

STATEMENT OF WORK

NETL Cooperative Agreement DE-FC26-02NT41476

Database and Analytical Tool for the Management of Data Derived from U. S. DOE (NETL) Funded Fine Particulate (PM_{2.5}) Research

PROJECT SCOPE

Advanced Technology Systems, Inc. with Ohio University and Texas A&M University - Kingsville as subcontractors, will develop a state-of-the-art, scalable and robust computer application for NETL to manage the extensive data sets resulting from the DOE-sponsored ambient air monitoring programs in the upper Ohio River valley region. Efforts will be made to include, to the greatest extent possible, ambient air data collected by other agencies in the upper Ohio River valley region, such as 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 will be placed on data collected in the upper Ohio River valley region, the computer application developed under this Agreement will be designed, to the greatest extent possible, to access data collected at NETL-sponsored ambient air monitoring sites outside the region, such as sites operated by the Tennessee Valley Authority in the Great Smoky Mountains (under DOE Interagency Agreement DE-AI26-98FT40406) and by Southern Research Institute in North Birmingham, AL (under DOE Cooperative Agreement DE-FC26-00NT40770). The data base and analytical tool development effort will also be coordinated, to the greatest extent possible, with similar tools being developed for use by U.S. EPA. This will ensure that the database and analytical tools produced under this Agreement will be readily accessible to a wide variety of stakeholders.

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.

TASK DESCRIPTIONS

The project will be conducted in two phases. In Phase 1, which will take twelve months to complete, a data repository and warehouse will be developed. Phase 1 will include the following tasks: (1) data inventory/benchmarking for data application, 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.

Phase 1: Development of the Data Repository and Warehouse

Task 1.1 - Data Inventory/Benchmarking for Database Applications: The first step to developing a large database is to determine the types and number of data it will contain. The Upper Ohio River Valley Project (UORVP) has generated ambient air PM_{2.5} and PM₁₀ mass and chemical species data along with ambient precursor gas and meteorological measurements. However, the other three NETL-sponsored ambient air monitoring programs in the region (Steubenville Comprehensive Air Monitoring Project, DE-FC26-00NT40771; Atmospheric Aerosol Source-Receptor Relationships: The Role of Coal-Fired Power Plants, DE-FC26-01NT41017; and NETL Office of Science and Technology research), with somewhat different objectives, have collected other data for parameters not measured in the UORVP. These include ambient air hydrocarbons, mold spore counts and other exotic species. Initially, an inventory will be taken of all the types of data collected in the four NETL programs. This will include the so-called “metadata,” which provides information about measurement and analytical methods associated with each type of data, and quality assurance information in the form of field sampling and laboratory analysis “flags” associated with each individual measurement. At this time a survey of the participants in the other three programs will be taken to determine the types of data analysis tools they are utilizing or might be interested in obtaining. This information will be used to develop the initial design of the Online Analytical Processing Package, which will be developed in Phase II of this project.

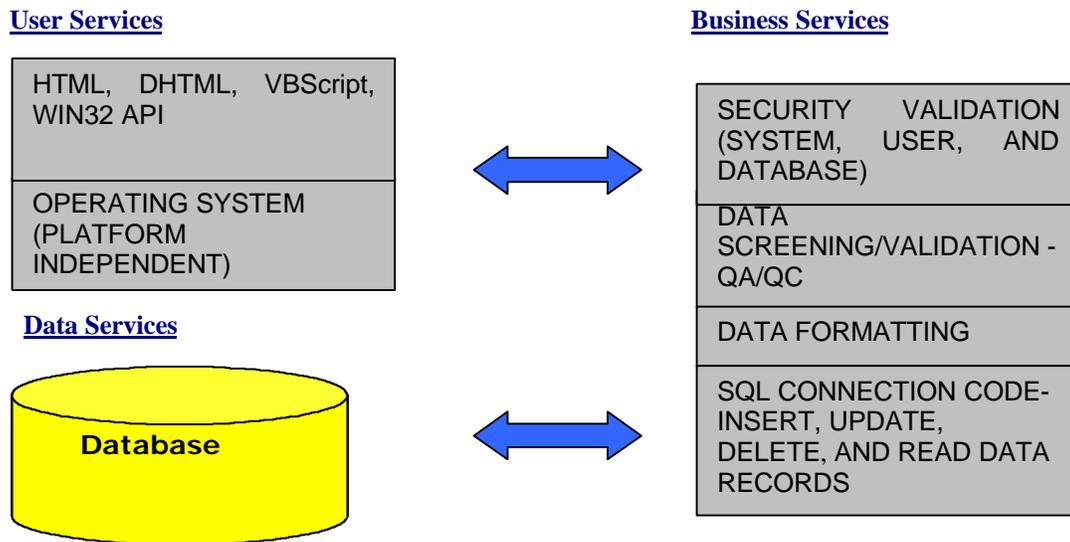
Participants in the NETL-sponsored air monitoring efforts described above will form the core of an external stakeholder group that will be maintained throughout the project. However, efforts will be made under Task 1.1 to broaden the stakeholder group to include other entities such as: (1) local, state, and Federal agencies involved with air monitoring in the upper Ohio River valley region, including the Pennsylvania Department of Environmental Protection (PA-DEP), West Virginia Division of Environmental Protection (WV-DEP), Ohio EPA, and the Allegheny County Health Department (ACHD); (2) NETL-sponsored organizations that operate ambient air monitoring sites outside the upper Ohio River Valley region, including the Tennessee Valley Authority (DOE Interagency Agreement DE-AI26-98FT40406) and Southern Research Institute (DOE Cooperative Agreement DE-FC26-00NT40770); and (3) personnel involved in the development of ambient air quality databases and analytical tools for use by U.S. EPA. This will ensure that the database and analytical tools compiled under this Agreement will be readily accessible to a wide variety of stakeholders.

At this stage of progress in Task 1.1, the approximate size of the database will have been determined and enough will have been learned about the data types to begin the design of the structure of the relational database. After obtaining the input from this scientific community, a benchmarking effort will be undertaken to determine if there are lessons to be learned from previous data development schemes that may have been developed by other groups. A literature search on such data management systems particularly those established by environmental organizations, whose focus is on regulatory activities and on providing public information, will be carried out. Although these types of organizations utilize larger databases with less complex inner structures than would be appropriate for the NETL studies, information about these types of databases will enable the determination of currently available software that may be applicable for both data management and data analysis. It may be possible to incorporate some of these software applications as “modules” into the data management system. Also, an examination of related public websites will provide us with examples of interactive systems, which will aid in the design of the interactive animated help feature in Phase II. On a more basic, detailed level, the knowledge gained from the literature search will be used to establish, if possible, a data format compatible with other large databases such as NARSTO and to institute established, widely used and accepted data handling standards such as file transfer protocols and internal QA/QC procedures as part of the project’s data management system. With the completion of Task 1.1, the specific information required to build the database will have been obtained and a foundation for the selection of specific analytical tools to be developed in Phase II work will have been established.

Task 1.2 - Develop Data Management System: The data management system will be duplicated at multiple sites around the country to provide maximum availability to end-users. By establishing redundant locations, a single line failure will not prevent remote users from accessing the online data store through Internet connections or modem calls. The system will have detailed built-in data querying and reporting tools, and will also allow the end user to define special queries and reports. The data querying and reporting tools will be linked to an online interactive data analysis package. The structure of the data repository will facilitate data reporting conventions, validation, and metadata, if possible, in a format compatible with NARSTO’s Data Management guidelines.

The application architecture will be a web-based solution with information stored in a relational database. Figure 1 provides an illustrative overview of the proposed architecture. The architecture sections shown in Figure 1 can either be all stored on a single computer, each stored on separate computers, or any combination of computers. This means that the application will be *scalable* because it can work with small data sets or large ones without changing the basic computer programming code defining the program flow.

Figure 1



Task 1.3 - Population of the Database: After the selection of the database software and hardware, the process of populating the database will begin by defining the internal database structure through the construction of a logical design diagram. Next, a detailed database table description based on the data inventories for each of the NETL-sponsored monitoring programs will be written. *If feasible, the database will include data from monitoring stations in the upper Ohio River valley operated by local, State, and other Federal agencies, and/or data from other NETL-sponsored monitoring sites external to the upper Ohio River valley region.* The database table description will provide standardized codes for each data point for the following parameters:

- Site ID (location, principal investigator, etc.)
- Sampling Methods (instrument, sampling time and duration, sampling media, audit methods, etc.)
- Analytical Methods
- Data quality flags (sampling and analytical)

The system will be standardized and, if possible, will utilize the NARSTO Data Management guidelines. The disparate data sets from the individual UORV-PM_{2.5} monitoring programs will be mapped where possible into a common coding system to provide a uniform data delivery system of transfer files. Following the inventory and development of the database table description, the population of the relational database will be undertaken. Detailed log tables will be maintained to track the population process for the database. These logs will be used as part of the quality control/audit procedures for data screening and will be utilized to prepare a summary report, highlighting the data

in the system available for downloading and analyzing. Written quality control procedures will be developed and automated where possible. The quality control procedures will be utilized to audit the integrity of the data population and data manipulation techniques used to produce table summaries and in the application of online analytical tools.

Task 1.4 - Develop Web-Based Retrieval Systems: A comprehensive web-based retrieval system will be developed to assist the various collaborating scientists, project sponsors, stakeholders and interested parties. The system will be two-tiered comprising a comprehensive analytical section for the scientists and other interested parties and a generic information dissemination tool for the citizens of the upper Ohio River valley region and at-large, details of which will be discussed later in the description of Phase 2. A web-based interface will be developed for remote database management with access through a standard Internet browser. Easy to understand Graphical User Interfaces (GUI) will be utilized for the data retrieval and management system. Extensive online help will support the retrieval system. The retrieval system will provide users with a series of drill-down queries to develop a concise data output file for viewing, online analysis, and/or downloading directly to their computers.

Some considerations during development will include data transfer protocols from the present data collection system, data screening/formatting and security. The data formats will be standardized to prevent data corruption and to facilitate automation of data imports from the participating program centers.

Task 1.5 - Develop Quality Assurance/Quality Control System: An approved Quality Assurance Project Plan (QAPP) will be developed prior to integrating the data sets from the UORV-PM_{2.5} air monitoring programs. The QAPP will detail all quality assurance/quality control items such as data manipulation, integration and audit methods. Where possible, the screening and quality assurance checks will be automated. The automated screening techniques will use validity checkpoints and rates of change to check each record entered into the database to identify errors. These automated scripts will also be utilized to evaluate online data analysis techniques prior to posting the information.

Phase 2: Development of an Online Analytical System (Data Analysis Package)

Task 2.1 - Develop Stakeholder Website: Stakeholders 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, sometimes referred to as “data mining,” 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)

Task 2.2 - Develop Public Website: A separate website connected to the data archive will be developed 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.

An interactive web page 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. The system 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.

Task 2.3 - Develop Online Help Feature: The online help feature will be developed to support both the Sponsor/Stakeholder and the Public Websites. The online help and instruction component of the web page 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

Task 2.4 - Provide Graphical Representation (Mapping) of the Data: The mapping feature will be developed in specific formats to support both the Stakeholder website and

the Public website. The mapping projects 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.)

Task 2.5 - Performance Test: This task will involve the β -testing of the developed database and analytical tools to insure that the intended objectives are met or exceeded. This effort may require revisiting and reworking the original designs. Consequently, this will be an ongoing exercise in Phase 2 of this project.

PROJECT SCHEDULE/MILESTONES

Phase 1: Development of the Data Repository and Warehouse

Tasks	Months after Project Start												
	1	2	3	4	5	6	7	8	9	10	11	12	
1.1 - Data Inventory/Benchmarking for Database Applications	█												
1.2 - Develop Data Management System			█										
1.3 - Population of Database				█									
1.4 - Develop Web-based Retrieval System					█								
1.5 - Develop QA/QC System						█							

Phase 2: Development of an Online Analytical System (Data Analysis Package)

Tasks	Months after Project Start												
	13	14	15	16	17	18	19	20	21	22	23	24	
2.1 - Develop Stakeholder Website	█												
2.2 - Develop Public Website		█											
2.3 - Develop Online Help Feature			█										
2.4 - Provide Graphical Representation of Data				█									
2.5 - Performance Test					█								