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

FINAL

TECHNICAL 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

Final Technical Report

REPORT PERIOD START DATE: August 12, 2002

REPORT PERIOD END DATE: August 11, 2007

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

DATE REPORT ISSUED: December 2007

DOE AWARD NUMBER: DE-FC26-02NT41476

CONTRACTOR: **Advanced Technology Systems, Inc.**
260 Airside Drive
Pittsburgh, PA 15108

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 was conducted in two phases. Phase One included 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 involved the development of a platform for on-line data analysis. Phase Two included 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 technically completed.

TABLE OF CONTENTS

LIST OF GRAPHICAL MATERIALS

LIST OF TABLES

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 were to gather the data from all these monitoring sites into a common database, and to develop analytical tools that would 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 included, 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 was placed on the upper Ohio River valley region, the database was to 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 latter data was not made available during this development effort. The database and analytical tool development effort was also being coordinated, to the extent possible, with similar efforts by U.S. EPA and others to develop relational databases for data collected at their "PM Study Sites". This coordination ensured that the database and analytical tools produced under the DOE-NETL effort would be readily accessible to a wide variety of stakeholders.

The data management system includes 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 provides graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system also provides multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

The project was conducted in two phases. The entire project was divided into ten primary tasks and those have been segmented into two primary phases. Phase One consisted of design and specification tasks related to designing, implementing and populating the primary database that would house the collected data. Phase Two consisted 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 were to gather the data from all these monitoring sites into a common database, and to develop analytical tools that would 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 included, 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 was placed on the upper Ohio River valley region, the database was to 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 latter data was not made available during this development effort. The database and analytical tool development effort was also being coordinated, to the extent possible, with similar efforts by U.S. EPA and others to develop relational databases for data collected at their "PM Study Sites". This coordination ensured that the database and analytical tools produced under the DOE-NETL effort would be readily accessible to a wide variety of stakeholders.

The data management system includes 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 provides graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system also provides multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

The project was conducted in two phases. The entire project was divided into ten primary tasks and those have been segmented into two primary phases. Phase One consisted of design and specification tasks related to designing, implementing and populating the primary database that would house the collected data. Phase Two consisted 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.

A. Project Goals and Objectives

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

The proposed data management system includes a web-based user interface that allows 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 provides graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system also provides 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 was conducted in two phases. The first phase included 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.

Table 1 - Phase One Progress Summary

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%	100%
1.4	Develop Web-based Retrieval System	100%	100%
1.5	Develop Internal QA/QC System	N/A	N/A

The second project development phase is currently being finalized. The tasks involved in this phase included the development of a stakeholder-specific website, a publicly accessible website and an online help feature. This phase also included the development of special analysis tools to provide graphical representation of the data and 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.

Table 2 - Phase Two Progress Summary

Task #	Description	Planned % completed	Actual % completed
2.1	Develop Stakeholder Website	100%	100%
2.2	Develop Public Website	100%	100%
2.3	Develop Online Help Feature	100%	90%
2.4	Provide Graphical Representation of Data	100%	100%
2.5	Performance Test	100%	99%

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

As stated in previous reports, data has already been populated by using a series of conversion scripts and data processing utilities to *pre-process*, or re-format the supplied data source files into application-specific formats, and a special application, the *PM Data Imports Utility*, was developed and used to create specialized data mappings and import a large amount of data received from Desert Research Institute (DRI). These specialized data mappings involve the mapping of DRI parameters to NARSTO parameters and sometimes involve creating new parameter records for items such as ‘Strong Acidity’ which equates an acidity level equivalent to levels of H_2SO_4, or Sulfuric Acid.

The project team has also imported all compatible data collected through the Pittsburgh Air Quality Study (PAQS), provided by researchers at Carnegie Mellon University (CMU). This data included a total of sixty-six (66) source files, containing data for ninety-seven (97) different parameters, for a total of one million, three hundred and sixteen thousand, eight hundred and ninety-four (1,316,894) new data records. All of the PAQS observation records, with the exception of the *Single Particle* data, have been imported into the project database.

Table 3 – UORVP Data Inventory details the quantity of data records obtained for each collection site within the UORVP data collection network.

Data Collection Site	Record Count
PITTSBURGH SUPERSITE	1,306,134
LAWRENCEVILLE PRIMARY	258,641
HOLBROOK PRIMARY	134,066
MORGANTOWN SATELITE	576

Table 3 - UORVP Data Inventory

TEOM data and data obtained from the Environmental Protection Agency (EPA) for data collection sites in Pennsylvania, Ohio, West Virginia and Kentucky have also been imported into the database. 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 is one remaining primary dataset not present in the database at this time. This is data collected by the Steubenville Comprehensive Air Monitoring Program (SCAMP) sites, which has not yet been received by the project team. Appendix I – Comprehensive Data Record Inventory contains a complete listing of all data records currently in the master database.

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, development efforts have primarily been focused on extending and improving the existing web-based data retrieval controls, discussed in detail in previous reports.

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. Previous semi-annual technical reports detail

the automation techniques used to verify data integrity during the database population processes.

In addition to verifying the data, the internal QA/QC system also allows for testing the website's integrity and functionality. Part of this scrutiny has resulted in the fixing of erroneous contact information and data inventory errors on the website.

The QA/QC process is an on-going exercise.

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 toolset 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)
- Back trajectory analysis using interactive mapping tools from Google

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.

As part of the development process, the project team continues to engage stakeholders to provide critical feedback so the development team can create *meaningful* and *useful* analysis tools. As stated in previous reports, a series of web casts have been conducted,

with more planned for the near future, and the capabilities and potential for this web-based application are presented and discussed at these forums.

Additionally, the project team has also taken the project *on-the-road* and made presentations about the project, including live demonstrations, to researchers attending several relevant conferences and seminars. The latest conference attended by the project team members was the NOAA-EPA Golden Jubilee conference held in Raleigh, North Carolina from September 19, through September 21, 2005. The project team continues to spread awareness of the project and to solicit input for improvements to both the data retrieval and data analysis tools.

The final product contains a redesigned overall look and feel of the stakeholder website, using updated graphics and a new and easier to use site development template. Future efforts will be able to build new sections of the site in a more streamlined and efficient manner as a result of this work. Figure 1 – Visible Changes to Stakeholder Application outlines the visible changes to the application.

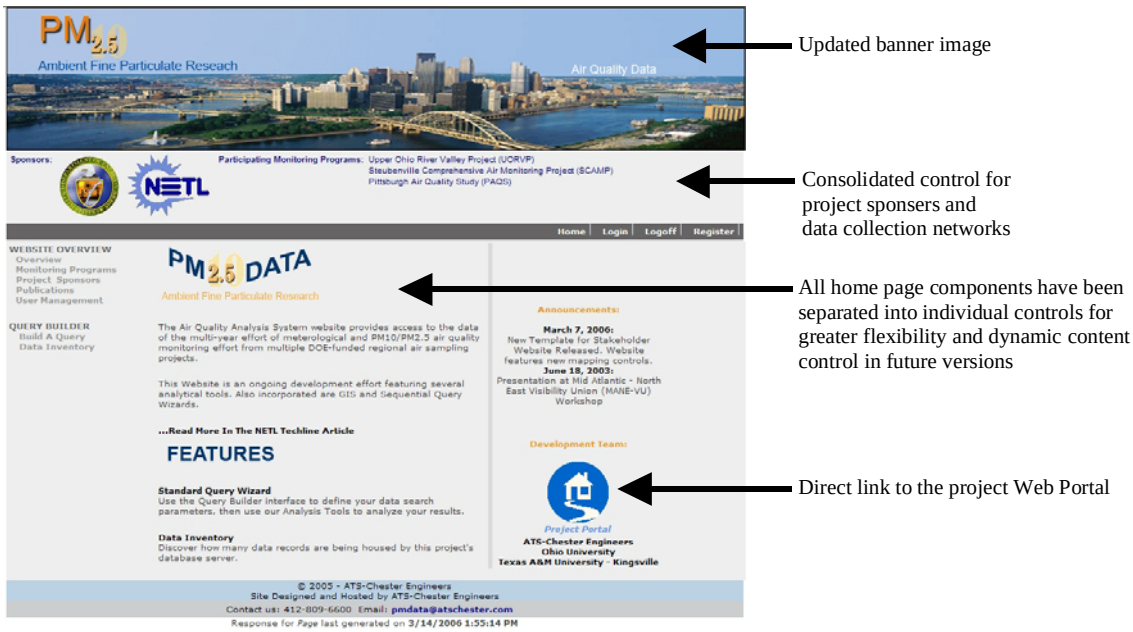


Figure 1 - Visible Changes to Stakeholder Application

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.

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.

Preliminary versions of the public website were 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.

Previous reports detailed several analysis tools used to graphically represent selected datasets:

1. data calculation tool, shown in Figure 2
2. time series plots, shown in Figure 3
3. box-whisker plots, shown in Figure 4
4. frequency histograms, shown in Figure 5
5. site comparison graphs using scatter plots, shown in Figure 6
6. filter composition tool, shown in Figure 7

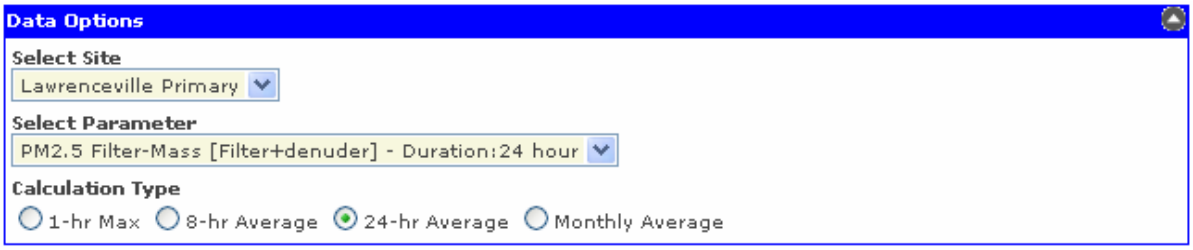


Figure 2 - Data Calculation Tool

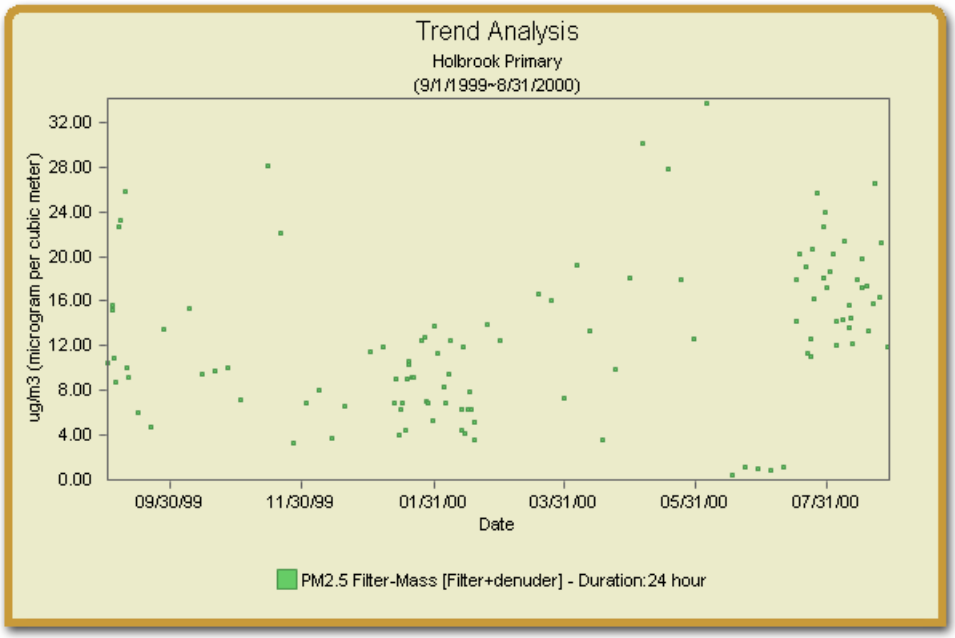


Figure 3 - Time Series Plots

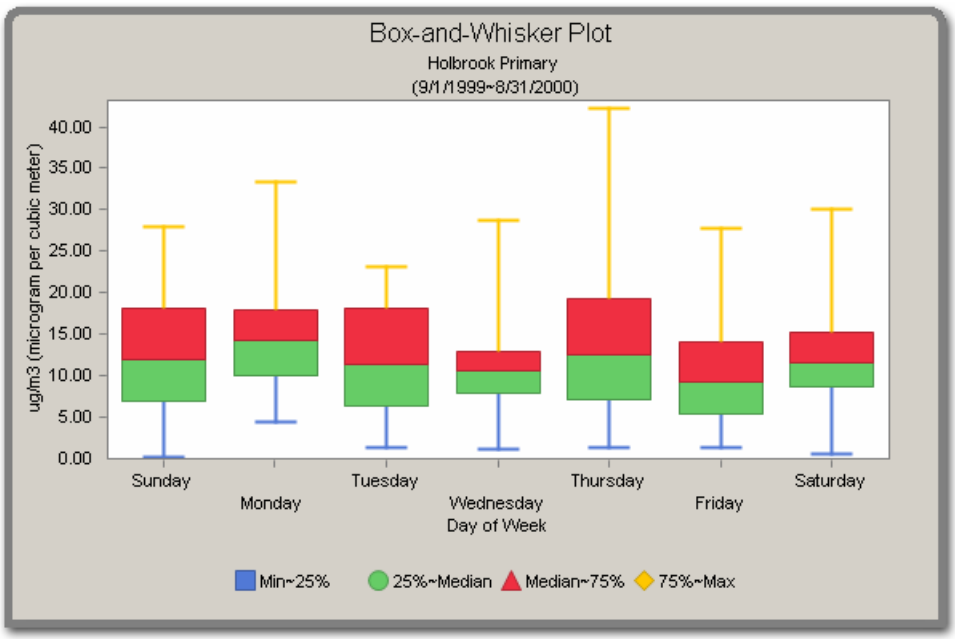


Figure 4 - Box-Whisker Plots

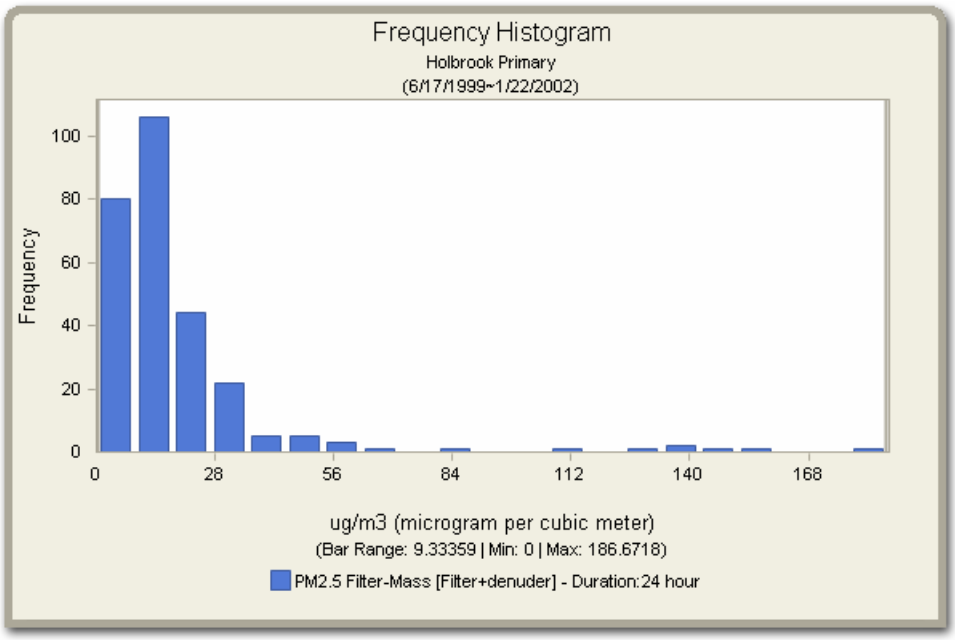


Figure 5 - Frequency Histograms

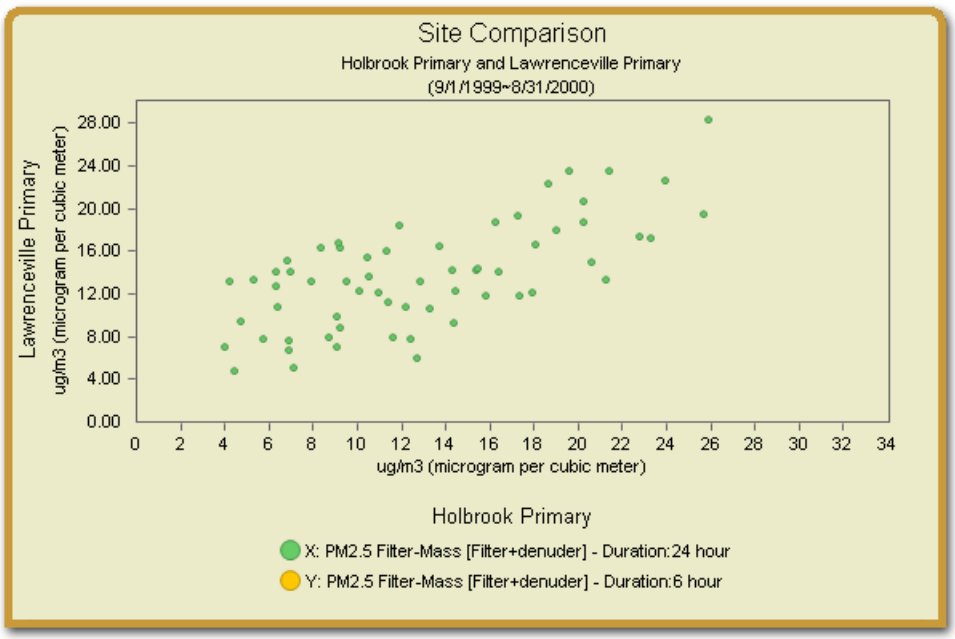


Figure 6 - Site Comparison Graphs

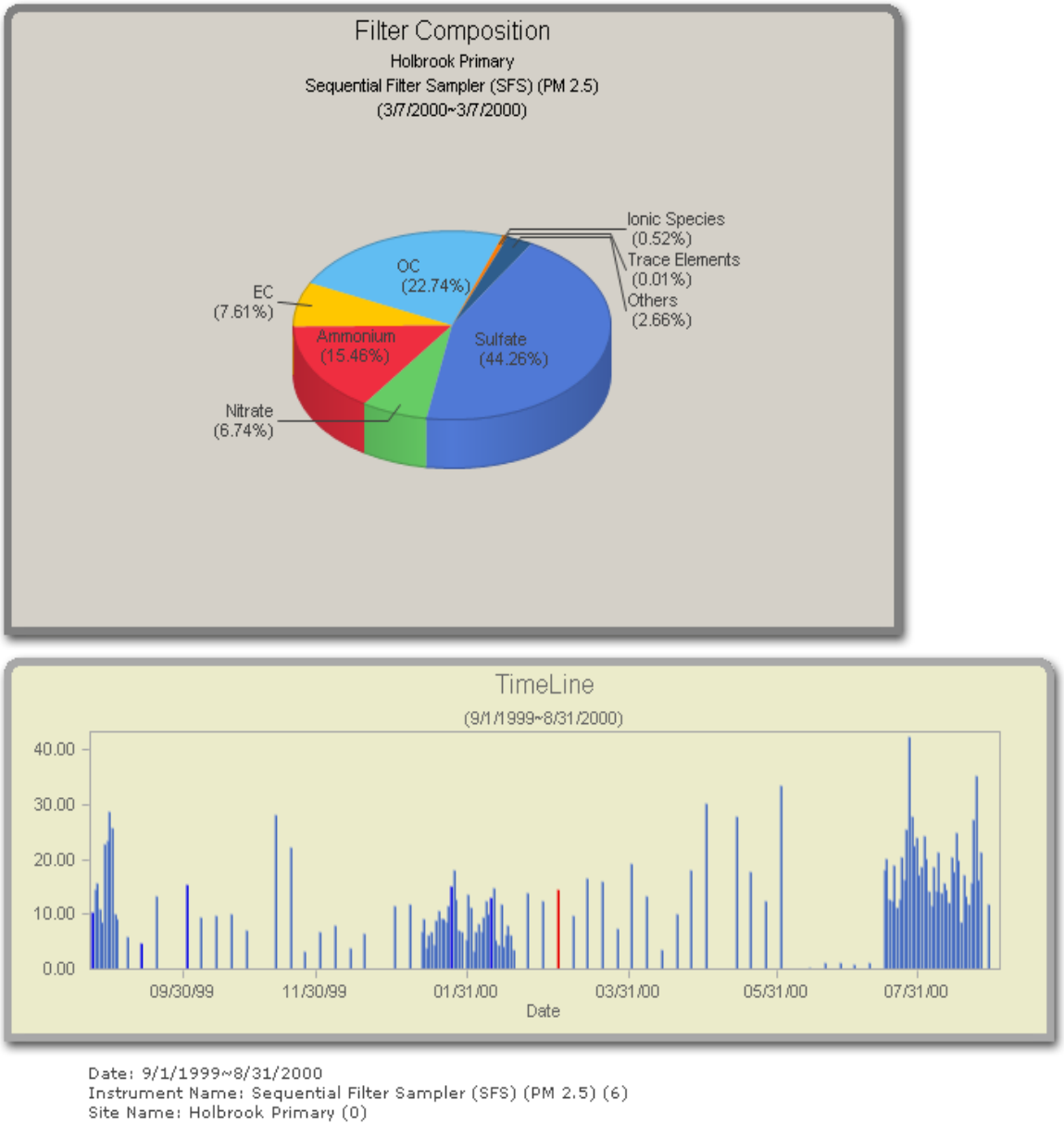


Figure 7 - Speciation Filter Composition Tool with Timeline Control

Work continues to improve the trajectory and cluster analysis tools, shown in Figures 7 and 9 and mentioned in previous reports. These tools require the use of pre-compiled trajectory data for each data collection site. The query operations used to develop a user-requested trajectory or cluster analysis have been redesigned to accommodate more users and more frequent usage and the Java-based map used in previous versions of the application with a new interactive Google Maps control.

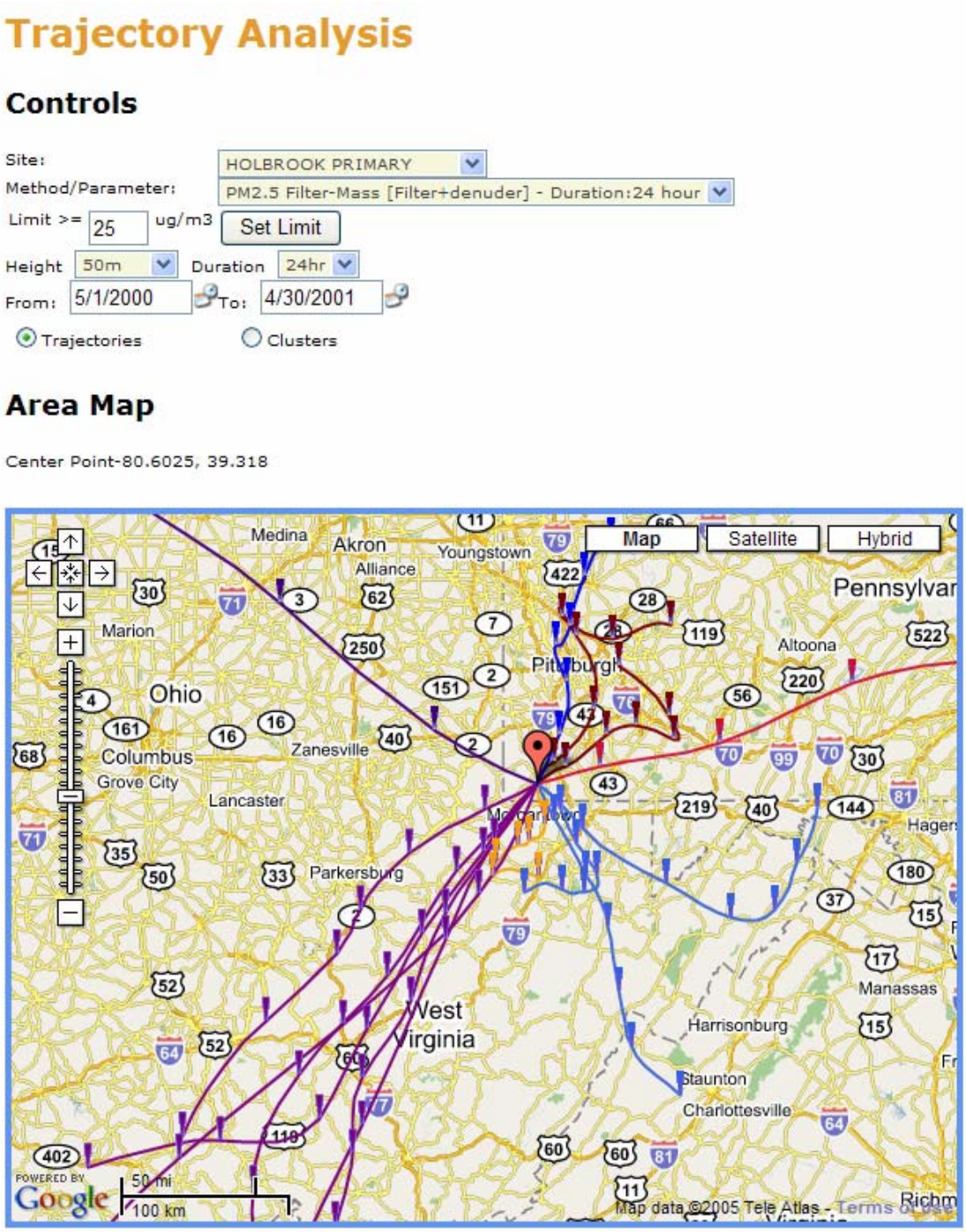


Figure 8 – Sample Trajectory Map

Our project team has also added more graphical tools to the stakeholder website such as the interactive map with data collection site metadata browser using a Google Maps interface.

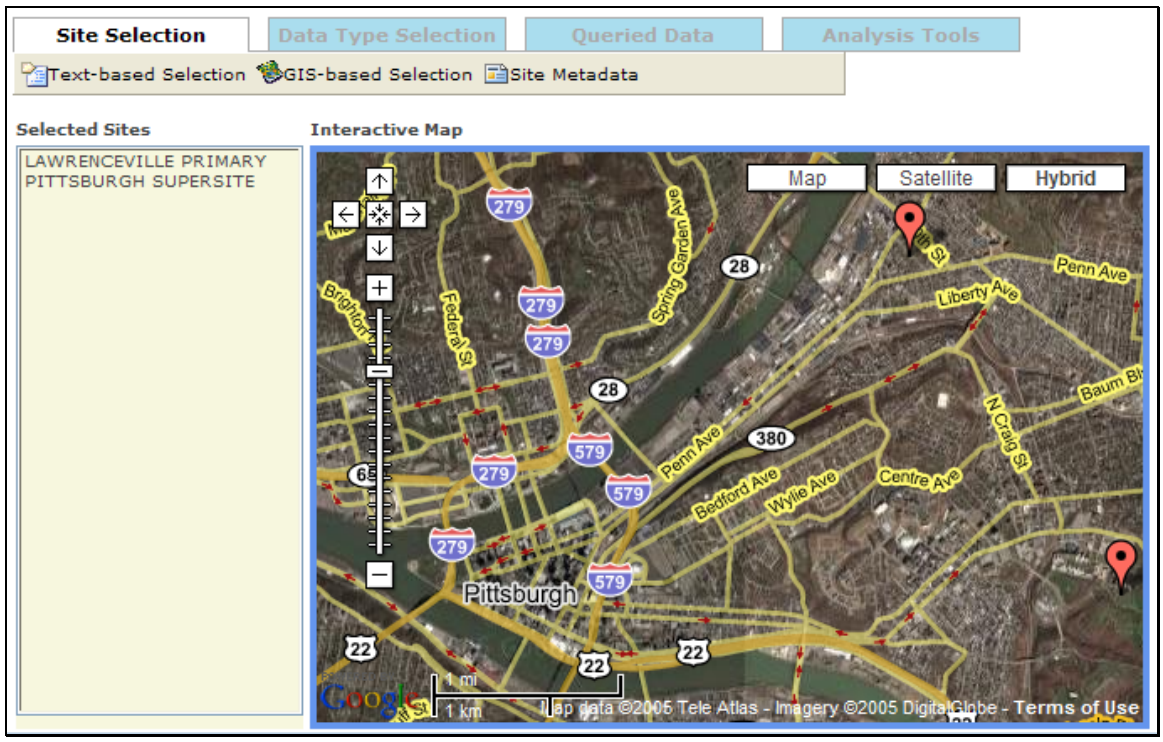


Figure 9 - Site Metadata Map

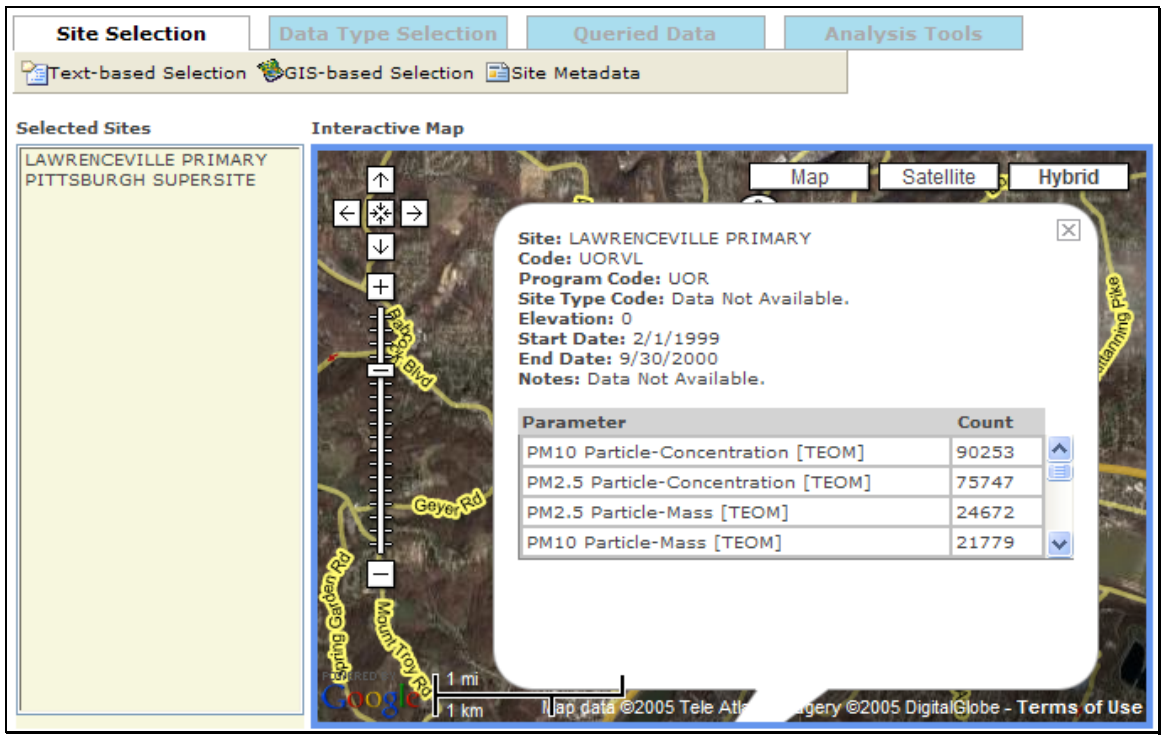


Figure 10 - Site Metadata Map with Info Window

Future efforts will continue to improve on the tools developed thus far and the project team can also work on new tools such as an improved trajectory and cluster analysis tool, depicted in Figure 7 – Sample Trajectory Map, which will use mapping overlays to depict trajectory and cluster vector graphics on an interactive Google Maps control.

Task 2.5 - Performance Testing

ATS, in coordination with all members of the external stakeholder group, conducted detailed testing program for the resulting application to verify the functionality and proper execution of all portions of this application. This testing program provided for interactive user feedback, discussion forums and periodic email notifications and announcements. This testing program helped insure that the intended objectives of this project were met or exceeded. This effort required revisiting and reworking some of the original designs, and consequently, was an ongoing exercise in Phase Two of this project.

Web cast participants were asked to participate in the beta-testing phases of development and links to the applications have been provided to researchers upon request. A Project Web Portal was implemented to facilitate this information exchange between team members and beta testers of the applications.

C. Project Web Portal

ATS-Chester Engineers has provided the project team with a Project Web Portal to provide the team members, client and stakeholders with a set of tools for monitoring the project tasks and for transferring files and data between registered portal users. The portal also provides the client with direct access to the team members, project discussions and status updates.

The Project Web Portal facilitates and fosters better teamwork, communication and collaboration, and file/data sharing; all of which provide DOE-NETL with a better set of products and services.

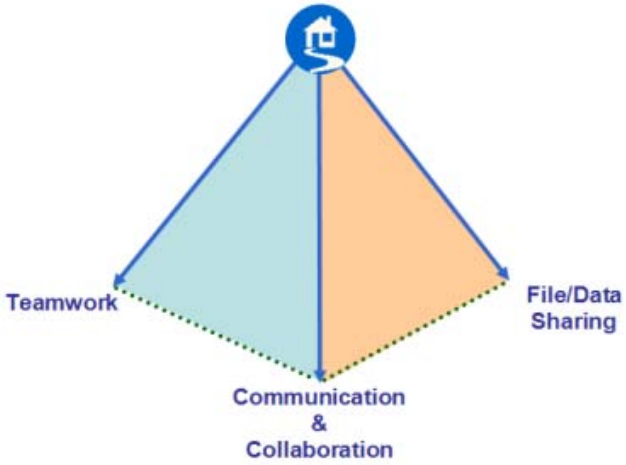


Figure 11 - Project Web Portal Cornerstones

Some of the key features of the portal include a project calendar, shown in Figure 12 – Project Calendar, for scheduling important project-related events; a task scheduler, shown in Figure 13 – Task Scheduler, for monitoring project tasks and schedules; a meeting center, shown in Figure 14 – Meeting Center, to track Action Items and meeting minutes; file and data exchange tools, including an FTP link to allow for file transfers greater than 4MB in size; a project mapping utility for use in testing new mapping controls before they are integrated into the project web sites; as well as miscellaneous modules such as a contacts list, project team members pictures with project role descriptions and a hyperlinks module used as a bookmarks listing for project-related web links.

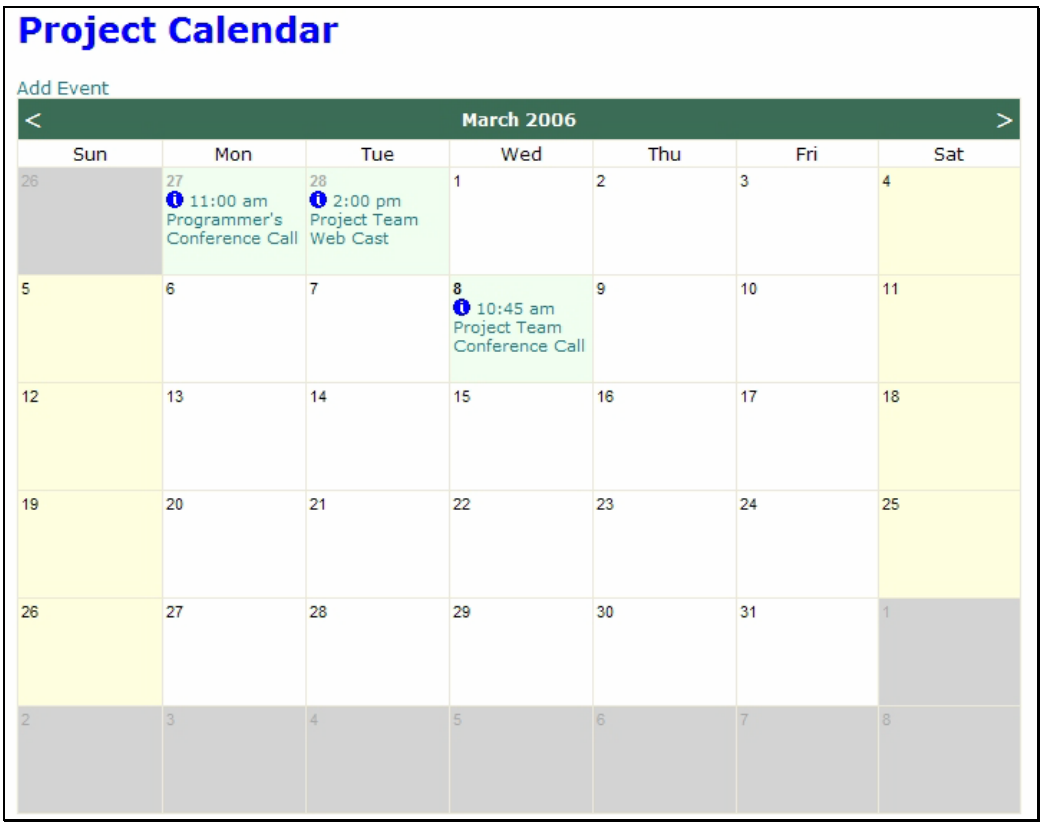
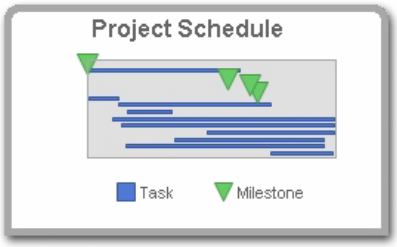


Figure 12 - Project Calendar

Project Tasks



Project tasks are organized by Task Number.

Clicking on the Task Number or Status values opens the Task Editor window if your login credentials permit Task edits. Clicking on the miniature Gantt chart opens a full-sized Gantt chart with accompaning task and milestone tables.

Users can add new tasks by clicking on the New Task button, or delete existing tasks by clicking on the trashcan icon in the far left column of the corresponding row in the task table.

	Task No.	Title	Duration (Days)	Start Date	Target Date	Status
	0s	Project Kickoff Meeting	1	09-25-2002	09-25-2002	Complete
	1.1	Data Inventory and Benchmarking	617	10-01-2002	02-11-2005	Complete
	1.1a	2004 Ohio Air Quality and Coal Research Symposium Presentation	2	12-02-2004	12-06-2004	Complete
	1.1b	CREST-RESSACA Conference -San Antonio, TX	4	04-12-2005	04-15-2005	Complete
	1.1c	IAGLR Conference 2005	5	05-23-2005	05-27-2005	Complete
	1.2	Develop Data Management System Architecture	121	10-01-2002	03-18-2003	Complete
	1.3	Population of Database	622	03-19-2003	08-11-2005	Complete
	1.4	Develop Web-based Retrieval System	181	05-12-2003	01-19-2004	Complete
	1.5	Develop Internal QA/QC System	901	02-17-2003	08-11-2006	Open
	2.1	Develop Stakeholder Website	866	04-07-2003	08-11-2006	Open
	2.2	Develop Public Website	513	08-12-2004	08-11-2006	Open
	2.3	Develop Online Help Feature	605	02-09-2004	06-15-2006	Open
	2.4	Provide Graphical Representation of Data	807	05-01-2003	06-15-2006	Open
	2.5	Performance Test	250	08-11-2005	08-01-2006	Open

New Task

Figure 13 - Task Scheduler

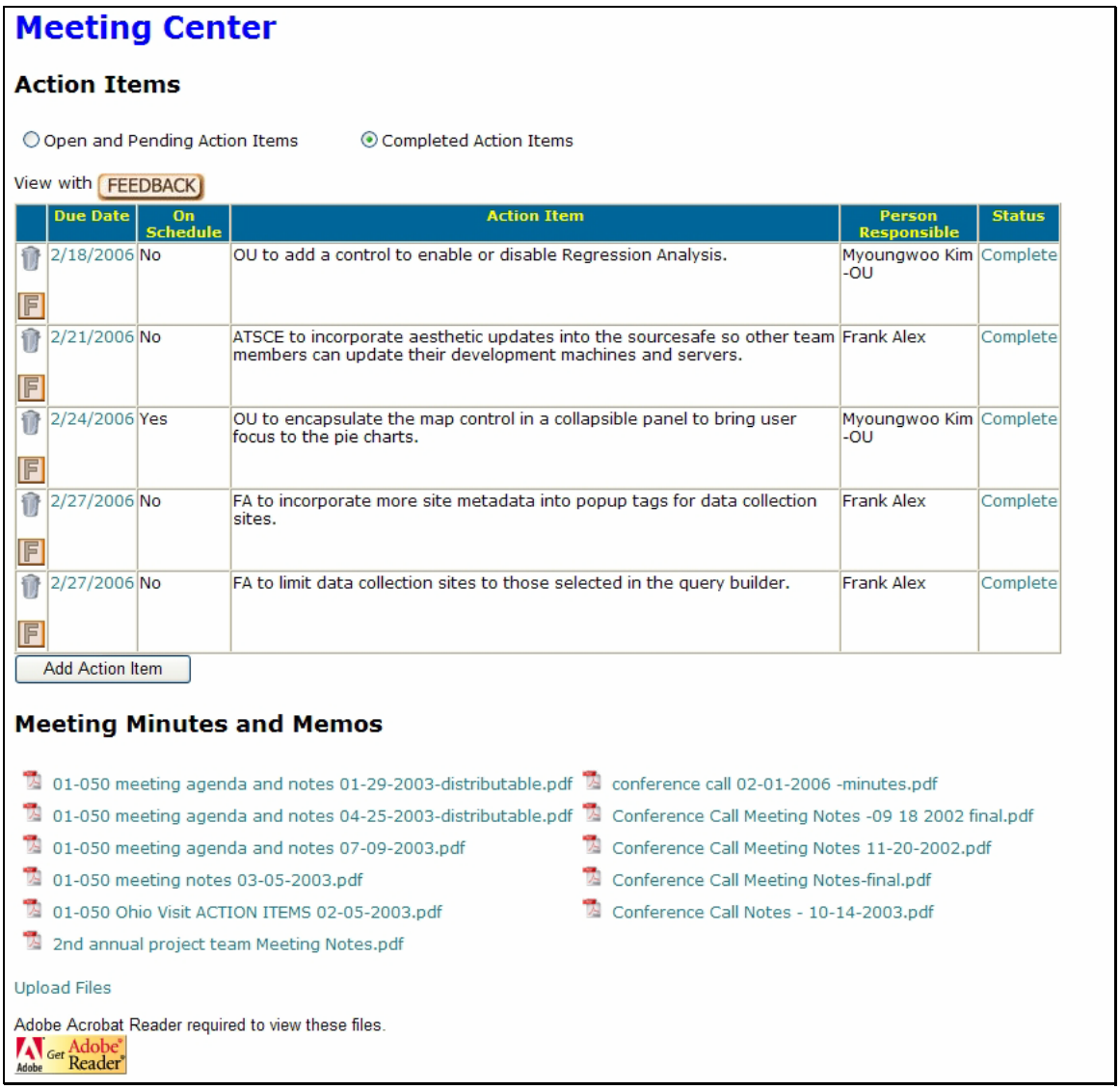


Figure 14 - Meeting Center

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 4:

Table 4 - Phase One Progress Summary

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%	100%
1.4	Develop Web-based Retrieval System	100%	100%
1.5	Develop Internal QA/QC System	N/A	N/A

The second project development phase is currently being finalized. 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 a series of performance tests designed to provide the best possible data management solution.

A progress summary for Phase Two is shown in Table 5.

Table 5 - Phase Two Progress Summary

Task #	Description	Planned % completed	Actual % completed
2.1	Develop Stakeholder Website	100%	100%
2.2	Develop Public Website	100%	100%
2.3	Develop Online Help Feature	100%	90%
2.4	Provide Graphical Representation of Data	100%	100%
2.5	Performance Test	100%	99%

IV. CONCLUSION

The development efforts on this project proceeded as expected given the size, complexity and the different stakeholders involved. Some target milestones were not been met in a timely manner primarily due to delays in acquiring input data from third party sources. This was especially so with the data inventory task, where data reformatting issues were also encountered. Delays were also 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 was believed that the ultimate success of this project would require a high degree of stakeholder confidence and subsequent participation in the website development process, and that such participation would be greatly enhanced if stakeholders were presented with a relatively “polished” product at the outset. Therefore, the extra programming effort was 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 was also actively involved in an on-going evaluation and β-testing of the developing website and analytical tools, providing critical feedback that was instrumental in modifications that have made the application more user-friendly and the navigation much more dynamic.

The tool was ready for delivery to NETL but Chester Engineers encountered a catastrophic failure of the website server in May 2007. This crash required the rebuilding of the server and the application tools from scratch and we are just now close to having a fully functional site. The redundancy we had employed in locating the website came in handy as the application at Ohio University remained uninterrupted and accessible.

The hurdles encountered, however, have not been insurmountable. Since the level of effort associated with the slowed tasks is still the same, the costs to complete this project have not been impacted either negatively or positively. The project is currently proceeding on a no-cost time extension, with the extra effort being provided through the in-kind component from project team, with no financial requirement or commitment from DOE-NETL.

The server, with a fully functional database website to NETL, will be delivered sometime in March, 2008.

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

2. **Professors Spyros Pandis, Cliff Davidson, and Allen Robinson, Carnegie Mellon University**
   EPA Supersite - Carnegie Mellon University
   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

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 $Pm_{10}/Pm_{2.5}$ Air Quality Study: Objectives and Associated Data Analysis and Modeling Approaches
   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

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

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

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

9. **Robinson P. Khosah, Ph.D., Charles G. Crawford Advanced Technology Systems, Inc.**
   Semi-Annual Technical Progress Report #5
   http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R05.pdf

10. **Robinson P. Khosah, Ph.D., Charles G. Crawford Advanced Technology Systems, Inc.**
    Semi-Annual Technical Progress Report #6
    http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R06.pdf

BIBLIOGRAPHY

Not applicable.

VI. 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.

Term	Definition
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.

Term	Definition
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

Appendix I – Comprehensive Data Record Inventory

PITTSBURGH SUPERSITE	1,306,134
Description	**Record Count**
MET Air Pressure	68,075
MET Ambient Temperature	68,075
MET Relative Humidity	67,991
MET Downwelling UV Radiation	67,991
MET Broadband Downwelling Solar Hemispheric Radiation	67,991
MET Sigma Theta-standard deviation of wind direction	67,991
MET Current Precipitation [Rain Gauge]	67,991
MET Unit Vecto Mean Wind Direction [Wind direction vane]	67,991
MET Horizontal Scalar Max Wind Speed [Anemometer--cup]	67,991
MET Horizontal Scalar Mean Wind Speed [Anemometer--cup]	67,991
PM2.5 Ionic-SO4 [Automated particulate sulfate monitor]	61,488
GAS SO2 [Pulsed fluorescence]	61,488
GAS CO1 [Infrared absorption]	61,488
GAS Nitric Oxide [Chemiluminescence]	61,488
PM2.5 Ionic-NO3 [Automated particulate nitrate monitor]	61,488
PM2.5 Particle-Mass [TEOM]	61,488
GAS NOX [Chemiluminescence]	61,488
GAS Ozone [Ultraviolet absorption]	61,488
GAS Total Peroxide [Fluorescence--other]	9,740
PM2.5 Light Scatter Co-efficient [Optical scattering--integrating nephelometer]	8,384
MET Instrument Temperature	8,384
MET Ambient Temperature	8,384
PM2.5 Ionic-CL [Continuous gas and particle speciation monitor]	5,508
PM2.5 Ionic-NA [Continuous gas and particle speciation monitor]	5,508
PM2.5 Ionic-NO2 [Continuous gas and particle speciation monitor]	5,508
PM2.5 Ionic-NO3 [Continuous gas and particle speciation monitor]	5,508
PM2.5 Ionic-NH4 [Continuous gas and particle speciation monitor]	5,508
PM2.5 Ionic-SO4 [Continuous gas and particle speciation monitor]	5,508
MOUDI Mass [Impactor]	3,555
PM2.5 Element-C [Continuous]	2,561
PM2.5 Organic-C [Continuous]	2,561
VOC o-Xylene [Continuous]	1,297
VOC Toluene [Continuous]	1,297
VOC Tetrachloroethene [Continuous]	1,297
VOC t-2-Butene [Continuous]	1,297
VOC N-Butane [Continuous]	1,297
VOC c-2-Butene [Continuos]	1,297
VOC Chloroform [Continuous]	1,297

VOC m-Xylene [Continuous]	1,297
VOC Benzene [Continuous]	1,297
VOC 1-Propene [Continuous]	1,297
VOC N-Pentane [Continuous]	1,297
VOC 2-methoxy-2-methyl-Propane [Continuous]	1,297
VOC 1-Pentene [Continous]	1,297
VOC Pentane, 2-Methyl- and Pentane, 3-Methyl- [Continuous]	1,297
VOC 2-Methyl-1-Butene [Continuos]	1,297
VOC N-Hexane [Continuous]	1,297
VOC Ethanol [Continuous]	1,297
VOC Methanol [Continuous]	1,297
VOC t-2-Pentene [Continuous]	1,297
VOC Acetone [Continuous]	1,297
VOC 1-Propyne [Continuous]	1,297
VOC 2-Propanol [Continuous]	1,297
VOC Cyclopentane [Continuos]	1,297
VOC Isopentane [Continuous]	1,297
VOC 3-Methyl-1-Butene [Continuous]	1,297
VOC Isobutane [Continuous]	1,297
VOC Methyl Ethyl Ketone [Continuous]	1,297
VOC p-Xylene [Continuous]	1,297
VOC Ethylbenzene [Continuous]	1,297
VOC 1-Butene [Continuous]	1,297
VOC Propane [Continuous]	1,297
GAS Hydrogen Peroxide [Fluorescence--other]	1,185
GAS Organic Peroxide [Fluorescence--other]	1,185
PM2.5 Organic-C [Filter - multiple]	1,032
VOC Acetaldehyde [Continuous]	662
VOC 1-Propene, 2-Methyl-	662
VOC Dimethyl Sulfide [Continuous]	635
VOC 3-Methylfuran [Continuous]	635
VOC Butanal [Continuous]	635
VOC Cyclopentene [Continuous]	635
VOC Dichloromethane [Continuous]	635
VOC Acetonitrile [Continuous]	635
VOC Pentanal [Continuous]	635
VOC Acetaldehyde [Continuous]	635
VOC Methacrolein [Continuous]	635
VOC Methyl Vinyl Ketone [Continuous]	635
VOC Alpha-Pinene [Continuous]	635
VOC Isoprene [Continuous]	635
PM2.5 Organic-C [Single Filter]	516
PM2.5 Element-C [Single Filter]	516

PM2.5 Ionic-NH4 [Filter - multiple]	504
PM2.5 Ionic-NO3 [Filter - multiple]	504
PM2.5 Ionic-SO4 [Filter - multiple]	504
GAS Nitric Acid [Filter - multiple]	504
GAS Ammonia [Filter - multiple]	504
PM2.5 Mass [FRM]	404
PM10 Filter-Mass [Single Filter]	377
PM2.5 Filter-Mass [Single Filter]	377
PM2.5 Element-C [Filter+denuder]	84
PM2.5 Organic-C [Filter+denuder]	84
CLO Ionic-CL [Cloud/Fog water--active]	8
CLO Ionic-SO4 [Cloud/Fog water--active]	8
CLO Ionic-K [Cloud/Fog water--active]	8
CLO Ionic-CA [Cloud/Fog water--active]	8
CLO Fog Water-Volume [Cloud/Fog water--active]	8
CLO Organic-C [Cloud/Fog water--active]	8
CLO Ionic-NO3 [Cloud/Fog water--active]	8
CLO Ionic-NH4 [Cloud/fog water--active]	8
CLO Fog Water-PH [Cloud/Fog water--active]	8
CLO Total Dissolved Organic-C [Cloud/Fog water--active]	8
CLO Ionic-NA [Cloud/fog water--active]	8
CLO Ionic-MG [Cloud/Fog water--active]	8
CLO Ionic-NITRITE [Cloud/Fog water--active]	8

LAWRENCEVILLE PRIMARY	**258,641**
Description	**Record Count**
PM10 Particle-Concentration [TEOM]	90,253
PM2.5 Particle-Concentration [TEOM]	75,747
PM2.5 Particle-Mass [TEOM]	24,672
PM10 Particle-Mass [TEOM]	21,779
PM2.5 Flow-Vol [Not assigned]	3,180
PM2.5 Element-AG [Filter+denuder]	2,389
PM2.5 Filter-Mass [Filter+denuder]	1,831
PM10 Flow-Vol [Other]	1,730
PM2.5 Element-AS [Filter+denuder]	1,347
PM2.5 Element-AL [Filter+denuder]	1,347
PM2.5 Ionic-CHL [Filter+denuder]	1,329
TSP Flow-Vol [Not assigned]	1,080
PM2.5 Element-AU [Filter+denuder]	946
PM2.5 Element-IN [Filter+denuder]	946
PM2.5 Element-HG [Filter+denuder]	898
PM2.5 Element-CR [Filter+denuder]	898

PM2.5 Element-BA [Filter+denuder]	898
PM2.5 Element-CA [Filter+denuder]	898
PM2.5 Element-BR [Filter+denuder]	898
PM2.5 Element-CU [Filter+denuder]	898
PM2.5 Element-CD [Filter+denuder]	898
PM2.5 Element-FE [Filter+denuder]	898
PM2.5 Element-CO [Filter+denuder]	898
PM2.5 Element-GA [Filter+denuder]	898
PM2.5 Element-C [Filter+denuder]	886
PM10 Filter-Mass [Filter+denuder]	758
TSP Element-NO3 [Filter--multiple]	524
TSP Element-NO3 [Denuder--cylinder+filter(s)]	524
PM2.5 Element-RB [Filter+denuder]	449
PM2.5 Element-SN [Filter+denuder]	449
PM2.5 Element-MO [Filter+denuder]	449
PM2.5 Element-V [Filter+denuder]	449
PM2.5 Element-TL [Filter+denuder]	449
PM2.5 Element-MG [Filter+denuder]	449
PM2.5 Element-ZR [Filter+denuder]	449
PM2.5 Element-SI [Filter+denuder]	449
PM2.5 Element-U [Filter+denuder]	449
PM2.5 Element-TI [Filter+denuder]	449
PM2.5 Element-LA [Filter+denuder]	449
PM2.5 Element-ZN [Filter+denuder]	449
PM2.5 Element-SB [Filter+denuder]	449
PM2.5 Element-S [Filter+denuder]	449
PM2.5 Element-PB [Filter+denuder]	449
PM2.5 Element-NA [Filter+denuder]	449
PM2.5 Element-P [Filter+denuder]	449
PM2.5 Element-SE [Filter+denuder]	449
PM2.5 Element-MN [Filter+denuder]	449
PM2.5 Element-PD [Filter+denuder]	449
PM2.5 Element-Y [Filter+denuder]	449
PM2.5 Element-NI [Filter+denuder]	449
PM2.5 Element-K [Filter+denuder]	449
PM2.5 Element-SR [Filter+denuder]	449
PM2.5 Organic-TOC [Filter+denuder]	443
PM2.5 Element-SO4 [Filter+denuder]	443
PM2.5 Ionic-NA [Filter+denuder]	443
PM2.5 Organic-C [Filter+denuder]	443
PM2.5 Element-NO3 [Filter+denuder]	443
PM2.5 Element-NH4 [Filter+denuder]	443
PM2.5 Ionic-K [Filter+denuder]	443

PM2.5 Acid-H2SO4 [Filter+denuder]	401
PM2.5 Gas-TOC [Filter+denuder]	401
PM2.5 Gas-C [Filter+denuder]	401
TSP Element-NH4 [Filter--multiple]	262
PM10 Element-AG [Filter+denuder]	159
PM10 Ionic-CHL [Filter+denuder]	123
PM10 Element-AS [Filter+denuder]	123
PM10 Element-AL [Filter+denuder]	123
PM2.5 Element-CL [Filter+denuder]	96
PM10 Element-FE [Filter+denuder]	82
PM10 Element-BR [Filter+denuder]	82
PM10 Element-CD [Filter+denuder]	82
PM10 Element-CR [Filter+denuder]	82
PM10 Element-IN [Filter+denuder]	82
PM10 Element-CU [Filter+denuder]	82
PM10 Element-CO [Filter+denuder]	82
PM10 Element-GA [Filter+denuder]	82
PM10 Element-AU [Filter+denuder]	82
PM10 Element-CA [Filter+denuder]	82
PM10 Element-HG [Filter+denuder]	82
PM10 Element-BA [Filter+denuder]	82
PM10 Element-C [Filter+denuder]	82
PM2.5 Acid-SAEC [Filter+denuder]	49
PM10 Element-NI [Filter+denuder]	41
PM10 Element-K [Filter+denuder]	41
PM10 Element-NO3 [Filter+denuder]	41
PM10 Element-ZN [Filter+denuder]	41
PM10 Element-LA [Filter+denuder]	41
PM10 Element-SB [Filter+denuder]	41
PM10 Element-PD [Filter+denuder]	41
PM10 Ionic-NA [Filter+denuder]	41
PM10 Element-MN [Filter+denuder]	41
PM10 Element-U [Filter+denuder]	41
PM10 Element-TI [Filter+denuder]	41
PM10 Element-SI [Filter+denuder]	41
PM10 Acid-H2SO4 [Filter+denuder]	41
PM10 Element-SO4 [Filter+denuder]	41
PM10 Element-Y [Filter+denuder]	41
PM10 Organic-TOC [Filter+denuder]	41
PM10 Element-SR [Filter+denuder]	41
PM10 Element-NA [Filter+denuder]	41
PM10 Element-MO [Filter+denuder]	41
PM10 Gas-C [Filter+denuder]	41

PM10 Element-S [Filter+denuder]	41
PM10 Element-V [Filter+denuder]	41
PM10 Element-P [Filter+denuder]	41
PM10 Element-TL [Filter+denuder]	41
PM10 Ionic-K [Filter+denuder]	41
PM10 Element-SN [Filter+denuder]	41
PM10 Element-RB [Filter+denuder]	41
PM10 Gas-TOC [Filter+denuder]	41
PM10 Element-NH4 [Filter+denuder]	41
PM10 Element-MG [Filter+denuder]	41
PM10 Organic-C [Filter+denuder]	41
PM10 Element-SE [Filter+denuder]	41

HOLBROOK PRIMARY	**134,066**
Description	**Record Count**
PM2.5 Particle-Concentration [TEOM]	88,903
PM2.5 Particle-Mass [TEOM]	32,122
PM2.5 Flow-Vol [Not assigned]	1,028
PM2.5 Element-AG [Filter+denuder]	740
PM10 Flow-Vol [Other]	638
PM2.5 Filter-Mass [Filter+denuder]	634

MORGANTOWN SATELITE	**576**
Description	**Record Count**
PM2.5 Flow-Vol [Not assigned]	330
PM2.5 Filter-Mass [Filter+denuder]	246