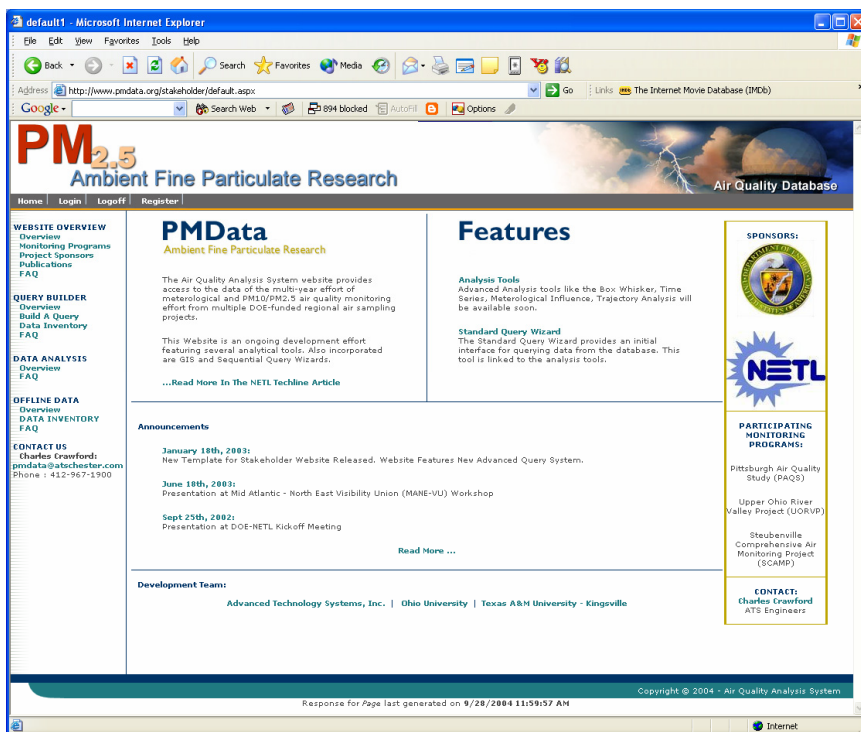


Figure 7 - Screenshot of Stakeholder Website

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. To date, a series of web casts have been conducted, with more planned for the near future, and the capabilities and potentials for this web-based application have been presented and discussed.

The focus audience members for these web casts are a select group of stakeholders who can provide 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 involve an online demonstration of the website and existing analysis tools. This group of stakeholders is being used as a focus group to set priorities and metrics for the development process and to test development efforts already underway.

## 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, which will require registered users to submit a username and password combination before access to the restricted website is granted. A screenshot of the public website is shown in figure 8.



Figure 8 - 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 fourth Semi-Annual Technical Progress Report detailed several analysis tools used to graphically represent selected datasets. These controls, including:

1. data calculation tool, shown in Figure 9
2. time series plots, shown in Figure 10
3. box-whisker plots, shown in Figure 11
4. frequency histograms, shown in Figure 12
5. site comparison graphs using scatter plots, shown in Figure 13

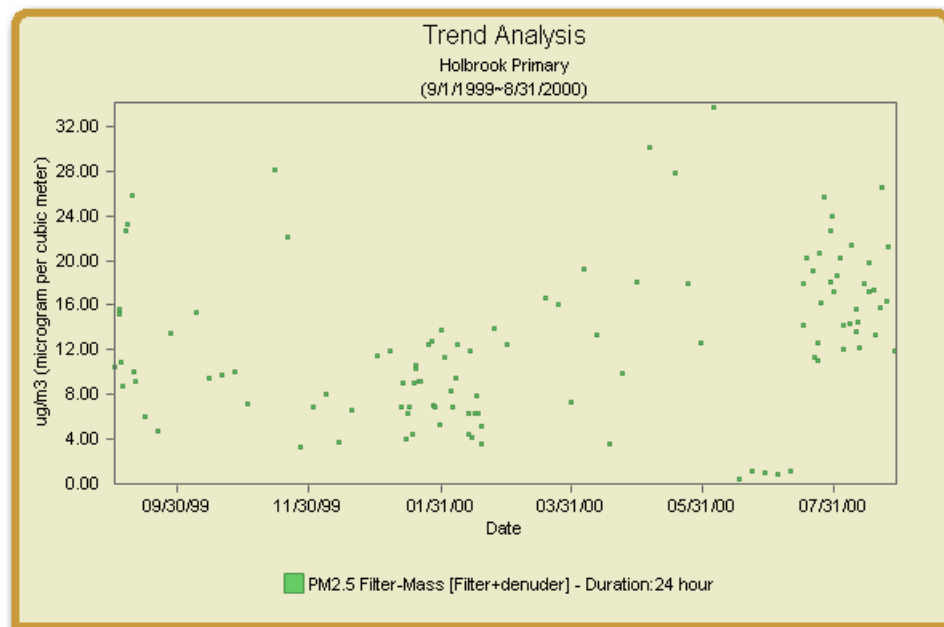
In the past six months, these tools have been augmented with a filter composition tool which groups data into broad categories and depicts a pie chart representing the

compositional makeup of all filter mass readings over a given time period. This tool, shown in Figure 14, also contains a timeline control which users see a bar chart with each vertical bar representing one day's readings. When the user first loads the tool, a pie chart is generated for the entire time period. Once the user selects a bar or day from the timeline control, the bar representing that day is colored red to distinguish it from the rest of the data, and the pie chart is regenerated to represent only the selected day's data.

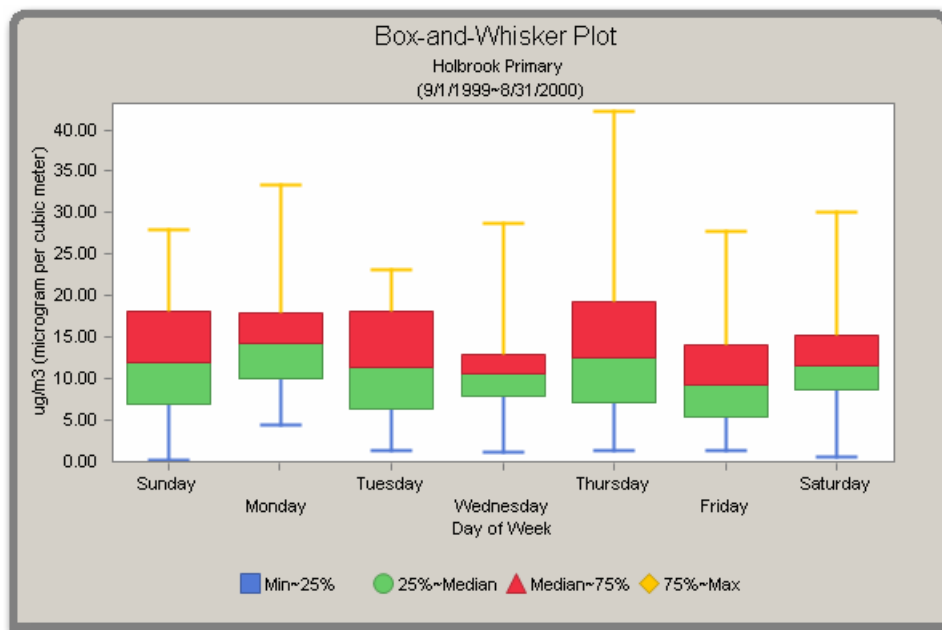


The screenshot shows a web-based interface titled "Data Options". It contains three main sections: "Select Site" with a dropdown menu showing "Lawrenceville Primary"; "Select Parameter" with a dropdown menu showing "PM2.5 Filter-Mass [Filter+denuder] - Duration:24 hour"; and "Calculation Type" with four radio button options: "1-hr Max", "8-hr Average", "24-hr Average" (which is selected), and "Monthly Average".

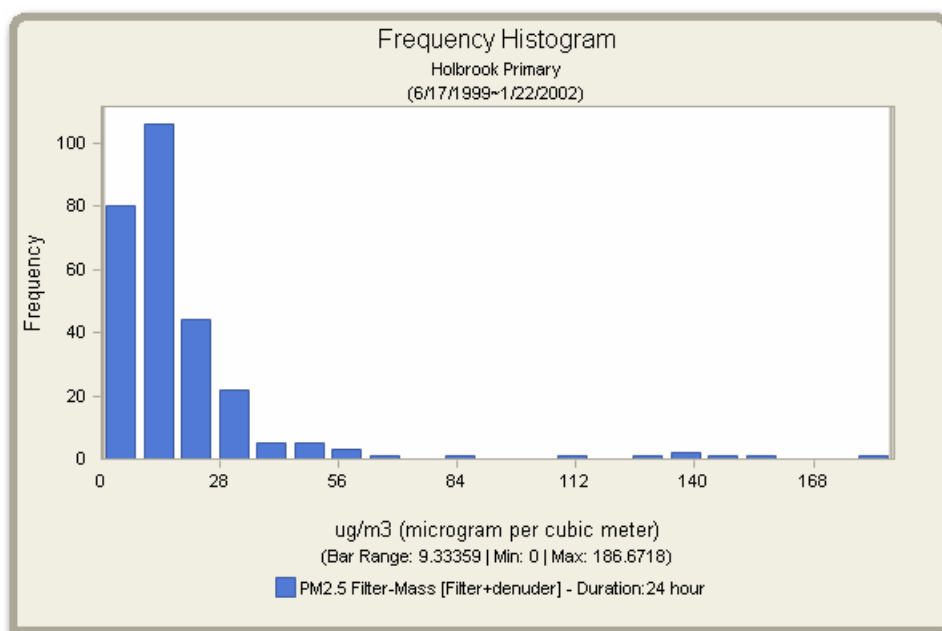
**Figure 9 - Data Calculation Tool**



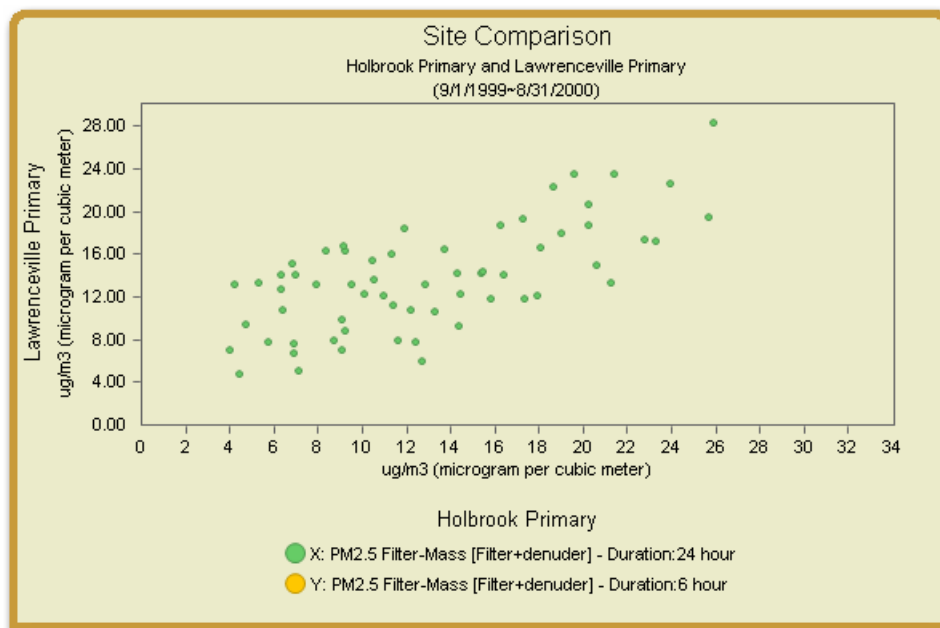
**Figure 10 - Time Series Plots**



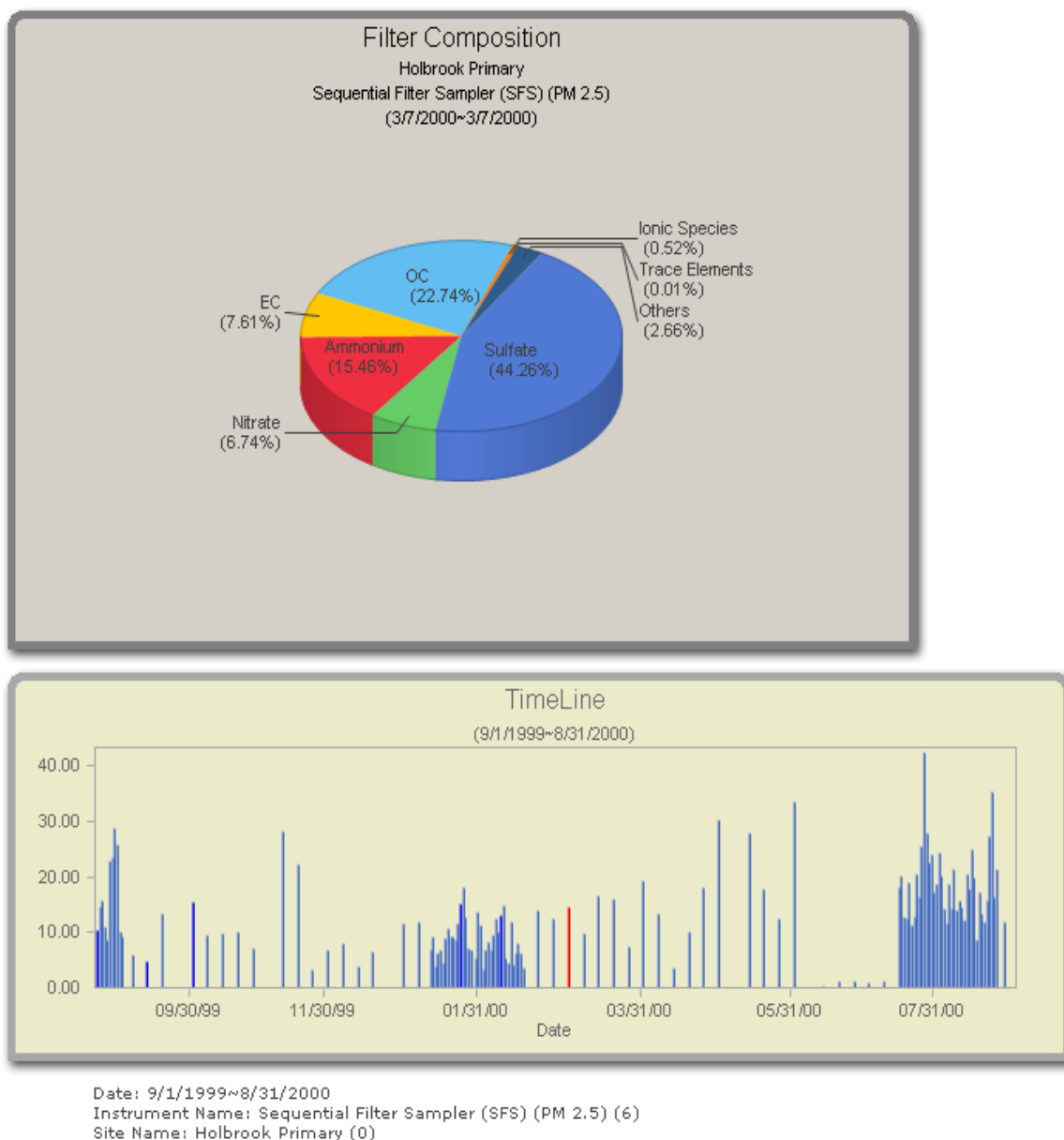
**Figure 11 - Box-Whisker Plots**



**Figure 12 - Frequency Histograms**

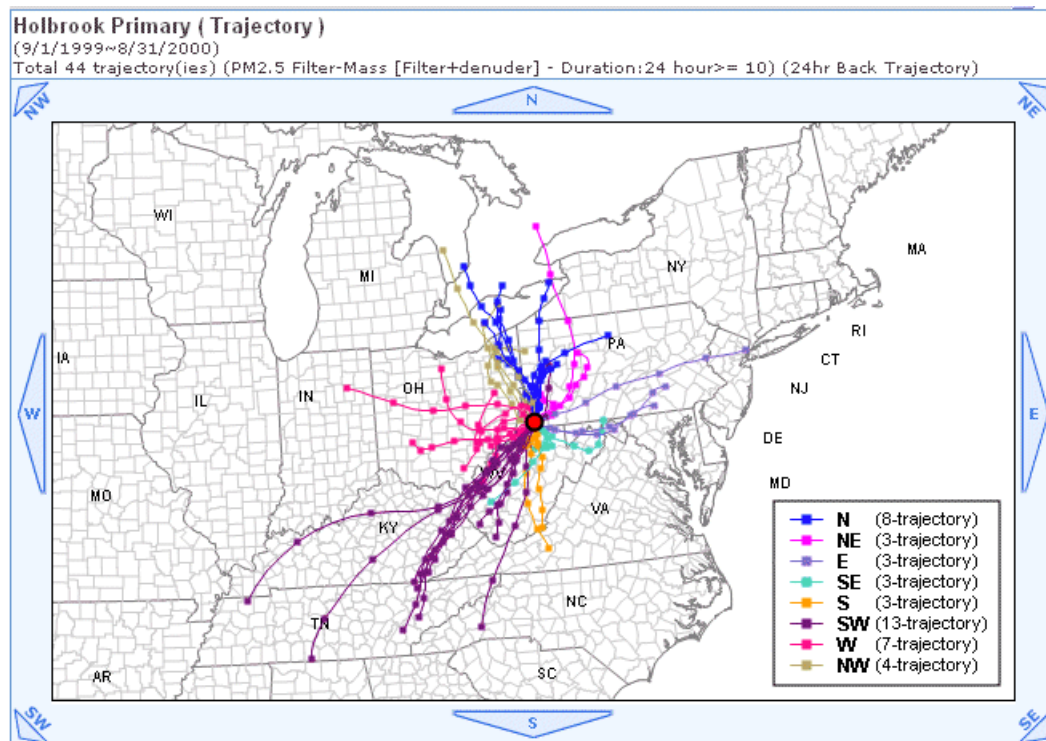


**Figure 13 - Site Comparison Graphs**

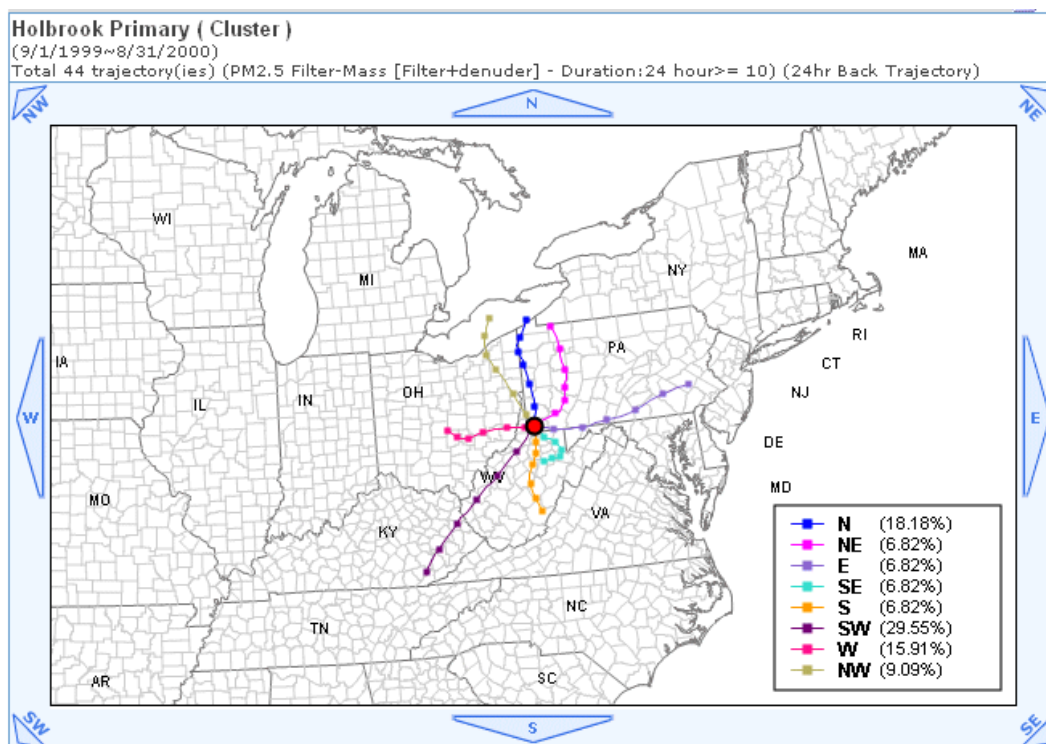


**Figure 14 - Speciation Filter Composition Tool with Timeline Control**

The Trajectory Generator is also new to this toolset. The Trajectory Generator allows users to calculate and display backward trajectories based on trajectory data obtained from NOAA. Users can specify parameter minimum reading levels, elevation, parameter and date range to use in generating the trajectory maps, a sample of which is shown in Figure 15. The user can also perform a cluster analysis on this set of trajectories and display a cluster analysis based on these results. A sample cluster analysis map is shown in Figure 16.

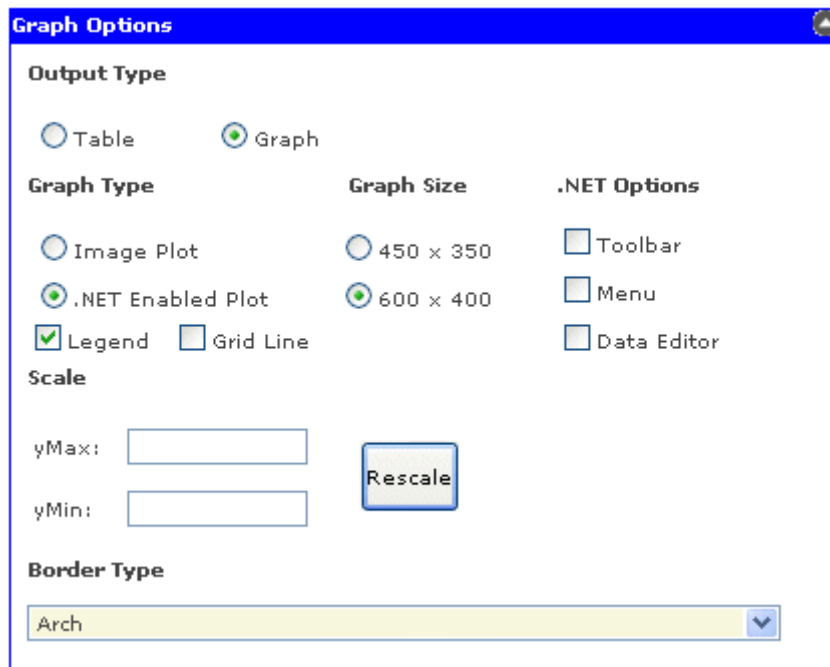


**Figure 15 – Sample Trajectory Map**



**Figure 16 - Sample Cluster Analysis Map**

The graphing tools developed using ChartFX software have been provided with a collapsible panel of controls used to alter the appearance and definable parameters of the graphs called the Graph Options panel, shown in Figure 17. Through this panel, users can specify the type of output desired, table or graphs; the size of the graph, 450 x 305 pixels or 600 x 400 pixels; y-axis minimum and maximum values; and the type of border to use when generating the graph. Users can also choose to generate .NET enabled graphs which offers users a lot of client-side processing controls which do not require round-trip data requests to the server.



**Figure 17 - Graph Options Panel**

The .NET-enabled plots do require that user computers be updated with the latest .NET software platform updates available from Microsoft, but once this requirement has been met, users have access to a set of controls which allow them to quickly change items such as chart or graph type (e.g. switching between scatter and line type charts), color schemes, turn labels and legends on and off and general graph options such as title, axis labels and tick marks.

Users also gain access to a dynamic zooming tool which allows them to quickly zoom in and out of charts to view subsets of data on the chart without having to regenerate a new query or chart. These tools are all available through a Toolbar, which can be displayed by selecting the appropriate checkbox in the Graph Options panel, or by right-clicking on a .NET enabled chart and choosing 'Toolbar' from the context menu. Figure 18 shows the Toolbar displayed with .NET enabled charts.



**Figure 18 - Toolbar for .NET Enabled Charts**

## 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 requires revisiting and reworking some of the original designs, and consequently, will be an ongoing exercise in Phase Two of this project.

A section of the stakeholder website will be dedicated for posting of feedback from beta-testers of this application. Webcast participants are being asked and selected for participation in the beta-testing phases of development and access to the applications described in this report will soon become available for beta testing.

## 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 Table 3:

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

**Table 3 - Phase One Progress Summary**

N/A – on-going activity.

The second project development phase will be completed in the second year of the project, which started in August 2003. 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 table 4.



Task #	Description	Planned completed %	Actual completed %
2.1	Develop Stakeholder Website	100%	85%
2.2	Develop Public Website	100%	65%
2.3	Develop Online Help Feature	100%	15%
2.4	Provide Graphical Representation of Data	100%	75%
2.5	Performance Test	100%	20%

**Table 4 - Phase Two Progress Summary**

## **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  $\beta$ -testing of the developing website and analytical tools, providing 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 five-month, no-cost, time extension from August to December of 2005 will be submitted to DOE-NETL.

## 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](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](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 Pm<sub>10</sub>/Pm<sub>2.5</sub> 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 #1](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R01.pdf)  
[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 #2](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R02.pdf)  
[http://www.netl.doe.gov/coalpower/environment/air\\_q/docs/41476R02.pdf](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R02.pdf)
7. **Robinson P. Khosah, Ph.D., Charles G. Crawford Advanced Technology Systems, Inc.**  
[Semi-Annual Technical Progress Report #3](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R03.pdf)  
[http://www.netl.doe.gov/coalpower/environment/air\\_q/docs/41476R03.pdf](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R03.pdf)
8. **Robinson P. Khosah, Ph.D., Charles G. Crawford Advanced Technology Systems, Inc.**  
[Semi-Annual Technical Progress Report #4](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R04.pdf)  
[http://www.netl.doe.gov/coalpower/environment/air\\_q/docs/41476R04.pdf](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R04.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