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

SEMI-ANNUAL

TECHNICAL PROGRESS REPORT

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

DOE AWARD NUMBER: DE-FC26-02NT41476

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

Semi-Annual Technical Progress Report

REPORT PERIOD START DATE: February 12, 2003

REPORT PERIOD END DATE: August 11, 2003

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

DATE REPORT ISSUED: September 2003

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-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 1, which is currently in progress and will take twelve months to complete, will include 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. In Phase 2, which will be completed in the second year of the project, a platform for on-line data analysis will be developed. Phase 2 will include 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 eleventh month of Phase 1 development activities.

TABLE OF CONTENTS

LIST OF GRAPHICAL MATERIALS

Not applicable.

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.

Over the past four years, research projects sponsored by DOE-NETL have collected large amounts of data on $PM_{2.5}$ and other air pollutants at ambient monitoring sites in the upper Ohio River valley region. 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 eleventh month of the Phase One development tasks and Phase Two is expected to begin in September, 2003. Phase One consists 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.

Over the past four years, research projects sponsored by DOE-NETL have collected large amounts of data on $PM_{2.5}$ and other air pollutants at ambient monitoring sites in the upper Ohio River valley region. 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 eleventh month of Phase One 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, which is currently in progress and will take twelve months to complete, includes data inventory, benchmarking and database population tasks, as well as the development of a data management architecture, a web-based retrieval system and an internal QA/QC system. A progress summary for Phase One is shown in the following table:

Task #	Description	Planned % completed	Actual % completed
1.1	Data Inventory/Benchmarking for Database Applications	100	90
1.2	Develop Data Management System Architecture	100	90
1.3	Population of Database	100	55
1.4	Develop Web-based Retrieval System	100	30
1.5	Develop Internal QA/QC System	100	50

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

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

Task #	Description	Planned % completed	Actual % completed
2.1	Develop Stakeholder Website	0%	1%
2.2	Develop Public Website	0%	0%
2.3	Develop Online Help Feature	0%	0%
2.4	Provide Graphical Representation of Data	0%	5%
2.5	Performance Test	0%	0%

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 Semi-Annual Technical Report for this project. Those included the CARB Data Management Project in California and the EPA Supersite Database Development Project. Additional benchmarking visits are planned for Phase Two, including examining what potential stakeholders and application users desire from an analysis toolset being designed from the ground up.

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 project team is continuously developing *stored procedures*, or static queries that are stored within the database structure. Appendix Three identifies and describes the existing *stored procedures* that have been developed to date.

Task 1.3 - Population of Database

A series of conversion scripts have been created to reformat the supplied data files into the application-specific format. Some data files have been submitted that have been formatted to comply with the published NARSTO standards. Since the NARSTO format utilizes specific headings and tables that do not directly correspond with the project-specific database structure, these files cannot be directly imported into the database. The conversion scripts allow the project team to convert this data into a suitable format that will allow the submitted data to be automatically merged with the existing data in the database.

The database structure, although complete in its present form, could undergo multiple changes and alterations while developing the database population source code. Once the population code has been completely tested, a final copy of all data will be processed into the database from the static, offline datasets stored on the server and submitted to the project team. The database population software is being coded to include a data tracking system so data quality will be more manageable.

Task 1.4 - Develop Web-Based Retrieval System

There are two primary transfer protocols available for large-scale data retrieval. These are File Transfer Protocol (FTP) and Hypertext Transfer Protocol (HTTP). Each has distinct advantages and disadvantages. With this application, HTTP transfers are the primary choice for dataset retrievals, while FTP transfers are the primary choice for dataset submissions.

Once the data is transferred to the database, users can specify which data is to be retrieved through two primary ad hoc query system interfaces. These are the *Standard Query Builder Interface* and the *GIS Query Builder Interface.* Each interface provides resulting datasets in a tabular format.

AD HOC QUERY SYSTEMS

Standard Query Builder Interface

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.

The Standard Query Interface provides a logical flow in the selection process so that the user can easily select the subset of data that the user wishes to view. Therefore, the user will be able to select from a range of input criteria as well as define their result sets based on user-specified output criteria. Input criteria will include selecting date ranges, collection site locations, monitoring networks, air quality parameters, collection principles, etc. Output criteria define what columns the user wishes to include in the resulting dataset, and include items such as date and time values, air quality parameters, durations, frequencies, and metaflags. The query interface also interacts with geographic information datasets to provide the user with the additional option of data selection based on geographic parameters.

GIS Query Builder Interface

In addition to the Standard Query Interface, the project team has developed an additional query method, which allows users to build datasets based on geographic extent or based on spatial relationships with other geographic features. Although this system requires Internet mapping software on the server, the client web browser requires no specialized software or plug-in. Within the GIS Query Interface, users have the option of selecting data from multiple monitoring networks using an interactive selection box or the extent of another geographic feature such as a county or state boundary. The selected monitors are added to a *data cart* and displayed in a table at the bottom of the browser window. The tabular display allows users to deselect from the selected set of monitors by clicking on the corresponding checkbox. Upon completion of the GIS monitor selection process, users are taken to an advanced query system where they select air quality parameters and a date range prior to data query execution.

Additional functionality of the GIS query interface includes basic GIS capabilities such as measuring distances, identifying features, creating buffers, feature on feature selection, panning, zooming, and toggling layer visibility. Professionals wishing to perform further spatial analysis on air quality data with commercial GIS software packages will be able to download any or all GIS layers in a given view extent with the single click of a button.

Tabular Results

The implementation of the tabular interface uses paging and allows the user to change the number of records in a page view. The number of pages in a result set is dependent upon the number of records returned by the query. The user can browse through the dataset and has the option of reordering the display by clicking on the header of the table columns.

Users can also download the resulting datasets for use as an *offline* dataset by clicking on a hyperlink placed above and below the results table. These offline datasets are available in a variety of formats, including Comma Separated Value (CSV) format, XML format and NARSTO format.

An archive will be maintained of all the query requests, similar to the *History* list in a web browser. To view previous query results, the user simply looks for previous queries in the history list and then re-executes them.

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.

As discussed in the first Semi-Annual Technical Report, the Automatic Population Module (APM) requires that data be submitted in a prescribed format. This ensures that each value or flag is positioned in the correct location within the data file and that each item is successfully added to the database. Should any problems with the data be discovered at a later date, then that entire data transaction can be removed, replaced or updated by submitting a replacement data transfer file or removal request to the Application Administrator. This transaction layer is being built to provide reverse-posting capabilities to the application, in addition to the standard transaction processing capabilities provided by Microsoft SQL Server 2000. This layer utilizes a series of timestamp values and log files to associate groups of data with file uploading operations and username activities.

B. Phase Two Tasks

Task 2.1 - Develop Stakeholder Website

As proposed by ***ATS***, each stakeholder will have access to the entire data analysis package while the general public will have access to selected features through the public website described in Task 2.2. The stakeholder website will provide the ability to view and develop graphical representation of the digital data online for reports and for data analysis. The data analysis package will be an interactive tool that will be embedded in

the data warehouse and repository. The querying of the data permits user-defined access and review of the data. Built-in online analytical tools for advanced data analysis will provide the following options:

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

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

The web directory structure for this website is defined in *Appendix One*. It identifies the location of key elements of the website and applications. A selection of screen captures, or *screenshots*, is included in *Appendix Two*.

Task 2.2 - Develop Public Website

ATS proposed to construct a separate website connected to the data archive for public outreach, providing the citizens of the upper Ohio River valley and at-large, along with legislative and regulatory authorities, a resource and an educational tool highlighting the extensive monitoring programs undertaken by NETL. Publicly accessible sections of the database application's web space will be available to everyone without log in. This portal will be different from the one for stakeholders, sponsors and developers, and will require registered users to submit a username and password combination before access to the restricted website is granted.

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, no items have been developed for the publicly accessible portions of this application. Development is slated to proceed with Task 2.2 in Phase Two of 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

To date, no items have been developed for the publicly accessible portions of this application. Most, or all, of the information contained in the appendices to this report will be contained in the online help system, at some level. Development is slated to proceed with Task 2.2 in Phase Two of this project.

Task 2.4 - Provide Graphical Representation of Data

A Geographic Information Systems (GIS) component has been proposed to accompany full graphing and charting components. The GIS component will be developed in specific formats to support both the stakeholder and public sections of the website. These components will initially focus on the criteria pollutants to provide spatial-temporal representations. The mapping project will involve linkages to Federal and regional efforts such as the LADCO and EPA/OAQPS ozone-mapping projects. Detailed urban maps of the UORV region showing spatial resolution of criteria pollutant concentrations will be part of a system that will be enhanced with point and click display of pollutant concentrations as defined by the pollutant isopleths.

The air quality maps will also be used as an interface for additional information or data analysis on a host of variables at a specific site or within an urban air shed. The system will provide information based on user input for the following:

- Historical trends in air pollution
- Influences of meteorology on air quality (including a link to a back trajectory wind calculation module)
- Population and potential health impacts of air quality
- Other health related air quality indices (pollutant standard index, pollen/mold, toxics, etc.)

In addition to the GIS component, graphing and charting components will also be developed to provide researchers with graphs and charts suitable to copy and paste into reports and other documents. The project team is actively seeking input from potential stakeholders as to which types of graphing tools would be most beneficial to end users.

To date, no items have been developed for the publicly accessible portions of this application. Development is slated to proceed with Task 2.2 in Phase Two of this project.

Task 2.5 - Performance Testing

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

III. RESULTS AND DISCUSSION

The first phase, which is currently in progress and expected to be completed by August 2003, includes data inventory, benchmarking and database population tasks, as well as the development of data management architecture, a web-based retrieval system and an internal QA/QC system. A progress summary for Phase One is shown in the following table:

Task #	Description	Planned % completed	Actual % completed
1.1	Data Inventory/Benchmarking for Database Applications	100	90
1.2	Develop Data Management System Architecture	100	90
1.3	Population of Database	100	55
1.4	Develop Web-based Retrieval System	100	30
1.5	Develop Internal QA/QC System	100	50

N/A – on-going activity.

The summary shows that performance on most of the tasks is on schedule.

The second project development phase will be completed in the second year of the project, which starts in September 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 the following table:

Task #	Description	Planned % completed	Actual % completed
2.1	Develop Stakeholder Website	0%	1%
2.2	Develop Public Website	0%	0%
2.3	Develop Online Help Feature	0%	0%
2.4	Provide Graphical Representation of Data	0%	5%
2.5	Performance Test	0%	0%

Some developmental work has been performed on Task 2.1 to complement efforts in Task 1.4 above.

IV. CONCLUSION

The development efforts have so far proceeded as expected. Some target milestones have not been met primarily due to lack of cooperation from third party sources, or from lack of required task data. This is especially so with the data inventory task, where data reformatting issues have also been encountered.

The hurdles encountered, however, have not amounted to show stoppers 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. All the tasks for this project will be performed in the timeframe proposed.

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
   http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R01.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; a 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	Extended Markup Language

APPENDICES

Appendix One

Stakeholder Website Directory Structure

The following diagram depicts the basic structure of the web directory space on the hosting server(s):

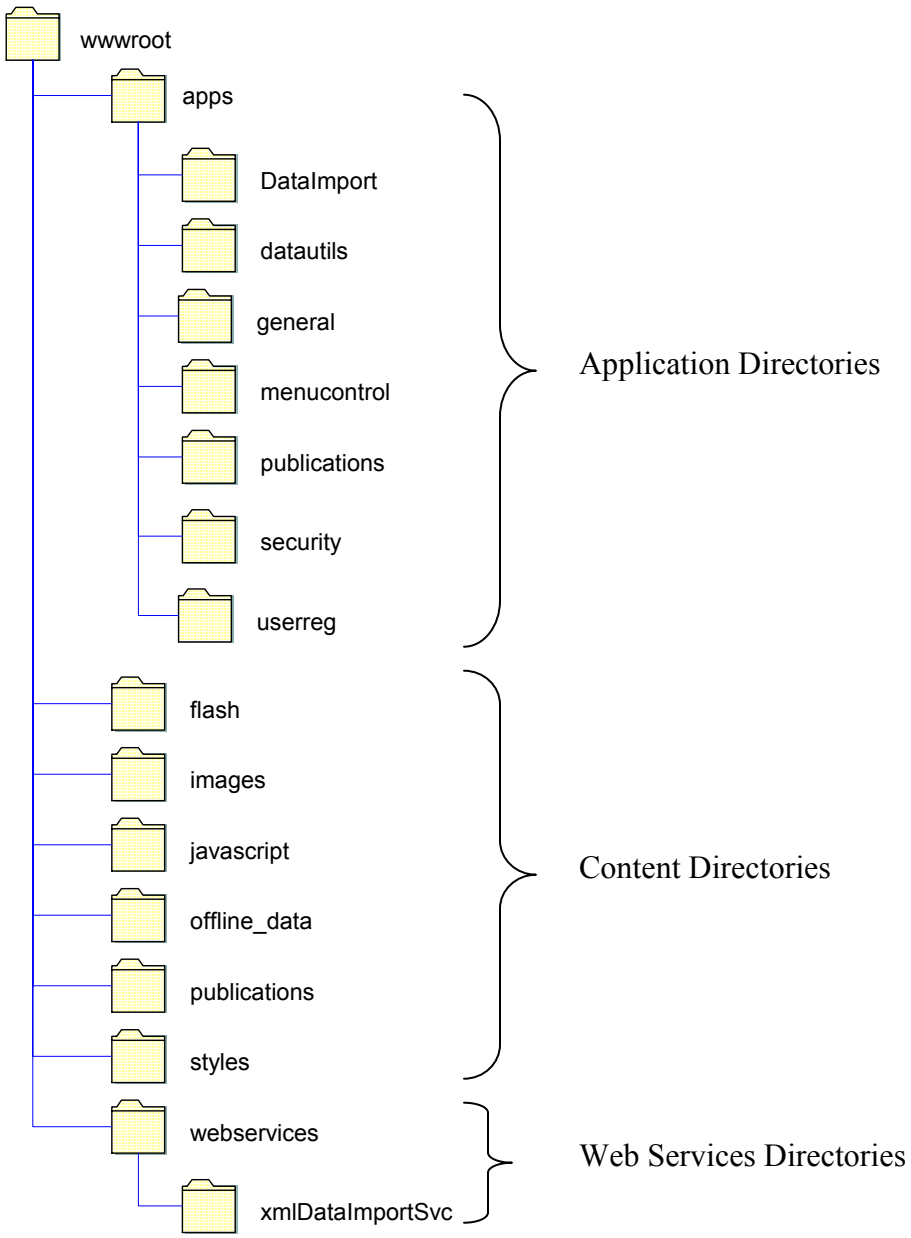


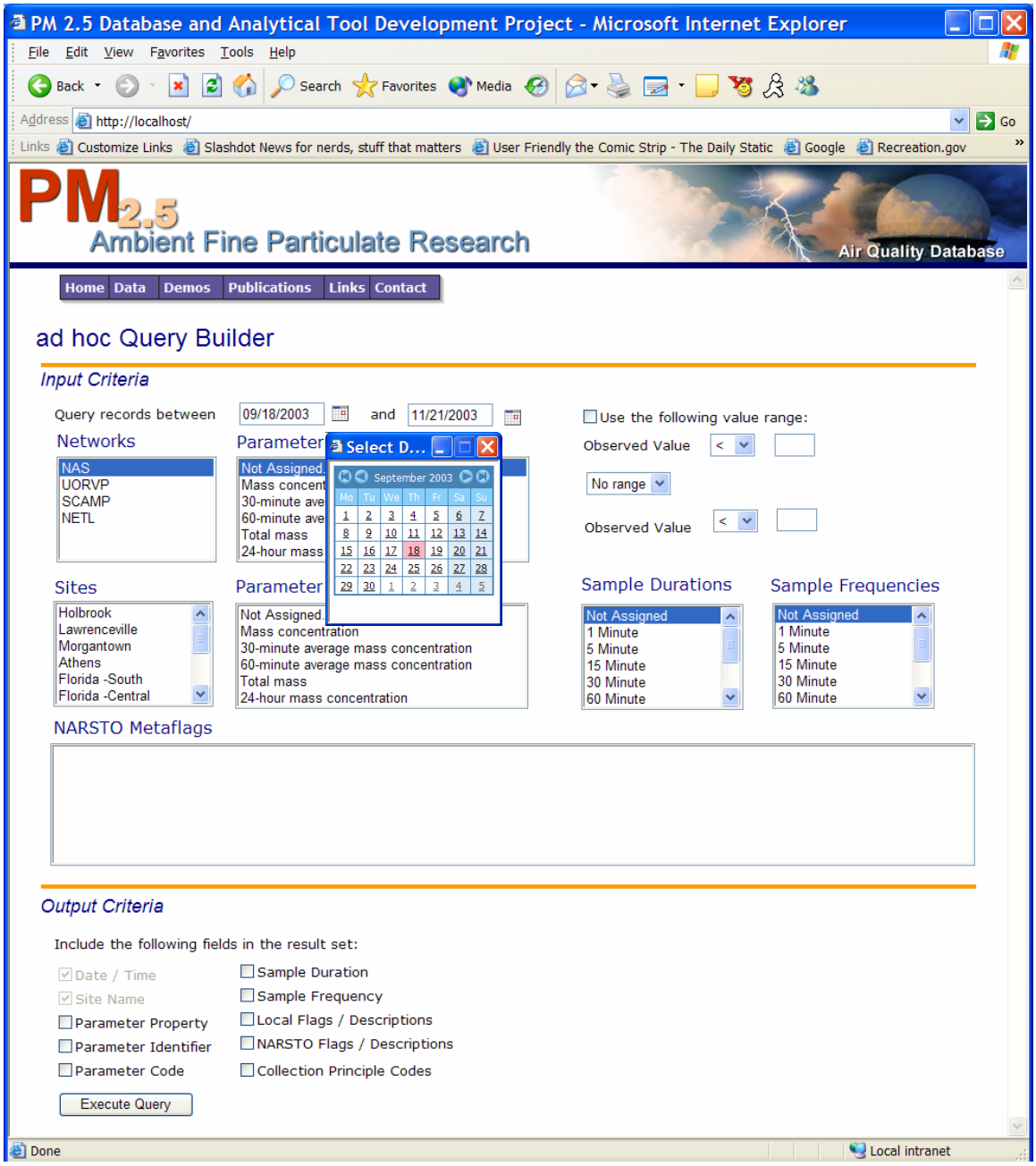
Figure 1 - Stakeholder Website Directory Structure

Appendix Two

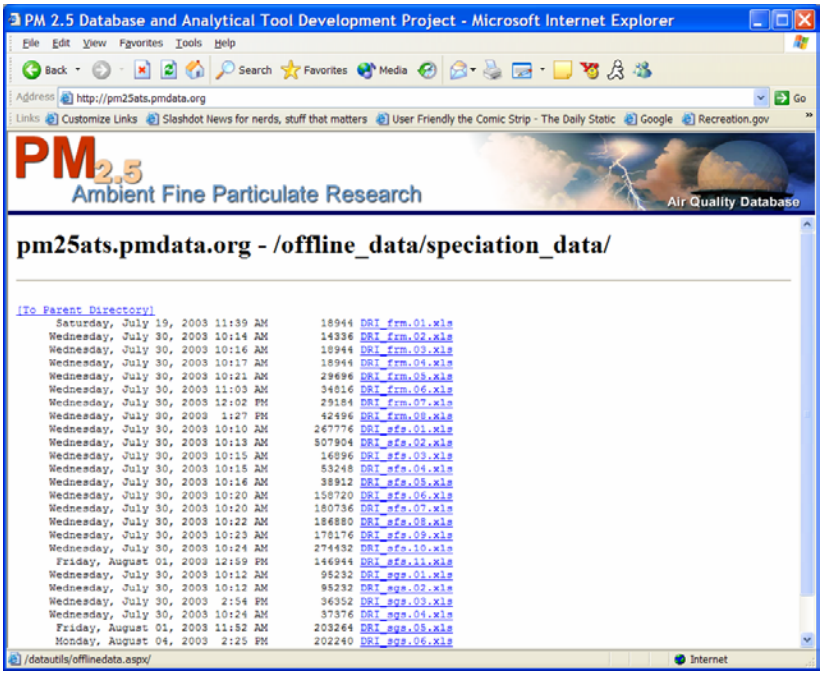
Application Screen Captures

Screenshot	Description
Screenshot 1	The application verifies a user's access level to the Stakeholder's website. If the user has not previously been authenticated, a logon prompt automatically appears and asks the user to provide a username and password.
Screenshot 2	Stakeholders will be able to customize their own queries based on specific criteria selected from the *Standard Query Builder* interface. To select dates, the user clicks on the small calendar icon to the right of the respective date fields, and then selects a date value from a popup calendar control. Other selectable items include *networks*, *sites*, *parameter properties*, *parameter identifiers*, *sample durations*, *sample frequencies* and *NARSTO Metaflags*. Users can also specify a value range if needed.
Screenshot 3	Authorized users can also download static, offline datasets from the server.
Screenshot 4	Example graphing tool: *Meteorological Influence on High PM Days*. Each data point can be removed from the plot as required.
Screenshot 5	Example graphing tool: Box-whisker plots.
Screenshot 6	Example graphing tool: Parameter comparisons. This tool is used to explore relationships between parameters for a given collection site.
Screenshot 7	Example graphing tool: GIS-style mapping capabilities provided by Java, utilizing *layer* browser technologies.
Screenshot 8	GIS query builder. This is a supplemental tool to the GIS query interface. From this screen, the user can specify a larger group of parameters for their GIS query.
Screenshot 9	Tabular results with paging enabled. To keep server-client response times low, only portions of the query results are returned per page. The user can utilize the navigational controls at the top of the table to page between results.
Screenshot 10	Tabular results with download option. The user has the option to save their results locally.
Screenshot 11	Tabular results in MS Excel. The user may open their result sets locally with MS Excel.
Screenshot 12	GIS query interface. From this screen, the user selects locations from which to retrieve data.
Screenshot 13	GIS query results. Example of GIS query results with the *data cart* displayed on same screen.

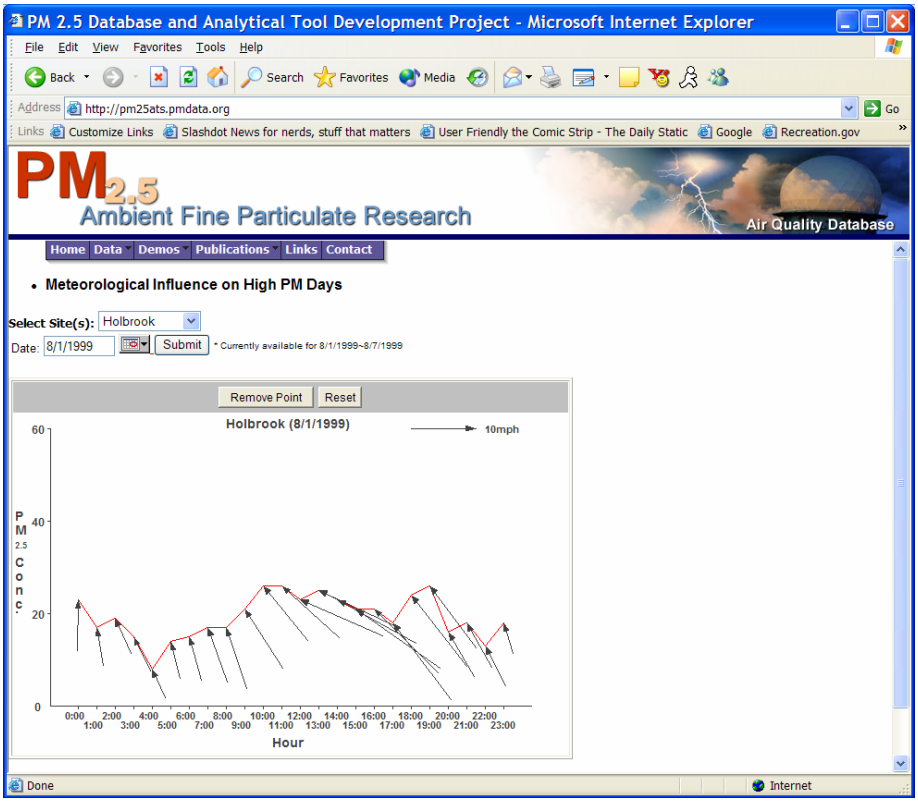
Screenshot 1 – Application Logon Prompt



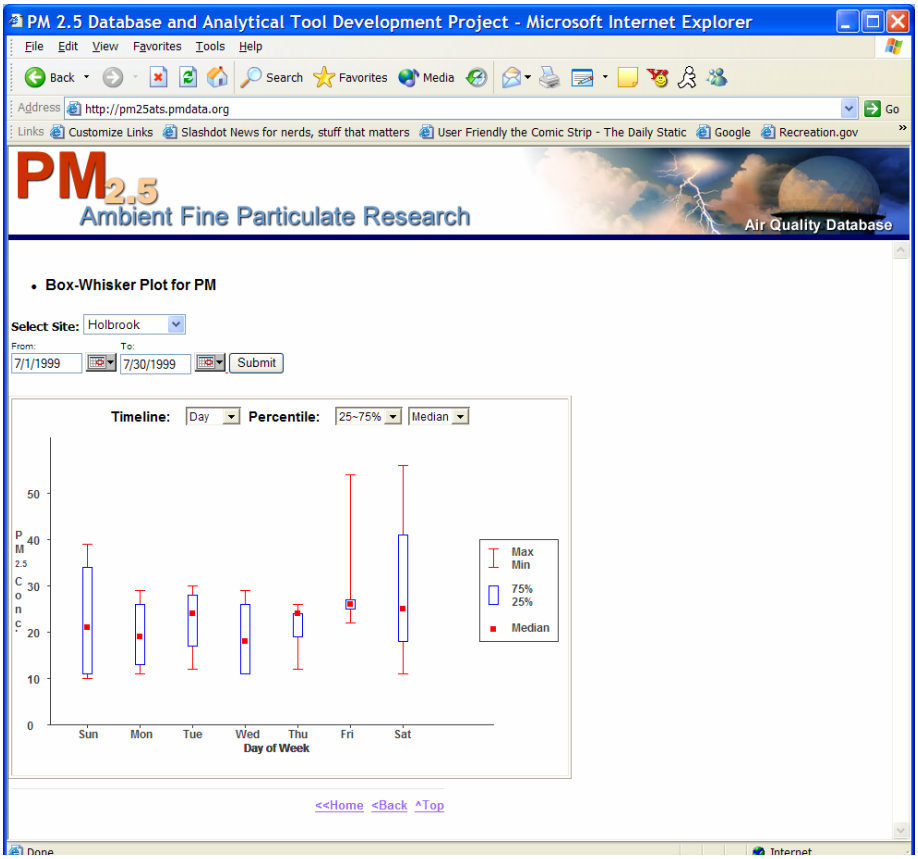
Screenshot 2 - Standard Query Builder



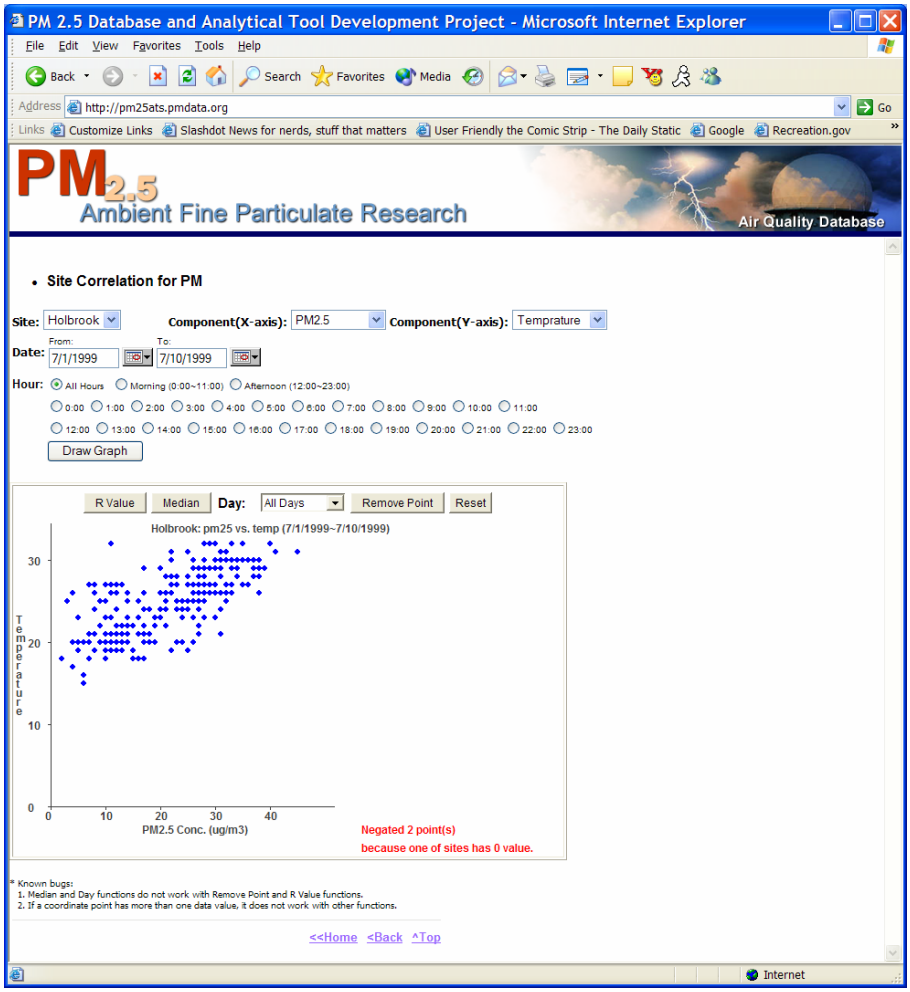
Screenshot 3 - Offline Data Downloads



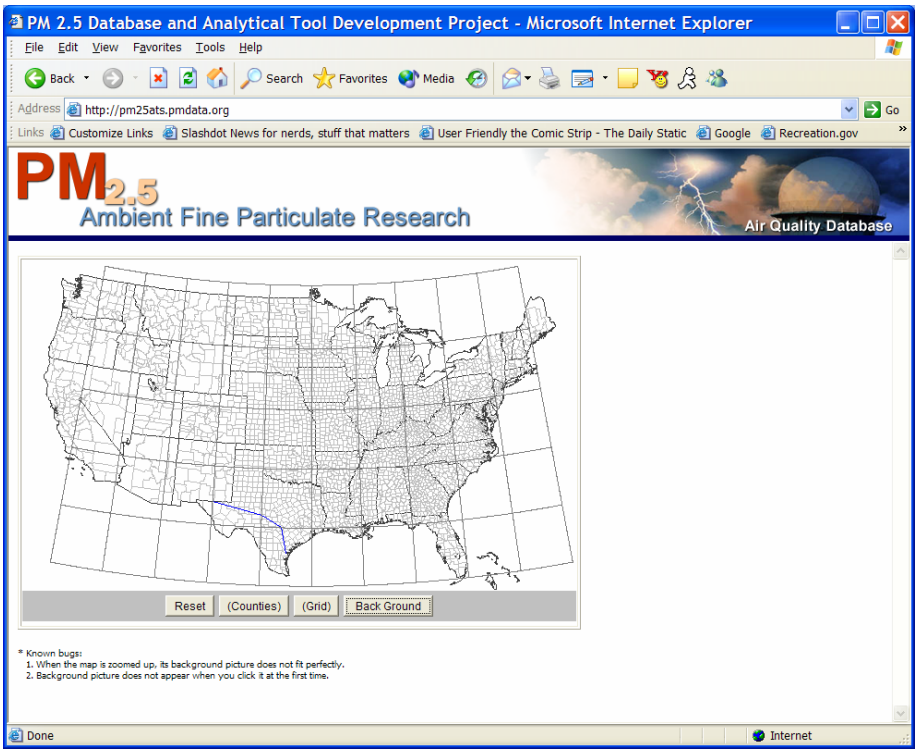
Screenshot 4 - Example Graphing Tool



Screenshot 5 - Example Graphing Tool



Screenshot 6 - Example Graphing Tool



Screenshot 7 - Example Graphing Tool

Screenshot 8 - GIS Query Builder

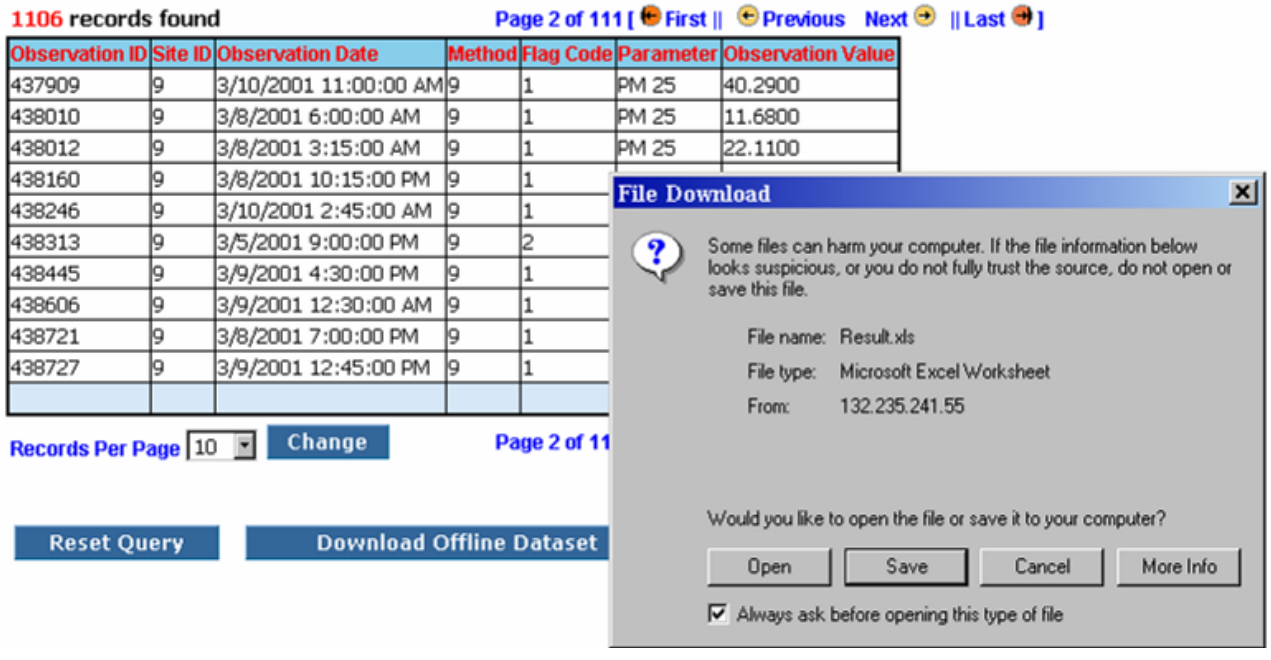
1106 records found Page 2 of 111 [First || Previous Next || Last]

Observation ID	Site ID	Observation Date	Method	Flag Code	Parameter	Observation Value
437909	9	3/10/2001 11:00:00 AM	9	1	PM 25	40.2900
438010	9	3/8/2001 6:00:00 AM	9	1	PM 25	11.6800
438012	9	3/8/2001 3:15:00 AM	9	1	PM 25	22.1100
438160	9	3/8/2001 10:15:00 PM	9	1	PM 25	28.9000
438246	9	3/10/2001 2:45:00 AM	9	1	PM 25	32.5200
438313	9	3/5/2001 9:00:00 PM	9	2	PM 25	30.6000
438445	9	3/9/2001 4:30:00 PM	9	1	PM 25	39.2900
438606	9	3/9/2001 12:30:00 AM	9	1	PM 25	29.7800
438721	9	3/8/2001 7:00:00 PM	9	1	PM 25	32.2600
438727	9	3/9/2001 12:45:00 PM	9	1	PM 25	37.6700

Records Per Page 10 Change Page 2 of 111 [First || Previous Next || Last]

Reset Query Download Offline Dataset

Screenshot 9 - Tabular Results with Paging Enabled



Screenshot 10 - Tabular Results with Download Option

Screenshot 11 - Tabular Results in MS Excel

1106 records found Page 2 of 111 [First || Previous Next || Last]

Observation ID	Site ID	Observation Date	Method	Flag Code	Parameter	Observation Value
437909	9	3/10/2001 11:00:00 AM	9	1	PM 25	40.2900
438010	9	3/8/2001 6:00:00 AM	9	1	PM 25	11.6800
438012	9	3/8/2001 3:15:00 AM	9	1	PM 25	22.1100
438160	9	3/8/2001 10:15:00 PM	9	1	PM 25	28.9000
438246	9	3/10/2001 2:45				
438313	9	3/5/2001 9:00:0				
438445	9	3/9/2001 4:30:0				
438606	9	3/9/2001 12:30				
438721	9	3/8/2001 7:00:0				
438727	9	3/9/2001 12:45				

Records Per Page 10 Change

Reset Query Dow

Microsoft Excel - Result[1].xls

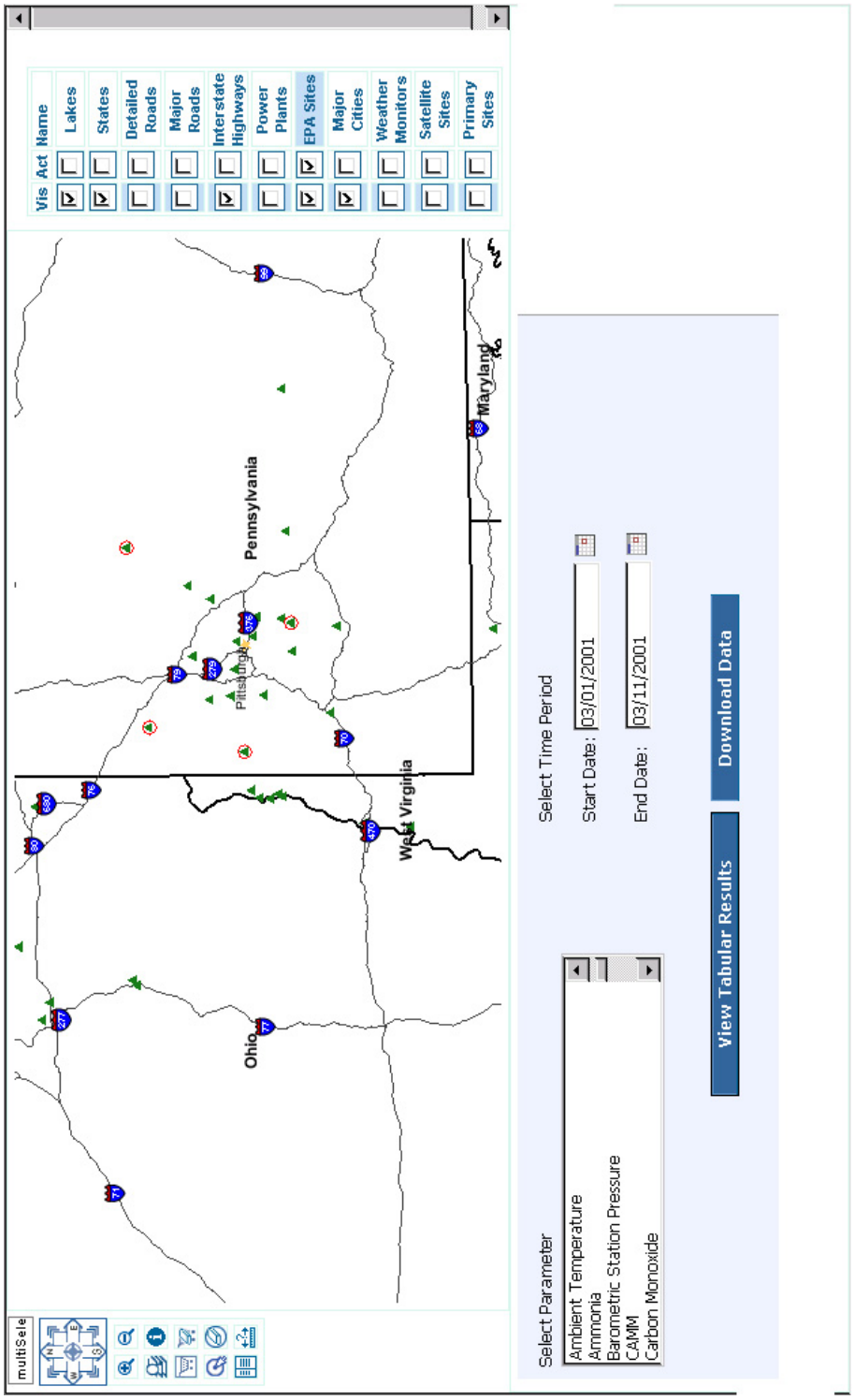
File Edit View Insert Format Tools Data Window Help Acrobat

A56 = 439048

	A	B	C	D	E	F	G	H	I
45	439009	9	3/10/2001 6:15	26.02000046	9				1
46	439012	9	3/4/2001 8:45	12.01000023	9				2
47	439013	9	3/1/2001 17:30	-999	9				1
48	439014	9	3/10/2001 3:30	28.82999992	9				1
49	439015	9	3/7/2001 21:45	24.44000053	9				1
50	439016	9	3/7/2001 14:00	38.75999832	9				1
51	439018	9	3/9/2001 6:45	38.18999863	9				1
52	439029	9	3/5/2001 11:15	27.36000061	9				2
53	439030	9	3/3/2001 11:30	29.44000053	9				2
54	439031	9	3/8/2001 4:00	57.47000122	9				2
55	439037	9	3/6/2001 6:45	33.58000183	9				2
56	439048	9	3/3/2001 13:45	33.18000031	9				2
57	439049	9	3/6/2001 0:30	30.30999947	9				2

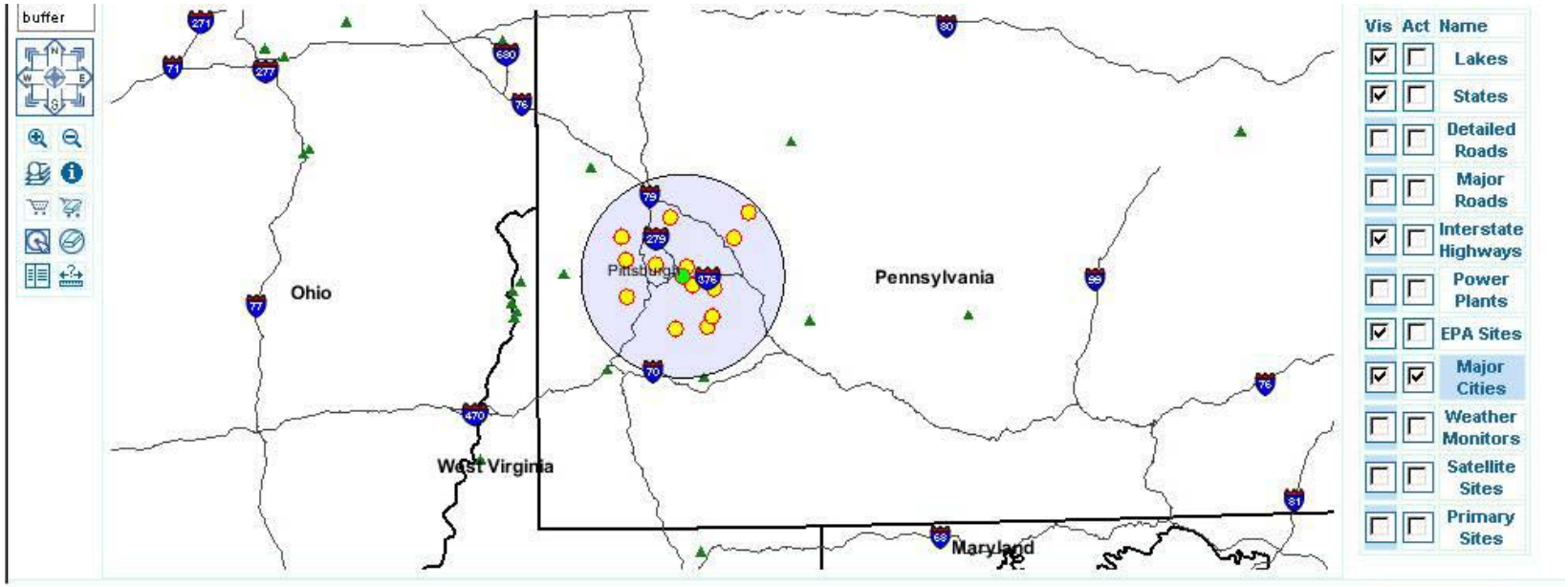
Result 1

Ready NUM



Screenshot 12 - GIS Query Interface

Screenshot 13 - GIS Query Results



Target layer: EPA Sites Buffer size: 20 Show/Change buffer Continue

EPA Sites

Select/Unselect	SITEID	LATITUDE	LONGITUDE	PM25	SO2	O3	STATE	ZSHAPEZ	ZIDZ	geoX	geoY
☑	420030008	40.465556	-79.961111	false	false	false	Pennsylvania	[Geometry]	6	588096.692752289	4479945.46112694
☑	420030116	40.473611	-80.077222	false	false	false	Pennsylvania	[Geometry]	31	578243.261311001	4480730.18845929
☑	420033007	40.294444	-79.886667	false	false	false	Pennsylvania	[Geometry]	34	594647.339943639	4461029.0761373
☑	420039002	40.546944	-79.783889	false	false	false	Pennsylvania	[Geometry]	41	602997.414744259	4489171.6398345
☑	420030131	40.289444	-80.005	false	false	false	Pennsylvania	[Geometry]	42	584595.644243847	4460354.40099155
☑	420030095	40.486944	-80.188056	false	false	false	Pennsylvania	[Geometry]	45	568834.124587036	4482117.81773471
☑	420030097	40.553056	-80.203333	false	false	false	Pennsylvania	[Geometry]	54	567473.036737472	4489444.47365003

Appendix Three

Stored Procedures

Stored Procedure Name	spAddNewUser	
Tables Referenced	tblUsers	
SQL Code	CREATE PROCEDURE [dbo].[spAddNewUser] @userName varchar(50), @passwordHash varchar(40), @salt varchar(10), @email varchar(150), @web_address varchar(150), @vCard_ID int, @phone varchar(50), @fax varchar(50), @mobile varchar(50), @pager varchar(50), @admin_Level_ID int, @browse_Level_ID int, @read_Level_ID int, @write_Level_ID int, @user_First varchar(75), @user_Last varchar(75), @user_Middle char(1) AS INSERT INTO tblUsers (username, passwd, salt, email, web_address, vCard_ID, phone, fax, mobile, pager, admin_Level_ID, browse_Level_ID, read_Level_ID, write_Level_ID, user_First, user_Last, user_Middle) VALUES(@userName, @passwordHash, @salt, @email, @web_address, @vCard_ID, @phone, @fax, @mobile, @pager, @admin_Level_ID, @browse_Level_ID, @read_Level_ID, @write_Level_ID, @user_First, @user_Last, @user_Middle) GO	
Required Parameters	Name	Type
	@userName	varchar(50)
	@passwordHash	varchar(40)
	@salt	varchar(10)
	@email	varchar(150)
	@web_address	varchar(150)
	@vCard_ID	int
	@phone	varchar(50)
	@fax	varchar(50)
	@mobile	varchar(50)
	@pager	varchar(50)
	@admin_Level_ID	int
	@browse_Level_ID	int
	@read_Level_ID	int
	@write_Level_ID	int
	@user_First	varchar(75)
	@user_Last	varchar(75)
	@user_Middle	char(1)
Comments	This stored procedure is used to add new user records to the database.	

Stored Procedure Name	spCheckDuplicateUsernames	
Tables Referenced	tblUsers	
SQL Code	CREATE PROCEDURE [dbo].[spCheckDuplicateUsernames] @userName varchar(50) AS SELECT username FROM tblUsers WHERE username = @userName GO	
Required Parameters	Name	Type
	@userName	varchar(50)
Comments	This stored procedure is used to prevent users from creating user accounts with duplicate usernames.	

Stored Procedure Name	spDataSubmitterRequest	
Tables Referenced	tblDataSubmitters	
SQL Code	CREATE PROCEDURE [dbo].[spDataSubmitterRequest] @user_ID int AS INSERT INTO tblDataSubmitters (user_ID) VALUES (@user_ID) GO	
Required Parameters	Name	Type
	@user_ID	int
Comments	This stored procedure is used to queue requests from users to become a member of the data submitters group.	

Stored Procedure Name	spStakeholderRequest	
Tables Referenced	tblStakeholders	
SQL Code	CREATE PROCEDURE [dbo].[spStakeholderRequest] @user_ID int AS INSERT INTO tblStakeholders (user_ID) VALUES (@user_ID) GO	
Required Parameters	Name	Type
	@user_ID	int
Comments	This stored procedure is used to queue requests from users to become a member of the stakeholder group.	

Stored Procedure Name	spSubscriberRequest	
Tables Referenced	tblSubscribers	
SQL Code	CREATE PROCEDURE [dbo].[spSubscriberRequest] @user_ID int AS INSERT INTO tblSubscribers (user_ID) VALUES (@user_ID) GO	
Required Parameters	Name	Type
	@user_ID	int
Comments	This stored procedure is used to allow users to subscribe to periodic email updates and announcements.	

Stored Procedure Name	spUserLookup	
Tables Referenced	tblUsers	
SQL Code	CREATE PROCEDURE [dbo].[spUserLookup] @userName varchar(50) AS SELECT passwd, salt, admin_Level_ID, browse_Level_ID, read_Level_ID, write_Level_ID FROM tblUsers WHERE username = @userName GO	
Required Parameters	Name	Type
	@userName	varchar(50)
Comments	This stored procedure is used to retrieve passwords, salt values and security settings upon an authentication request by the user.	