

**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: August 12, 2002

REPORT PERIOD END DATE: February 11, 2003

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

DATE REPORT ISSUED: March 13, 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 sixth month of Phase 1 development activities.

DISCLAIMER

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

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	iii
I. INTRODUCTION.....	4
A. PROJECT GOALS AND OBJECTIVES.....	4
B. PROJECT PHASE DEVELOPMENT.....	5
II. EXPERIMENTAL.....	6
A. PHASE ONE TASKS.....	6
Task 1.1 – Data Benchmarking/Inventory for Database Applications.....	6
Task 1.2 – Develop Data Management System Architecture.....	11
Task 1.3 – Population of Database.....	19
Task 1.4 – Develop Web-based Retrieval System.....	21
Task 1.5 – Develop QA/QC System.....	23
B. PHASE TWO TASKS.....	23
Task 2.1 – Develop Stakeholder Website.....	23
Task 2.2 – Develop Public Website.....	24
Task 2.3 – Develop Online Help Feature.....	25
Task 2.4 – Provide Graphical Representation of Data.....	25
Task 2.5 – Performance Testing.....	26
III. RESULTS AND DISCUSSION.....	26
IV. CONCLUSIONS.....	27
V. REFERENCES.....	27
APPENDIX 1 – DATA FILE INVENTORY LISTS.....	28
C. UORVP DATA FILES.....	28
D. PITTSBURGH SUPERSITE DATA FILES.....	34
E. NETL SITE DATA FILES.....	37
F. SCAMP DATA FILES.....	37
APPENDIX 2 – SITE SPECIFIC DATA CHARTS.....	38
A. UORVP MONITORING PROGRAM.....	38
B. SCAMP MONITORING PROGRAM.....	40
C. PITTSBURGH SUPERSITE PROGRAM.....	43
D. NETL SITE.....	43
APPENDIX 3 – DATABASE TABLE DEFINITIONS.....	44
A. SYSTEM TABLES.....	44
B. SECURITY TABLES.....	47
C. OBSERVATION TABLES.....	52
D. SUPPORT TABLES.....	64
APPENDIX 4 – STORED PROCEDURE DEFINITIONS.....	69
APPENDIX 5 – DATABASE RELATIONSHIP DIAGRAMS.....	70
APPENDIX 6 – DATA INPUT FILE REQUIREMENTS.....	74
A. REQUIRED INPUT FILE HEADER FIELDS.....	74
B. TABULAR FORMAT REQUIRED TAGS AND KEYWORDS.....	77
C. XML FORMAT REQUIRED TAGS.....	78
D. EXAMPLE OF PROPERLY FORMATTED TABULAR DATA INPUT FILE.....	84
E. EXAMPLE OF PROPERLY FORMATTED XML DATA INPUT FILE.....	86
APPENDIX 7 – QUALITY CONTROL DEFINITIONS.....	91
APPENDIX 8 – DEFAULT SECURITY LEVELS.....	92

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 sixth month of the Phase One development tasks and Phase Two is expected to begin in August, 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 sixth 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%	75%
1.2	Develop Data Management System Architecture	100%	55%
1.3	Population of Database	55%	55%
1.4	Develop Web-based Retrieval System	36%	10%
1.5	Develop Internal QA/QC System	14%	14%

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%	0%
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.

a) **BENCHMARKING**

Several areas were examined by the *ATS* project team in relation to this project task. Efforts underway in other states, such as the California Air Resources Board's development of a database application for air quality data obtained through their various collection efforts were evaluated. This provided much insight into what would be required of the project team in developing a database application to store, search and retrieve datasets resulting from the various NETL-sponsored data collection projects.

(1) **CARB Data Management Project**

The California Air Resources Board (CARB) is sponsoring a project very similar to this project. As part of our benchmarking efforts, the *ATS* project team initiated a conference call to members of the CARB Data Management Project team to discuss their development efforts to date.

Several issues were discussed, and the findings can be summarized as follows:

- The project has been underway for three years.
- To date the entire project, including data collection efforts, has cost a total of approximately forty-four million dollars.
- The data collection process lasted a total of fourteen months.
- The data engine used is MS SQL Server
- Program allows for Quality Assurance Automation and ad hoc Queries.
- The project is in the last/final phase of design
- The database contains approximately one hundred million records.
- The program provides users with raw data.
- The interface still needs to be completed.

The most critical input obtained from the CARB data management project was:

- Standards must be established early in the development phase.
- Standards are perhaps the most important part of the project.
- Be prepared to assist data submitters with the QA/QC and formatting operations.
- They came out with a standard input format in order to automate the file screening and uploading process, which saves the data manager the arduous task of deciphering various formats and writing programs to convert it to the common database format.

- Normalization allows for different kinds of data like meteorological data and air quality data to be stored in similar table structures.
- Though normalization of the tables causes the size to increase, it provides a huge benefit for applications.
- They defined a “method” to describe how an instrument was used to make a measurement. This helps in tracking all measurements made by a single instrument.

(2) EPA Supersite Data Management Project

The DOE/*ATS* project team initiated a conference call on September 24, 2002, to discuss ongoing efforts by the University of Maryland (UMD) to develop a relational database for ambient air data collected at the “PM Supersites” sponsored by the U.S. Environmental Protection Agency (EPA). It was determined that the efforts of the EPA/UMD project team were at approximately the same stage of development as those of the DOE/*ATS* team; therefore, no experience or “lessons learned” by the EPA/UMD team were available to guide the efforts of the DOE/*ATS* team. However, the two project teams agreed to continue to communicate periodically during the course of our parallel development efforts. All significant results of these interactions will be reported in subsequent Technical Progress Reports.

b) INVENTORY

The first step to developing a large database is to determine the types and number of data points 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; CMU’s Project: Atmospheric Aerosol Source-Receptor Relationships: The Role of Coal-Fired Power Plants, DE-FC26-01NT41017; and the 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 would be taken of all the types of data collected in the four NETL programs, including *metadata*. The metadata 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.

As the project team began to examine the data files provided by the data collection teams, one trend was very clear: no one standard was being utilized to catalog and archive the data being collected. The data collection teams needed a common format to use for submitting data.

The data files collected were a variety of Microsoft Excel Worksheets, Microsoft Access database files and text files. Each collection team utilized their own set of data parameters, headings, flags and formats. Therefore, a common specification had to be developed to provide a smooth experience when transferring data to the database server from the data collection teams and to protect the data integrity of the database.

(1) UORVP Data

The overall goal of the *Upper Ohio River Valley Project* (UORVP) was to investigate the nature and composition of fine particulate (PM_{2.5}) and its precursor gases in the Upper Ohio River valley and provide a better understanding of the relationship between coal-based power system emissions and ambient air quality in this region. Accordingly, primary ambient fine particulate matter (PM_{2.5}) and the gaseous precursors (e.g., SO₂ and NO_x) including secondary fine particles (e.g., ammonium

sulfates and nitrates) and condensable species such as ammonia and nitric acid vapor were all monitored.

Two urban and two rural monitoring sites were included in the UORVP. The two urban site included Lawrenceville in Pittsburgh, Pennsylvania, while the second urban site was located at Morgantown, West Virginia. The two rural monitoring sites were located at Holbrook, Pennsylvania and Athens, Ohio.

Along with the discrete filter-based samplers used for intermittent sampling, continuous PM_{2.5} mass measuring instruments were in operation at the Lawrenceville and Holbrook sites. Tapered Element Oscillating Mass balances (TEOM) provided continuous mass measurements of PM_{2.5} and PM₁₀.

Surface meteorological data were collected at the Lawrenceville and Holbrook sites. Wind speed and direction, temperature, barometric pressure, relative humidity, precipitation and solar radiation sensors were operational at Lawrenceville while wind speed and direction, and temperature data were collected at Holbrook. The following table summarizes the quantity of data records associated with the UORVP sites:

Site Name	Number of Parameters Tracked		Start Date	End Date	Number of Records ¹
Holbrook Primary	Continuous:	9	Feb 1999	Sep 2002	207,360
	Discrete:	115			95,335
Athens Satellite	Continuous:	0	Feb 1999	Sep 2002	0
	Discrete:	115			16,905
Lawrenceville Primary	Continuous:	13	Feb 1999	Sep 2002	299,520
	Discrete:	115			220,225
Morgantown Satellite	Continuous:	0	Feb 1999	Sep 2002	0
	Discrete:	115			17,710

¹ Values represent estimates of observation data only, based on data submitted to date.

(2) SCAMP Data

The Steubenville Comprehensive Air Monitoring Project (SCAMP) monitoring program was conducted from May 2000 through May 2002. The objective of SCAMP was to measure the concentrations of PM_{2.5} and other potential air pollutants at ambient monitoring stations in and around Steubenville, OH, and relate them to the pollutant concentrations in air that is actually breathed by people living in the area. The structure of the outdoor ambient monitoring network at Steubenville and the four satellite sites varied in the amount and type of data they collected.

The SCAMP monitoring program primary site was located at Steubenville. At Steubenville, PM_{2.5} levels were determined everyday and chemically characterized for ions (SO₄²⁻, Cl⁻, NO₃⁻, NH₄⁺) and water-soluble elements every fourth day. PM₁₀ levels were determined every day and chemically characterized for ions (SO₄²⁻, Cl⁻, NO₃⁻, NH₄⁺) and water-soluble elements every fourth day. A PM_{2.5} Speciation (4 channel unit) was operated every fourth day to compliment day-4 chemical characterizations. This unit consisted of a Nylon filter (in series with an MgO denuder) to determine NO₃⁻ artifacts; additional Teflon® filter for total elements; a quartz filter to determine EC/OC/CC and TC; and an extra Teflon® filter for any additional chemical characterization. Furthermore, ambient gases, TEOM measurements and metrological data were collected continuously and Pollen/Mold/Spores collected on weekly tapes.

The SCAMP monitoring program also collected data at the four remote sites located at Hopedale, Wheeling, Tomlinson Run State Park (TRSP) and St. Vincent College. At these sites PM_{2.5} levels were determined everyday and were chemically characterized for ions (SO₄²⁻, Cl⁻, NO₃⁻, NH₄⁺) and water soluble elements every fourth day.

Owing to the variety in the data collected at various sites, it was deemed necessary to conduct telephone conferences with the contractor (CONSOL) to understand the data collection methodology. CONSOL provided the data in an MS Access (.mdb) database format, and with the help of the clarifications obtained from the telephone conferences and a fairly extensive documentation provided by CONSOL, scripts were generated to convert this format into the input format specification of the automated population module. The first priority was given to the mapping of the flagging system of the SCAMP monitoring program to the NARSTO flags as was required for the database population. With the scripts for conversion of the data from the program to the input format in place, we await the final installment of the data from the two-year SCAMP monitoring program.

Site Name	Number of Parameters Tracked		Start Date	End Date	Number of Records ¹
Steubenville Primary Site	Continuous:	33	May 2000	May 2001	911,435
	Discrete:	34			10,215
TRSP Satellite	Continuous:	0	May 2000	May 2001	0
	Discrete:	26			3,192
St Vincent College	Continuous:	0	May 2000	May 2001	0
	Discrete:	26			3,190
Wheeling Jesuit University Satellite	Continuous:	0	May 2000	May 2001	0
	Discrete:	26			3,297
Hopedale Satellite	Continuous:	0	May 2000	May 2001	0
	Discrete:	26			1,479

¹ Values represent estimates of observation data only, based on data submitted to date.

(3) Pittsburgh Supersite Data

Carnegie Mellon University, in conjunction with DOE, RJ Lee Group, University of Maryland College Park, University of California at Davis, University of Delaware, Clarkson University, Ohio University, Brigham Young University and Rutgers University collected ambient air quality data for a period of eighteen months (May 2001 – October 2002) and this data is being incorporated into the PM_{2.5} database. The collection efforts are part of the larger Pittsburgh Air Quality Study and the purpose was to collect data to be part of an extensive database of ambient

PM data. The study included two summers and consisted of regular measurement periods and three 14-day intensive collection periods.

Site Name	Number of Parameters Tracked		Start Date	End Date	Number of Records ¹
	Continuous:				
Pittsburgh Supersite	Continuous:	18	July 2001	October 2002	4,335,408
	Discrete:	0			0

¹ Values represent estimates of observation data only, based on data submitted to date.

Since the Carnegie-Mellon site is one of EPA’s “PM Supersites”, data from this site will be included in both the DOE/*ATS* and the EPA/UMD databases. This common dataset will facilitate future collaboration between the two database project teams, especially when evaluating options for querying and retrieving data.

(4) NETL Site Data

The NETL Site is monitored and maintained by NETL’s in-house Research and Development Ambient Air Quality Research Program at the Office of Science & Technology (OST), located in Pittsburgh, PA. The OST collaborates with CONSOL Energy, Inc. and the Allegheny County Health Department to analyze the PM_{2.5} FRM sampler performance. Monitored data includes O₃, SO₂, NH₃, NO_y, NO_x, CO, H₂S, Peroxide, PM_{2.5} particulates and meteorological data. OST’s core capabilities and competencies include characterizations of organic and inorganic compounds. New additions to the resources located at OST’s two in-house laboratories include a Kratos MS50 high-resolution mass spectrometer for detailed characterization of organic compounds.

Site Name	Number of Parameters Tracked ²		Start Date	End Date	Number of Records ¹
	Continuous:				
NETL Supersite	Continuous:	3	Jan 2001	Dec 2001	13,196
	Discrete:	0			0

¹ Values represent estimates of observation data only, based on data submitted to date.

² Eight (8) parameters reported, three (3) parameters submitted.

(5) Southern Fine Particulate Monitoring Project

The monitoring site being operated by the Southern Research Institute (SRI) in Birmingham, AL, is part of a nationwide NETL-sponsored network of monitoring stations, set up to aid in the monitoring of fine particulate matter (PM_{2.5}). The location of the site in the Deep South with the region’s climatologic and topographic differences, as well as a variety of emission sources, serve to complement the PM_{2.5} data collected in the UORV. These differences are all possible factors that could affect the chemical make-up and airborne transport of fine particulates in ways that are different than factors present in dissimilar regions. SRI has supplemented their NETL funding by establishing collaborative relationships with the Jefferson County Health Department (JCHD), EPA and the Electric Power Research Institute/Southern Company Southeastern Aerosol

Research and Characterization (SEARCH) project. The data derived from the SRI monitoring efforts will augment data collected from other monitoring efforts. Such data collection projects include the Interagency Monitoring of Protected Visual Environments (IMPROVE) program at national park sites and the Southern Oxidant Study (SOS).

Site Name	Number of Parameters Tracked		Start Date	End Date	Number of Records ¹
North Birmingham	Continuous:	Not available.	Not available.	Not available.	Not available.
	Discrete:	Not available.			Not available.

¹ Values represent estimates of observation data only, based on data submitted to date.

c) *Inventory Summary*

Altogether, the data for this project, in consideration of any future growth of continuous data, as well as metadata associated with various records, is estimated to be over 10 million entries.

Program Name	Parameter Type	Number of Records ^{1, 2}
UORVP	Continuous:	506,880
	Discrete:	350,175
SCAMP	Continuous:	911,435
	Discrete:	21,373
Pittsburgh Supersite	Continuous:	4,335,408
	Discrete:	0
NETL Site	Continuous:	13,196
	Discrete:	0
SRI	Continuous:	0
	Discrete:	0
TOTAL RECORDS	ALL	6,138,467

¹ Values represent estimates of observation data only, based on data submitted to date.

² Values represent submitted quantities only.

Task 1.2 – Develop Data Management System Architecture

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.

The system will have detailed built-in data querying and reporting tools, and will also allow the end user to define ad hoc 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 and, if possible, in a format compatible with the NARSTO Data Management guidelines.

d) COMPUTER HARDWARE AND SOFTWARE

The project team decided early in the planning stage to utilize an Internet-connected cluster of three Windows 2000 Advanced Server computers. The primary server is located at Ohio University's (OU) Ilgard Technology Center in Athens, Ohio and another is located at *ATS's* corporate headquarters in Pittsburgh, Pennsylvania. Another is planned to be located at the Texas A&M University –Kingsville (TAMUK) campus.

Each server is running Microsoft Windows 2000 Advanced Server and Microsoft SQL Server 2000 Enterprise Edition. This software configuration provides excellent database management and development solution capabilities and is fully compatible with the GIS software planned for integration into the project in Phase II.

The Ohio University members of the project team have implemented the database on a custom built server which utilizes two Intel® Xeon™ 2.4GHz central processing units (CPU) featuring 3GB of ECC DDR Random Access Memory (RAM). The server has a dual channel RAID controller and is set up with two 18 GB system drives on one RAID channel and six 36 GB data drives on the other. It has a SONY AIT 360GB auto loading tape drive for the daily differential backups and weekly full backups to build on the failsafe implementation of the server.

A Dell PowerEdge 2600 server has been installed at *ATS's* corporate headquarters in Pittsburgh, Pennsylvania. This server has two Intel® Xeon™ 2.0GHz CPUs and has over 180GB of hard drive space available through a RAID-5 drive array of six 36GB 10,000 RPM disk drives. A Dell PowerVault 110T tape drive completes the package. This tape drive makes a differential backup of this server each Monday through Thursday evening, and a full backup on Friday evenings.

e) DATABASE TABLE STRUCTURE DESIGN

The tables are the most important part of any Relational Database Management System (RDBMS). The design of the tables needs to reduce the chances of duplicate data storage as much as possible. Therefore, great care has been exercised while designing the underlying architecture of the database.

The tables are grouped into four major categories: System Tables, Security Tables, Observation Tables and Support Tables. Each contains a distinct set of related data and each are linked to other tables by means of unique foreign key identifiers.

(1) System Tables

The System Tables Group provides management data for the application to utilize in normal housekeeping activities as well as QA/QC operations, as discussed in greater detail later in this document. The System Tables Group consists of the following tables (more details regarding each table's design can be found in **Appendix 3 – Database Table Definitions**):

tblColumnDefinitions	tblLogs	tblTableDefinitions
tblEvents	tblSystemStamps	

Design data such as table and column definitions are contained in the System Tables Group along with events, logs and timestamp data regarding all events, transactions and warnings.

The table and column definitions tables contain the underlying data structure for the database. Each column of each table is fully described in these tables to facilitate the use of a *Data Dictionary* which provides users with a detailed data map of the database which can be used in constructing special applications and data queries.

The events and log tables (`tblEvents` and `tblLogs`) contain predetermined events that are desirable to track and detail log entries that allow for user and data activity tracking, reverse posting and disaster recovery. Since each event is associated with a unique integer identifier, this identifier is used in the logs table to classify events as they are stored or read from the logs table, thus reducing wasted space by storing lengthy messages once and referring to them with the use of a unique identifier field.

The system stamps table contains a list of timestamps that are created each time specific events occur. Anytime specified transactions occur, a timestamp is recorded in this table and its unique identifier value is recorded in the foreign key identity field of the record being written or altered in the targeted table.

(2) Security Tables

The Security Tables Group contains the necessary data to control user access to restricted data and application functions. By establishing security levels on a table and column basis, a value can be assigned in the `tblTableDefinitions` and `tblColumnDefinitions` system tables that are compared with permission levels assigned to a specific user. The Security Tables Group consists of the following tables (more details regarding each table's design can be found in **Appendix 3 – Database Table Definitions**):

tblAdminLevels	tblReadLevels	tblUsers
tblBrowseLevels	tblStakeholders	tblWriteLevels
tblDataSubmitters	tblSubscriptions	

Data submitting authority and stakeholder status are restricted user lists that require approval to be added to the list of authorized users. Therefore, two tables are used to record membership requests as well as approval notifications. These tables, or lists, are used to determine who can submit data to the database and who will receive special access to Stakeholder Only sections of the application.

The users table (`tblUsers`) is a list of all registered users of the application. Most information that is unique to individual users is stored in this table. Electronic notification of news and updates are available through automated mailings and users that choose to subscribe to these periodic mailings are listed in the subscriptions table (`tblSubscriptions`).

(3) Observation Tables

The Observation Tables Group (OTG) contains all measured data and accompanying metadata submitted to the database. The data values themselves are organized in one table and categorized according to different criteria. The Observation Tables consist of the following tables (more details regarding each table’s design can be found in **Appendix 3 – Database Table Definitions**):

<code>tblCollectionPrinciples</code>	<code>tblMethods</code>	<code>tblParameterProperties</code>
<code>tblDataClasses</code>	<code>tblNARSTOMetaflags</code>	<code>tblParameters</code>
<code>tblFilters</code>	<code>tblNetworks</code>	<code>tblQCStatus</code>
<code>tblInstrumentLocations</code>	<code>tblObservationFlags</code>	<code>tblSampleDurations</code>
<code>tblInstruments</code>	<code>tblObservations</code>	<code>tblSampleFrequencies</code>
<code>tblMetaflags</code>	<code>tblParameterIdentifiers</code>	<code>tblSites</code>

Observational data is first assigned a *Data Classification*. A data classification might be ‘Meteorological’ or ‘SGS Speciation.’ A master table of classifications holds all possible classification abbreviations and descriptions and is referenced by a unique primary key value.

Next, a series of table entries is used to specifically describe the data collection process. These tables are the *parameters*, *instrument location*, *parameter identifiers*, *collection principles* and the *parameter property* tables. Each table contains a single element of the collection process description. For example, let us assume that data collected from a site is classified as a ‘Gas’ with the parameter ‘Sulfur’ being monitored. The data is collected from an instrument that is at ground level and is stationary. The parameter identifier would be ‘Sulfur’. The data was collected using the collection principle ‘Infrared Absorption’ and the parameter property is ‘Total Sulfur.’

The following table depicts this example’s dataset and shows how data is separated and identified according to values, parameters and properties.

Example:

DATA CLASS	PARAMETER DESCRIPTION	INSTRUMENT LOCATION DESCRIPTION	PARAMETER IDENTIFIER DESCRIPTION	COLLECTION PRINCIPLE DESCRIPTION	PARAMETER PROPERTY
Particles	PM _{2.5} Weight	Surface-Mobile	PM _{2.5}	TEOM	PM _{2.5} Weight
Gas	Sulfur	Surface-Stationary	Sulfur	Infrared Absorption	Total Sulfur
Metrology	WD-Min	Aloft-tower	WD	Wind Direction Vane	UCOMP

Each record will also have a local flag which is associated with a standardized NARSTO flag. The local flags have been developed by the agencies and organizations responsible for collecting and submitting the datasets. Localized flag systems provide site-level details that may be unique to the collection site. Since each site can potentially have a unique set of flags, the data from multiple sites cannot be adequately analyzed unless each set of localized flags is mapped to a *common* flag set. The NARSTO flag set has been chosen as the *common* flag set to reduce the time required to convert datasets between the format required for data submission to this project and the standard NARSTO data submission format.

Some data warrant special tracking tables. Examples of such data include FRM, SFS, SGS and other filter identification data, instrument identification data and frequency and duration data.

The OTG provides excellent data input and retrieval capabilities to the project. This application utilizes these tables, in conjunction with the System, Security and Support Table Groups, to organize and collate the collected data into datasets provided by XML Web Services data providers located on the database and web servers.

(4) Support Tables

Ancillary, or support tables are part of the Support Tables Group (STG). This group of tables provides supplemental or support data to the application. Data items such as addresses and city, county, state and country codes are contained in this group of tables. Time zone and unit conversion factors, codes and descriptions are also part of the Support Tables Group and provide useful conversion data to view data in common date/time zones as well as providing support for both US and metric units. The Support Tables consist of the following tables (more details regarding each table's design can be found in **Appendix 3 – Database Table Definitions**):

tblCities	tblStates	tblUnits
tblCounties	tblThreadedNotes	tblVCards
tblCountries	tblTimeZones	

Addresses are stored as virtual, or 'V' cards. Cities, counties, states and countries are stored in individual tables, and foreign key relationships link them to each other and the V-Cards table.

A time zone table keeps track of all of the various time zones with a Greenwich Mean Time (GMT) offset value used to view datasets across multiple time zones in a common date/time system. By using this offset value, these values are easily converted to GMT and local time, as required.

The units table keeps track of various units, their abbreviations and conversion factors so that values can easily be viewed in both metric and English units. Users will be able to select data to be in one format or the other by default, or convert specific units individually.

f) DATABASE TABLE RELATIONSHIPS

Data tables, as stated earlier, are the most important parts of any RDBMS. This is because the tables are linked to each other by unique identification fields, or Primary and Foreign Keys (PK, FK). This linkage is called a relationship and can represent different types of relationships, such as many-to-one, one-to-many, one-to-one, etc.

By storing data in this format, space is conserved by not having to repeat data throughout a table. If data is collected from one site, it is much more efficient to store the site name, location, etc. in one table, along with a unique site identification number, and then only store the site identification number, instead of the name, location, etc. with each record collected from each site.

This database application has several key relationships and every table has a unique primary key field. Records can be easily located by using this numeric field in combination with database indexes, queries and stored procedures, but care must be taken when inserting and updating records, because these relationships create *forced* dependencies that require certain data be present in linked tables before a record link can be established.

As the datasets grow, a periodic check is required to ensure that duplicate records are not created for referenced items. This is especially important in regards to any observation data, since multiple entries for the same items in the parameter tables would cause incomplete datasets when requesting data based on any of these parameters. A complete set of diagrams can be found in **Appendix 6 – Database Relationship Diagrams**.

g) DATA INPUT FILE FORMATTING SPECIFICATIONS

The large variance in data file formats has necessitated the development of an input file format specification. This specification is designed to help maintain the data integrity of the database and ensure that quality data results are available to all stakeholders, researchers and anyone else using this application to retrieve particulate matter data.

There are two primary formats available to use when creating input files. A modified NARSTO format allows data submitters to quickly convert files to one of the formats required by this application from an existing file already formatted according to NARSTO specifications. For the purposes of this application, we will call this modified NARSTO format the *tabular* format.

Each input file must include a header section that contains several required items. This data is primarily stored in the log table and other data fields throughout the database, and is used to display or report metadata fields relating to a specific sites or data items. These field names are listed in Appendix 7.

(1) *Tabular Format*

The tabular format specification requires all files include “*beginning of file*” and “*end of file*” tags that are contained within brackets. Each file submitted in the tabular format must begin with the tag [bof] and end with the tag [eof].

Each section of a transfer file must also include “*begin*” and “*end*” tags. The header section is preceded by the [begin header] tag and is closed with the [end header] tag. Between these two tags, header fields and values are arranged in comma separated, pairs utilizing a new line for each pair and quotation marks to enclose each value item. For example, the header field COMMENT would appear in the transfer file as follows:

Example:

```
...  
COMMENT, "This is a comment regarding this particular input file."  
...
```

All other sections of a transfer files begin with the [begin table] tag and end with the [end table] tag. Each table section of the transfer file must contain some basic information that adequately describes the data being submitted. This information includes items such as column names, references to other tables in the same transfer file and quality control flags and comments for *each* data item contained in *each* table. Column names must be defined within the database prior to use in the submission files, however, and a complete list of available column names is contained in **Appendix 3 – Database Table Definitions**. Only data from the Observation and Support Tables Groups can be submitted in this format. All other data requires Administrative permissions to enter, change or remove.

A table section within the transfer file must contain a table name and a list of column names. These are also arranged in a comma delimited format. The first line within the table section consists of the keyword `table_Name`, a comma and the name of the table, enclosed within quotation marks.

The first position of the next line will contain the keyword or field name `column_Names`, followed by a list of comma separated values that indicate to which column a corresponding data item belongs.

Line three of a table section contains a list of reference table names. If items in a particular column are extracted from a particular source table within the transfer file, then the name of that corresponding table is required in the same column position as the column name in the row above. For example, a transfer file might contain a table devoted to collection site location information.

Every other line, except for the final one, of a table section begins with the keyword `data_Row` and all data values for that row of the table, separated by commas. This tells the file parsing application to treat all values after this keyword as a data item corresponding to the column name listed in the same column position in line two of the same table section. The example below should clarify these requirements. A complete list of all required tags and keywords is included in **Appendix 7 – Data**

Input File Requirements.

Example:

In this example, Table One contains a list of all sites, with each site being identified by a unique numeric identifier. Table Two consists of a set of values that utilizes the referenced site table to identify a particular site.

When viewing this data in a spreadsheet, it would look similar to the following two tables:

Table One

table Name	"sites"	
column Names	"site Name"	"site ID"
reference Table	"N/A"	"N/A"
values	"Holbrook"	1
values	"Lawrenceville"	2

Table Two

table Name	"FRM values"	
column Names	"site ID"	"value"
reference Table	"sites"	"N/A"
values	1	2.886
values	1	3.001

...

```
[begin table]
table_Name, "sites"
column_Names, "site_Name", "site_ID"
reference_Table, "N/A", "N/A"
data_Row, "Holbrook", 1
data_Row, "Lawrenceville", 2
[end table]
[begin table]
table_Name, "FRM values"
column_Names, "site_ID", "value"
reference_Table, 1, 2.886
data_Row, 1, 2.886
data_Row, 1, 3.001
[end table]
...
```

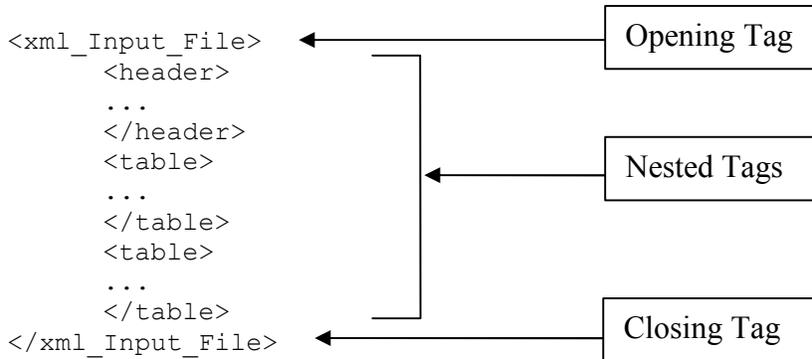
(2) XML Format

The other format acceptable for submitting data is an XML-style file format. XML, or Extensible Markup Language, is a very broad file format definition. XML files are *text* files that adhere to a specific *schema*. There are several versions of XML schema, which are essentially an interpretation map or translation reference that defines the keywords and tags required for this application.

The XML files submitted to the database server for inclusion in the database tables must meet the same component requirements as the tabular formatted data files. That is to say that both XML and tabular formatted documents require *header* and *table* sections. XML formatted files differ from the tabular formatted files primarily in the type, number of and positioning of special *tags* throughout the document. Each element within an XML formatted data input file requires both an *opening* and *closing* tag. Closing tags utilize the same keyword as the corresponding opening tag, but the keyword is preceded by a forward slash character (/).

XML data files require an opening and closing tag for each element. These can be *nested* tags to associate parent and child elements with each other. Below is an example of a nested set of XML tags. All required XML tags are listed in **Appendix 7 – Data Input File Requirements**, along with examples of properly formatted files.

Example:



Task 1.3 – Population of Database

During the database development period, datasets cannot be fully loaded into the database since the structure is likely to change drastically at any given moment. Therefore, an initial set of data files was obtained from each submitting contractor or organization. These contractors and organizations included NETL, Carnegie Mellon, CONSOL and *ATS*.

These initial datasets were used to establish an inventory of the data being managed and to develop data submission methods and procedures to automatically check new datasets for quality and then upload the datasets to the database tables.

A series of checks and balances are being integrated into this system to ensure that only authorized individuals are able to submit new datasets or change existing ones. Usernames and passwords, in combination with a data submitter registration and identification process, prevent unauthorized uploads of datasets and preserves this project’s data integrity. Users login through a login screen and then the User Authentication Module (UAM) verifies the user’s username and password against stored data in the database. Users can then upload their data, and the Automated Population Module (APM), a server-only application, detects changes to the upload target directory’s contents and processes the data, provided it is in the correct format, and sends it to the database.

h) INITIAL DATASETS

The datasets submitted to the project team varied in format. Data submitted by Carnegie Mellon, for example, were submitted in NARSTO format, while data submitted by CONSOL were in a Microsoft Access database.

To preserve database integrity in an automated data submission environment, a file format standard has been developed to ensure users have access to quality data that is easily manageable

by the hosting organization. Before any data can be submitted through the Automated Population Module of this application, the submitting user and organization must first register, login and have their username and password authenticated by the User Authentication Module.

i) USER AUTHENTICATION AND AUTHORIZATION MODULE

The User Authentication Module (UAM) handles all validation tasks for accessing the data and restricted web site sections. The usernames and passwords provide simple authentication tasks, as well as perform extensive and complex authorization tasks. Since no direct interaction occurs between the user and the database server and server file system, system usernames and passwords do not need to be assigned and managed.

The web application must rely on username and password combinations to identify, or *authenticate* a particular user. Since the most crucial operations involve uploading data to the database, care must be taken to ensure that only users who are authorized gain access to the data upload utilities. To accomplish this, the user must first obtain a username and password. For security reasons, usernames are required to be a minimum of six alpha-numeric characters (a-z and 0-9) with no spaces between characters. Passwords are required to be a minimum of eight (8) characters, using any standard character available from a standard 104-key keyboard.

When the user first loads the application in their standards-compliant (4.0) web browser (Internet Explorer 5.0 or later, Netscape 5.0 or later, etc.), the main entry page is loaded. From this point, the user clicks on a link to log in. At the login prompt, the user enters their username and password combination. The UAM receives the password by way of a Secure Sockets Layer (SSL) connection and then queries the database to retrieve an encrypted version of the password from the database to compare with the encrypted version received from the user's login response.

If the user enters the correct password for the given username, then the user proceeds to the file upload section of the application. If not, then the user's browser is redirected to the login prompt again with a message that the password entered does not match. The Application Administrator can control the number of times a user can enter incorrect passwords within a specific time period to prevent *brute force* server attacks by automated scripts.

j) AUTOMATED POPULATION MODULE

The Automated Population Module (APM) is a Windows Service application that runs continuously and maintains watch on the upload target folder for changes. As new files are added, a virus scan is conducted on the inbound files, then the APM checks to see if the file name is formatted correctly.

If the file name adheres to the data file naming specification and the file contains ASCII encoded text and the file begins with a properly formatted header, then the data contained in the file is uploaded to the primary database, the file is transferred to an archive directory, the transaction is logged into the event and log tables and the process begins all over again with the APM becoming idle while waiting for another file to be uploaded to the server.

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 Hyper Text Transfer Protocol (HTTP). Each has distinct advantages and disadvantages. Users build custom datasets of the data they desire, using an online ad hoc query system. The query builder utilizes XML datasets that display the data in grids, or tables. Each dataset is available for download via HTTP hyperlinks, or can be queued in a download directory for multiple file transfers. To queue a file for later download, the user must have the appropriate permissions assigned to them by the Application Administrator.

a) FTP DATA FILE RETRIEVAL

Users can retrieve static data transfer files directly from the FTP server. Registration is only required for data uploads. For downloads only, anonymous FTP access will be granted to allow users to connect and download these datasets. FTP file transfers are easier to utilize when retrieving multiple files, but the user's software must be configured correctly. Specifically, the user's FTP client application must utilize a "Passive" connection. Users will find detailed information regarding this and other settings for their client software within the help sections of the FTP client application.

There are several programs that act as FTP clients when connected to an FTP server, including Internet Explorer and Netscape. There are also several specialty FTP client applications that offer much greater flexibility to the end user if they are working with multiple FTP sites and folders. These applications offer more features, but are not required for use with this application.

b) HTTP DATA FILE RETRIEVAL

If the user cannot connect via FTP, the static datasets will be available via Hyper Text Transfer Protocol (HTTP). HTTP file transfers provide file size information when the client begins downloading a file, so that the estimated remaining time can easily be calculated using the file size information provided by the server and the bandwidth usage information available to the web browser's underlying engine. The primary limitation of using this method is that only one file can be selected per download.

c) AD HOC QUERY SYSTEMS

(1) Query 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 query interface should provide a logical flow in the selection process so that the user can easily select the subset of the database that the user wishes to view. The query interface also interacts with geographic information datasets to provide the user with the additional option of selection based on geographical location in the Internet Mapping System (IMS).

(2) GIS Integration

The project team has defined two possible avenues for integrating the geographic information datasets with air quality datasets in a web-based data retrieval system. The first involves using the geographic information datasets to produce static maps with air quality monitoring sites stored as html hyperlinks to separate web pages. Each linked webpage would contain a detailed map of the site location and additional links for site description and data retrieval. Advantages of using this method include: ease of use, low bandwidth requirements, and very little customized programming outside of html. Disadvantages include: inability to display query results on a map, perpetual generation of new static maps to reflect any changes in data layers, and GIS data cannot be incorporated into other mapping packages via internet connection without downloading each data layer.

The second retrieval system utilizes the geographic information datasets in a much more robust manner but requires Internet mapping software on the server. The client web browser requires no specialized software or plug-in. This method would allow the user to completely interact with geographic features within the project area and visualize spatial relationships between air quality query results and other geographic features. Identification of any feature within any layer can be performed and customized maps can be printed based on the user's specifications. Additional functionality includes basic GIS software capabilities such as measuring distances, creating buffers, feature on feature selection, panning, zooming, turning layers on and off, and manual selection of multiple monitoring sites for data query purposes. Professionals wishing to perform further spatial analysis on air quality data will be able to directly connect to the geographic information datasets, or download all GIS layers in a given view extent with the single click of a button. For example, a researcher would be able to connect to the online data management application, download summarized air data, interpolate the data into a surface using their own software package, and instantly overlay this data with all GIS layers from the GIS server in addition to adding any of their own layers. Disadvantages of the Internet Mapping System (IMS) include high bandwidth requirements at the client-end, increased programming needs at the server-end and some level of familiarity with GIS software by the end-user (which will be lessened slightly by a detailed help module being developed in Phase Two of this project). As advancements continue within future IMS software releases some of these disadvantages will be mitigated.

A separate GIS database to be used for web-based mapping and spatial analysis of air quality data for the Upper Ohio River Valley (UORV) has been constructed. Data has been gathered for a large region, approximately all land east of the Mississippi river, in order to completely encompass the project area and enable visualization of dispersion and other models, which may go beyond the extent of the UORV. Geographically referenced data layers composing the database include, but are not limited to, roads, populated places, political boundaries, digital elevation datasets, electric utilities, and air quality and weather monitoring stations. Thus far, the data has been used in the generation of digital maps for illustrative purposes and used in multiple "test" Internet mapping websites.

(3) Tabular Results

The tabular query result interface provides an option for the user to view the results of his query in a tabular page-wise format before downloading the resulting dataset to their personal computer. 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 database. The user can browse through the dataset and has the option of reordering the display by clicking on the header of the table columns.

Each result set can alternatively be downloaded for use as an “*offline*” dataset by clicking on a hyperlink placed above and below the results table. An archive would 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.

Each submitting authority is responsible for flagging data according to locally established flag designations or NARSTO flag designations. If the submitting authority chooses to utilize a local system of flags, each local flag must be mapped to a corresponding NARSTO flag. This ensures that a common set of flags are available when comparing data submitted by multiple submitting authorities.

As previously discussed, the APM requires that data be submitted in a prescribed format, or standard. This ensures that each value or flag is positioned in the correct location within 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, other than 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)

A basic design framework has been developed to host a data retrieval utility, using XML Web Services. This allows the developers to make changes to the visual interface for the application without affecting the database business logic code that updates and retrieves datasets for users.

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

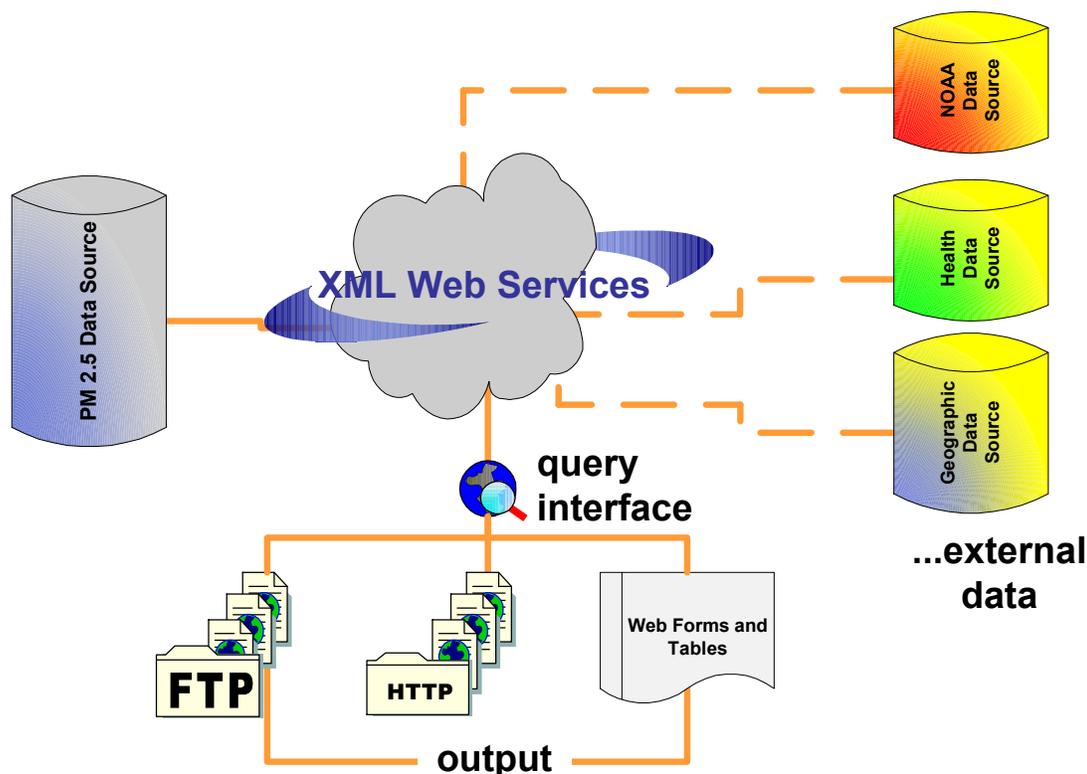


Figure One

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.

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%	75%
1.2	Develop Data Management System Architecture	100%	55%
1.3	Population of Database	55%	55%
1.4	Develop Web-based Retrieval System	36%	10%
1.5	Develop Internal QA/QC System	14%	14%

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%	0%
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. CONCLUSIONS

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

APPENDIX 1 – DATA FILE INVENTORY LISTS

C. UORVP Data Files

DRI Data Files			
NAME	TYPE	SIZE	MODIFIED
dri_frm_01.xls	Microsoft Excel	18 KB	Thu 10/24/2002 2:18 PM
dri_sfs_01.xls	Microsoft Excel	275 KB	Wed 10/23/2002 2:50 PM
dri_sfs_sgs_02.xls	Microsoft Excel	674 KB	Wed 10/23/2002 3:35 PM
DRIFilterPackIDs.1.xls	Microsoft Excel	15 KB	Mon 11/25/2002 11:36 AM
FileNamesDRIAnalyses.2.xls	Microsoft Excel	14 KB	Mon 11/25/2002 11:36 AM
Ohcon01s.rev.xls	Microsoft Excel	60 KB	Wed 6/20/2001 9:32 AM
ohcon02.xls	Microsoft Excel	47 KB	Tue 11/21/2000 2:15 PM
OHCON04F.CC.xls	Microsoft Excel	18 KB	Wed 10/23/2002 3:36 PM
OHCON04S.xls	Microsoft Excel	155 KB	Wed 10/23/2002 3:40 PM
OHCON05X.xls	Microsoft Excel	195 KB	Fri 9/22/2000 6:18 PM
ohcon07x.xls	Microsoft Excel	207 KB	Fri 9/28/2001 11:36 AM
UORVP06.xls	Microsoft Excel	179 KB	Thu 10/24/2002 2:17 PM
UORVP Mass data Jan2001 through Aug2001.xls	Microsoft Excel	192 KB	Wed 1/8/2003 9:23 AM
UORVPSPEC1.xls	Microsoft Excel	364 KB	Thu 10/24/2002 2:13 PM
UORVPSPEC1_columndefs.csv	Comma Separated Values File	8 KB	Tue 10/1/2002 3:53 PM

LABCO Data Files			
NAME	TYPE	SIZE	MODIFIED
00-005.xls	Microsoft Excel	23 KB	Mon 1/24/2000 4:06 PM
00-011.xls	Microsoft Excel	28 KB	Thu 3/2/2000 7:33 AM
00-029.xls	Microsoft Excel	28 KB	Thu 10/25/2001 6:38 PM
00-057.xls	Microsoft Excel	28 KB	Thu 10/25/2001 6:38 PM
00-070.xls	Microsoft Excel	21 KB	Thu 10/25/2001 6:38 PM
00-099.xls	Microsoft Excel	22 KB	Wed 7/26/2000 2:03 PM
00-117.xls	Microsoft Excel	28 KB	Thu 6/29/2000 8:59 AM
00-169.xls	Microsoft Excel	29 KB	Wed 8/16/2000 3:10 PM
00-194.xls	Microsoft Excel	24 KB	Tue 8/22/2000 1:34 PM
00-239.xls	Microsoft Excel	30 KB	Thu 10/25/2001 6:38 PM
00-291.xls	Microsoft Excel	33 KB	Thu 11/30/2000 3:48 PM
00-328.xls	Microsoft Excel	25 KB	Tue 1/9/2001 5:34 PM
01-017.xls	Microsoft Excel	27 KB	Wed 1/8/2003 10:26 AM
01-033.xls	Microsoft Excel	23 KB	Wed 1/8/2003 10:27 AM
01-074.xls	Microsoft Excel	27 KB	Wed 1/8/2003 10:27 AM
01-108.xls	Microsoft Excel	22 KB	Wed 1/8/2003 10:27 AM
01-142.xls	Microsoft Excel	23 KB	Thu 10/25/2001 6:38 PM
01-167.xls	Microsoft Excel	24 KB	Wed 1/8/2003 10:11 AM
01-189.xls	Microsoft Excel	42 KB	Fri 1/10/2003 10:10 AM
01-227 Complete.xls	Microsoft Excel	29 KB	Wed 1/8/2003 10:12 AM
01-227 No Carbon.xls	Microsoft Excel	22 KB	Thu 1/9/2003 8:33 AM
99-111.xls	Microsoft Excel	27 KB	Wed 1/8/2003 10:26 AM
99-116.XLS	Microsoft Excel	27 KB	Thu 3/2/2000 7:25 AM

99-140.XLS	Microsoft Excel	25 KB	Tue 8/31/1999 1:22 PM
99-153.XLS	Microsoft Excel	22 KB	Tue 9/14/1999 2:38 PM
99-179.xls	Microsoft Excel	25 KB	Wed 3/1/2000 4:03 PM
99-204.xls	Microsoft Excel	29 KB	Fri 11/19/1999 3:43 PM
99-223.xls	Microsoft Excel	26 KB	Fri 12/17/1999 2:43 PM
tm-labnet-1page-1011991.tif	TIF Image	13 KB	Mon 10/11/1999 9:07 AM

MET Data Files			
NAME	TYPE	SIZE	MODIFIED
HW00R-apr-0504.met	MET File	80 KB	Thu 5/4/2000 3:04 PM
HW00R-aug-0901.met	MET File	83 KB	Fri 9/1/2000 10:33 AM
HW00R-jul-0802.met	MET File	83 KB	Wed 8/2/2000 12:50 PM
HW00R-jun-0707.met	MET File	80 KB	Fri 7/7/2000 3:08 PM
HW00R-may-0607.met	MET File	83 KB	Wed 6/7/2000 2:41 PM
HW00R-sep-1002.met	MET File	80 KB	Mon 10/2/2000 11:11 AM
HWR-apr-050102.met	MET File	80 KB	Wed 5/1/2002 2:59 PM
HWR-apr-050201.met	MET File	80 KB	Tue 5/1/2001 9:47 PM
HWR-aug-090401.met	MET File	83 KB	Tue 9/4/2001 12:20 PM
HWR-dec-010802.met	MET File	83 KB	Tue 1/8/2002 11:46 AM
HWR-dec-122800.met	MET File	83 KB	Thu 12/28/2000 10:04 AM
HWR-feb-030502.met	MET File	75 KB	Tue 3/5/2002 11:15 AM
HWR-feb-031201.met	MET File	75 KB	Mon 3/12/2001 8:27 AM
HWR-jan-020101.met	MET File	83 KB	Thu 2/1/2001 10:57 AM
HWR-jul-080101.met	MET File	83 KB	Wed 8/1/2001 4:43 PM
HWR-jul-080102.met	MET File	83 KB	Thu 8/1/2002 5:57 PM
HWR-jun-070201.met	MET File	80 KB	Tue 7/3/2001 3:21 AM

HWR-jun-070302.met	MET File	66 KB	Wed 7/3/2002 3:53 PM
HWR-mar-040102.met	MET File	83 KB	Mon 4/1/2002 4:32 PM
HWR-may-060101.met	MET File	83 KB	Fri 6/1/2001 3:18 AM
HWR-may-060302.met	MET File	83 KB	Mon 6/3/2002 11:51 AM
HWR-nov-120400.met	MET File	80 KB	Mon 12/4/2000 8:27 AM
HWR-nov-120701.met	MET File	101 KB	Fri 12/7/2001 10:39 AM
HWR-oct-110701.met	MET File	83 KB	Wed 11/7/2001 11:43 AM
HWR-oct-111300.met	MET File	83 KB	Mon 11/13/2000 10:55 AM
HWR-sep-100101.met	MET File	80 KB	Mon 10/1/2001 6:11 PM
LW00R-apr-0504.met	MET File	101 KB	Thu 5/4/2000 11:36 AM
LW00R-aug-0911.met	MET File	105 KB	Mon 9/11/2000 3:12 PM
LW00R-feb-0306.met	MET File	98 KB	Mon 3/6/2000 1:40 PM
LW00R-jan-0201.met	MET File	90 KB	Tue 2/1/2000 3:21 PM
LW00R-jul-0801.met	MET File	105 KB	Tue 8/1/2000 3:15 PM
LW00R-jun-0707.met	MET File	101 KB	Fri 7/7/2000 11:20 AM
LW00R-mar-0404.met	MET File	105 KB	Tue 4/4/2000 12:13 PM
LW99R-dec-0121.met	MET File	90 KB	Fri 1/21/2000 2:18 PM
LW99R-nov-1213.met	MET File	87 KB	Mon 12/13/1999 11:11 AM
LWH-jul-100499.met	MET File	90 KB	Mon 10/4/1999 4:59 PM
LWR-apr-050101.met	MET File	108 KB	Tue 5/1/2001 2:16 PM
LWR-aug-090401.met	MET File	112 KB	Tue 9/4/2001 10:34 AM
LWR-dec-010201.met	MET File	105 KB	Tue 1/2/2001 2:08 PM
LWR-feb-030101.met	MET File	101 KB	Thu 3/1/2001 12:44 PM
LWR-feb-030102.met	MET File	101 KB	Fri 3/1/2002 9:34 AM
LWR-jan-020101.met	MET File	105 KB	Thu 2/1/2001 11:10 AM
LWR-jan-020102.met	MET File	112 KB	Fri 2/1/2002 9:42 AM

LWR-jul-080101.met	MET File	112 KB	Wed 8/1/2001 11:14 AM
LWR-jul-080102.met	MET File	112 KB	Thu 8/1/2002 12:44 PM
LWR-jun-070201.met	MET File	108 KB	Mon 7/2/2001 10:15 AM
LWR-mar-040102.met	MET File	112 KB	Mon 4/1/2002 10:56 AM
LWR-may-060101.met	MET File	112 KB	Fri 6/1/2001 11:01 AM
LWR-nov-120400.met	MET File	101 KB	Mon 12/4/2000 12:21 PM
LWR-oct-110600.met	MET File	105 KB	Mon 11/6/2000 9:17 AM
LWR-sep-100101.met	MET File	108 KB	Mon 10/1/2001 11:16 AM

Report Data Files			
NAME	TYPE	SIZE	MODIFIED
all_teom-7-1-2002 and 7-31-2002.xls	Microsoft Excel	2 MB	Fri 10/18/2002 5:01 PM
lw_2.5_teom-7-1-2002 and 7-31-2002.xls	Microsoft Excel	1 MB	Fri 10/18/2002 5:05 PM

TEOM 10 Data Files			
NAME	TYPE	SIZE	MODIFIED
LT10R.013101-030101.tom	TOM File	156 KB	Thu 3/1/2001 12:41 PM
LT10R.033100-050400.tom	TOM File	181 KB	Thu 5/4/2000 11:12 AM
LT10R.033101-050101.tom	TOM File	242 KB	Tue 5/1/2001 2:46 PM
LT10R.043000-060100.tom	TOM File	171 KB	Thu 6/1/2000 5:33 PM
LT10R.053100-070700.tom	TOM File	194 KB	Fri 7/7/2000 11:16 AM
LT10R.060702-070202.tom	TOM File	173 KB	Tue 7/2/2002 9:46 AM
LT10R.061401-070201.tom	TOM File	140 KB	Mon 7/2/2001 10:12 AM
LT10R.063000-080100.tom	TOM File	170 KB	Tue 8/1/2000 3:00 PM
LT10R.063001-080101.tom	TOM File	255 KB	Wed 8/1/2001 11:11 AM
LT10R.063002-080102.tom	TOM File	255 KB	Thu 8/1/2002 12:35 PM
LT10R.083100-100300.tom	TOM File	234 KB	Tue 10/3/2000 2:47 PM

LT10R.083101-100101.tom	TOM File	247 KB	Mon 10/1/2001 12:22 PM
LT10R.103100-120400.tom	TOM File	184 KB	Mon 12/4/2000 12:18 PM
LT10R.113000-010201.tom	TOM File	176 KB	Tue 1/2/2001 1:58 PM
LT10R.123100-020501.tom	TOM File	194 KB	Mon 2/5/2001 8:56 AM
LT10R.123101-020102.tom	TOM File	255 KB	Fri 2/1/2002 9:39 AM

TEOM 2.5 Data Files			
NAME	TYPE	SIZE	MODIFIED
HT25R.013101-030101.tom	TOM File	152 KB	Thu 3/1/2001 8:34 AM
HT25R.013102-030502.tom	TOM File	263 KB	Tue 3/5/2002 10:13 AM
HT25R.033100-050400.tom	TOM File	184 KB	Thu 5/4/2000 2:59 PM
HT25R.033101-050201.tom	TOM File	254 KB	Tue 5/1/2001 9:44 PM
HT25R.033102-050102.tom	TOM File	246 KB	Wed 5/1/2002 2:42 PM
HT25R.042902-060302.tom	TOM File	271 KB	Thu 10/17/2002 9:17 AM
HT25R.043000-060200.tom	TOM File	179 KB	Fri 6/2/2000 5:24 PM
HT25R.043001-060101.tom	TOM File	250 KB	Fri 6/1/2001 3:34 AM
HT25R.053100-062800.tom	TOM File	164 KB	Wed 6/28/2000 5:36 PM
HT25R.053101-070201.tom	TOM File	254 KB	Tue 7/3/2001 3:17 AM
HT25R.053102-070302.tom	TOM File	263 KB	Wed 7/3/2002 3:46 PM
HT25R.060100-070700.tom	TOM File	196 KB	Fri 7/7/2000 3:05 PM
HT25R.063000-080100.tom	TOM File	169 KB	Tue 8/1/2000 12:28 PM
HT25R.063001-080101.tom	TOM File	254 KB	Wed 8/1/2001 4:40 PM
HT25R.063002-080102.tom	TOM File	255 KB	Thu 8/1/2002 5:54 PM
HT25R.073001-090401.tom	TOM File	283 KB	Tue 9/4/2001 12:19 PM
HT25R.073100-090100.tom	TOM File	169 KB	Fri 9/1/2000 10:30 AM
HT25R.083100-100200.tom	TOM File	168 KB	Mon 10/2/2000 11:03 AM

HT25R.083101-100101.tom	TOM File	248 KB	Mon 10/1/2001 6:03 PM
HT25R.103100-120400.tom	TOM File	180 KB	Mon 12/4/2000 8:21 AM
HT25R.113000-010201.tom	TOM File	169 KB	Tue 1/2/2001 8:30 AM
HT25R.122900-020101.tom	TOM File	184 KB	Thu 2/1/2001 10:45 AM
LT25R.013101-030101.tom	TOM File	206 KB	Thu 3/1/2001 12:34 PM
LT25R.033100-050400.tom	TOM File	242 KB	Thu 5/4/2000 11:24 AM
LT25R.033101-050101.tom	TOM File	242 KB	Tue 5/1/2001 2:32 PM
LT25R.043000-060100.tom	TOM File	225 KB	Thu 6/1/2000 5:25 PM
LT25R.043002-060302.tom	TOM File	260 KB	Mon 6/3/2002 10:33 AM
LT25R.053100-070700.tom	TOM File	256 KB	Fri 7/7/2000 11:07 AM
LT25R.053101-070201.tom	TOM File	255 KB	Mon 7/2/2001 10:03 AM
LT25R.063000-080100.tom	TOM File	224 KB	Tue 8/1/2000 3:09 PM
LT25R.063001-080101.tom	TOM File	255 KB	Wed 8/1/2001 10:59 AM
LT25R.063002-080102.tom	TOM File	255 KB	Thu 8/1/2002 12:21 PM
LT25R.083100-100300.tom	TOM File	174 KB	Tue 10/3/2000 12:55 PM
LT25R.083101-100101.tom	TOM File	246 KB	Mon 10/1/2001 12:08 PM
LT25R.103100-120400.tom	TOM File	242 KB	Mon 12/4/2000 12:07 PM
LT25R.113000-010201.tom	TOM File	235 KB	Tue 1/2/2001 1:48 PM
LT25R.123101-020501.tom	TOM File	255 KB	Mon 2/5/2001 8:46 AM
LT25R.123102-020102.tom	TOM File	255 KB	Fri 2/1/2002 9:26 AM

D. Pittsburgh Supersite Data Files

NAME	TYPE	SIZE	MODIFIED
PAQS_SURF-CONC2-DATA_SCHP-20010701-31_all_V1.4.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20010801-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM

PAQS_SURF-CONC2-DATA_SCHP-20010901-30_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20011001-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20011101-30_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20011201-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020101-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020201-28_all_V1.1.xls	Microsoft Excel	1 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020301-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020401-30_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020501-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020601-30_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020701-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC2-DATA_SCHP-20020801-31_all_V1.1.xls	Microsoft Excel	2 MB	Fri 11/15/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20010701-31_all_V1.2.xls	Microsoft Excel	2 MB	Tue 11/12/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20010801-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20010901-30_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20011001-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20011101-30_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20011201-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM

PAQS_SURF-CONC-DATA_SCHP-20020101-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020201-28_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020301-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020401-30_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020501-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020601-30_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020701-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020801-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20020901-30_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-CONC-DATA_SCHP-20021001-31_all_V1.1.xls	Microsoft Excel	2 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20010701-31_all_V1.2.xls	Microsoft Excel	2 MB	Tue 11/12/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20010801-31_all.xls	Microsoft Excel	3 MB	Thu 9/26/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20010901-30_all.xls	Microsoft Excel	3 MB	Thu 9/26/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20011001-31_all.xls	Microsoft Excel	4 MB	Thu 9/26/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20011101-31_all.xls	Microsoft Excel	3 MB	Thu 9/26/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20011201-31_all.xls	Microsoft Excel	4 MB	Thu 9/26/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020101-31_all.xls	Microsoft Excel	4 MB	Thu 9/26/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020201-28_all.xls	Microsoft Excel	3 MB	Thu 9/26/2002 12:00 AM

PAQS_SURF-MET-DATA_SCHP-20020301-31_all_V1.1.xls	Microsoft Excel	4 MB	Mon 10/28/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020401-30_all.xls	Microsoft Excel	4 MB	Fri 9/27/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020501-31_all.xls	Microsoft Excel	4 MB	Fri 9/27/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020601-30_all.xls	Microsoft Excel	4 MB	Fri 9/27/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020701-31_all_V1.1.xls	Microsoft Excel	4 MB	Fri 9/27/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020801-31_all.xls	Microsoft Excel	4 MB	Tue 10/22/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20020901-30_all.xls	Microsoft Excel	4 MB	Tue 10/22/2002 12:00 AM
PAQS_SURF-MET-DATA_SCHP-20021001-31_all.xls	Microsoft Excel	4 MB	Mon 10/28/2002 12:00 AM

E. NETL Site Data Files

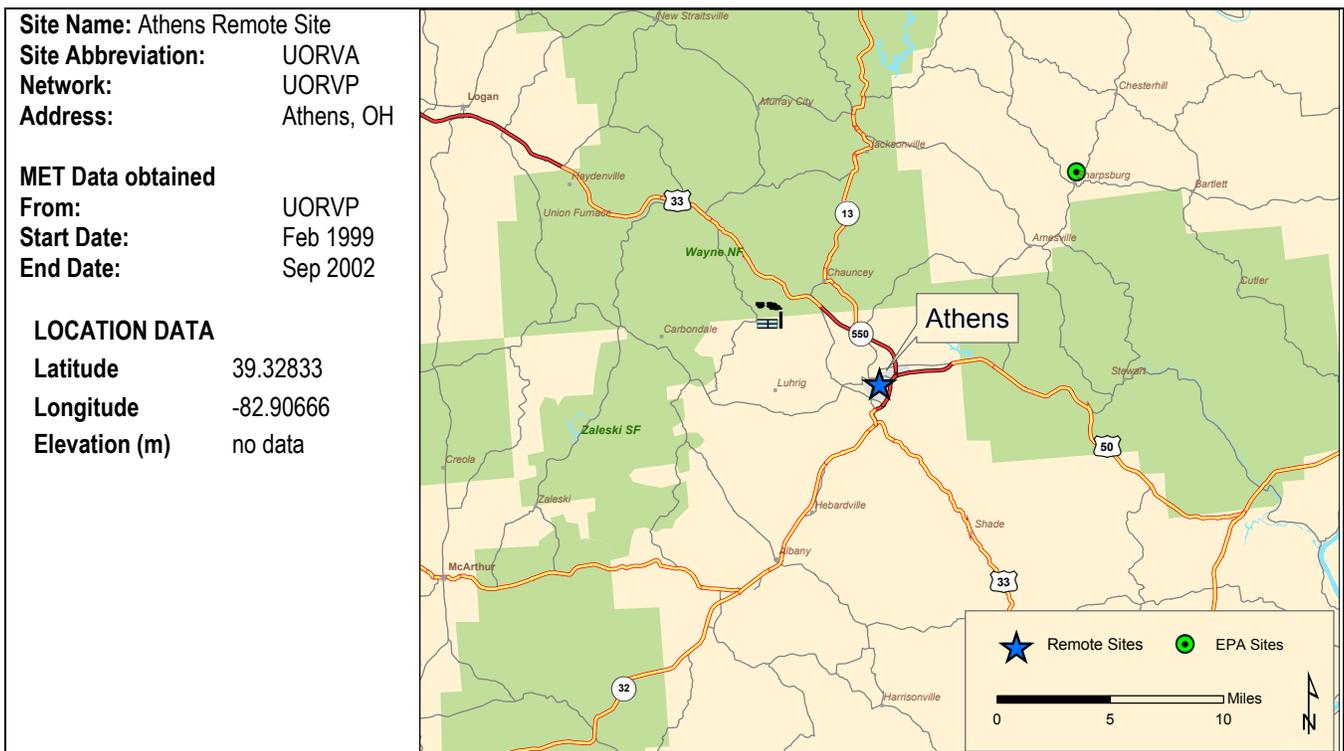
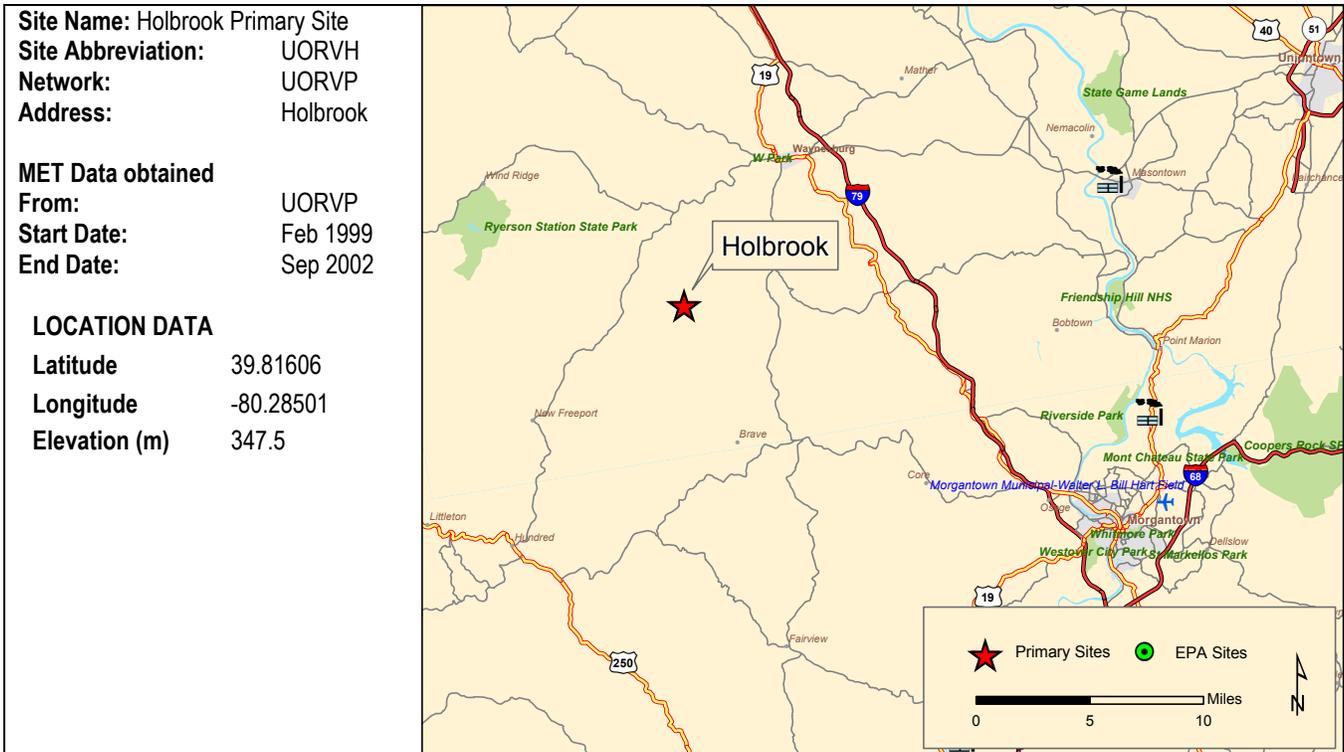
NAME	TYPE	SIZE	MODIFIED
NETL_OST_SURF_TEOM_PM25_Q1_2001.xls	Microsoft Excel	1 MB	Wed 12/18/2002 12:05 PM
NETL_OST_SURF_TEOM_PM25_Q2_2001.xls	Microsoft Excel	1 MB	Wed 12/18/2002 1:59 PM
NETL_OST_SURF_TEOM_PM25_Q3_2001.xls	Microsoft Excel	1 MB	Wed 12/18/2002 2:34 PM
NETL_OST_SURF_TEOM_PM25_Q4_2001.xls	Microsoft Excel	1 MB	Wed 12/18/2002 3:09 PM

F. SCAMP Data Files

NAME	TYPE	SIZE	MODIFIED
Additional_Speciation.mdb	Microsoft Access	332 KB	Wed 1/22/2003 5:21 PM
CEM_Calibration_Data.xls	Microsoft Excel	746 KB	Thu 9/26/2002 3:21 PM
Data_Validation.htm	HTM File	45 KB	Thu 9/26/2002 3:14 PM
Table_Definitions.htm	HTM File	34 KB	Thu 9/26/2002 3:14 PM
Units.htm	HTM File	82 KB	Thu 9/26/2002 3:14 PM
SCAMP Outdoor Data Release 2.mdb	Microsoft Access	51 MB	Fri 2/7/2003 12:30 PM
Scamp020603.mdb	Microsoft Access	3 MB	Fri 2/7/2003 4:42 PM

APPENDIX 2 – SITE SPECIFIC DATA CHARTS

A. UORVP Monitoring Program

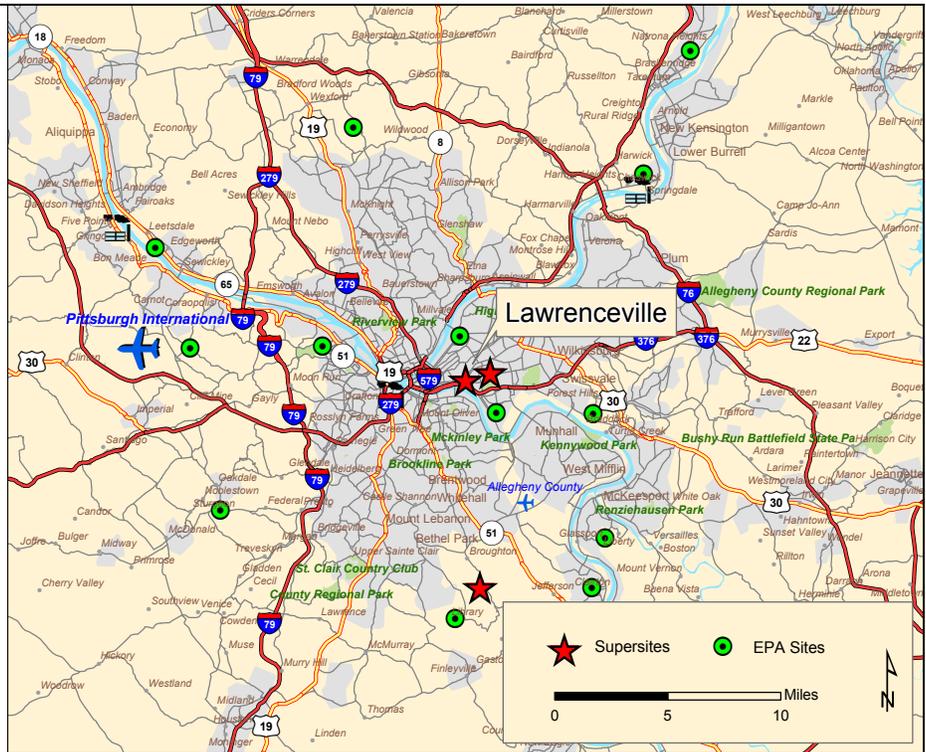


Site Name: Lawrenceville Primary Site
Site Abbreviation: UORVL
Network: UORVP
Address: Lawrenceville

MET Data obtained From: UORVP
Start Date: Feb 1999
End Date: Sep 2002

LOCATION DATA

Latitude 40.43772
Longitude -79.96217
Elevation (m) no data



Site Name: Morgantown Remote Site
Site Abbreviation: UORVM
Network: UORVP
Address: Morgantown

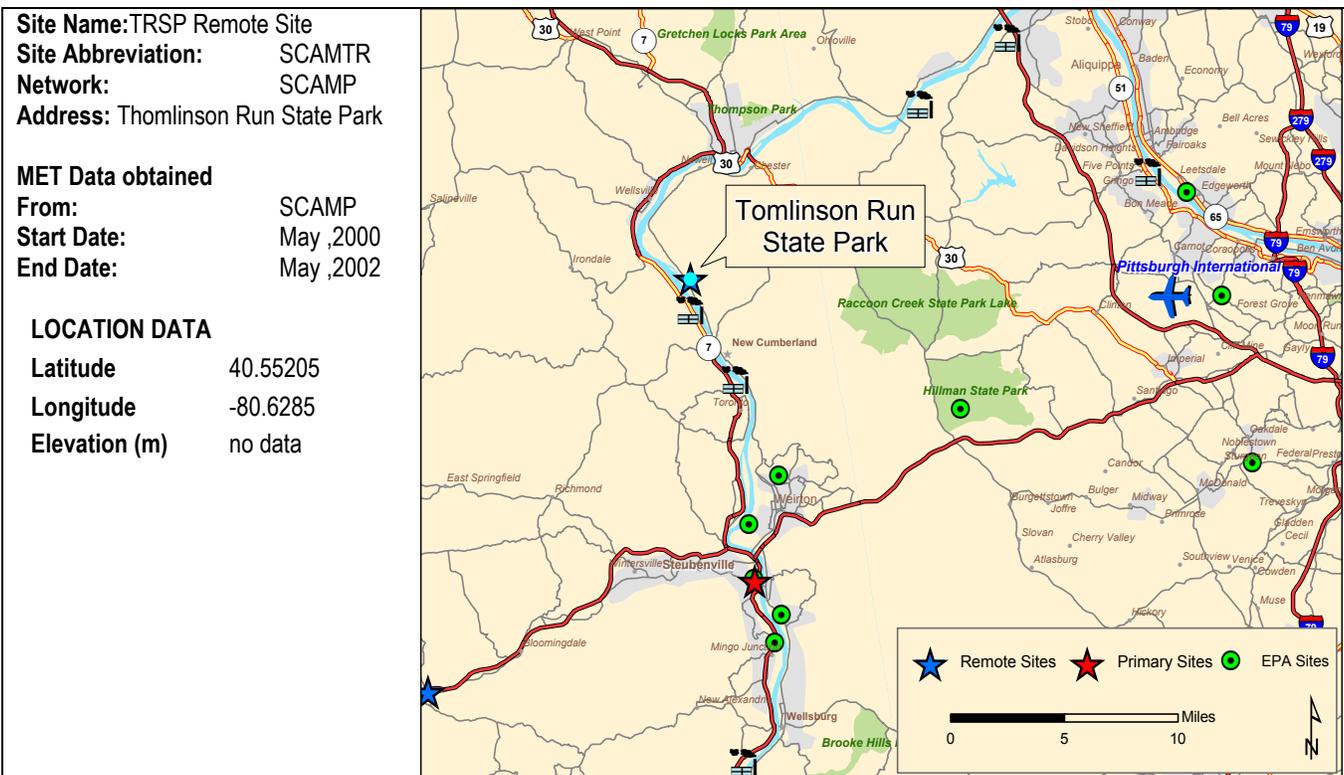
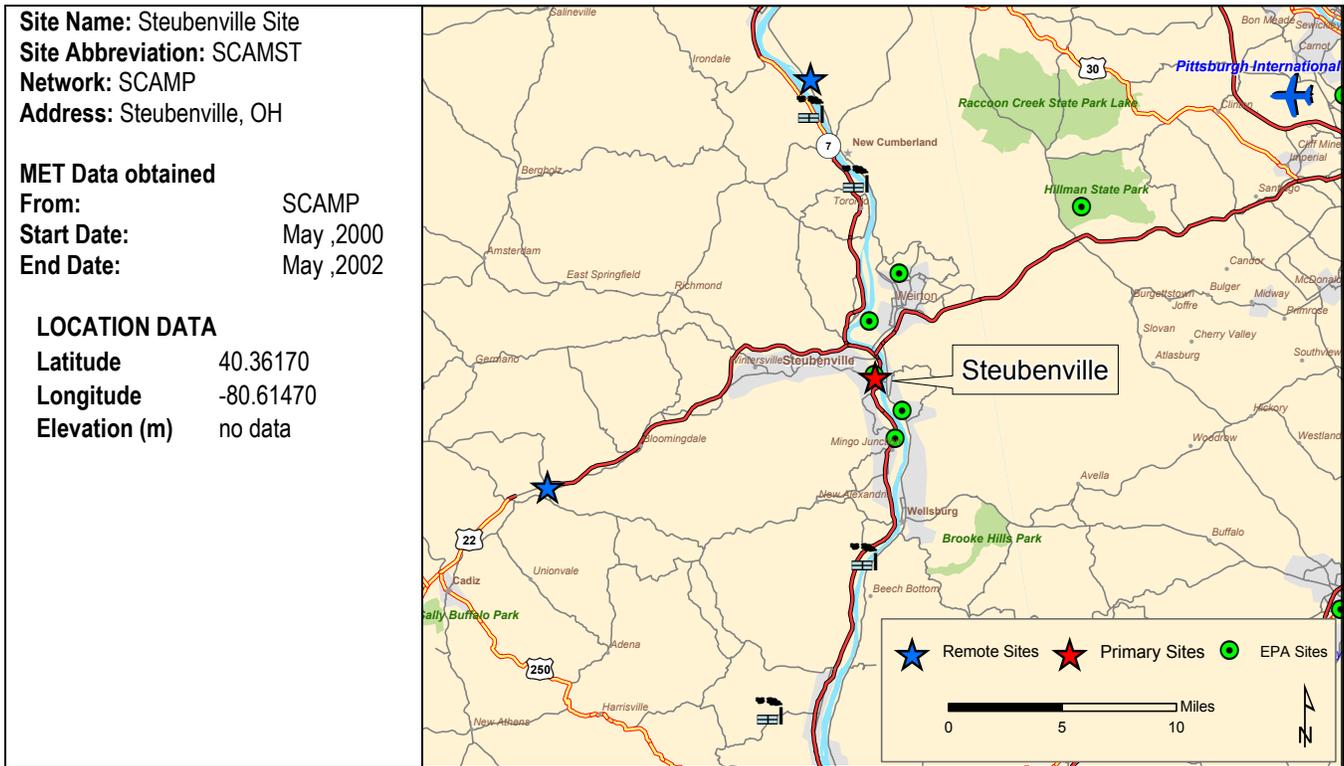
MET Data obtained From: UORVP
Start Date: Feb 1999
End Date: Sep 2002

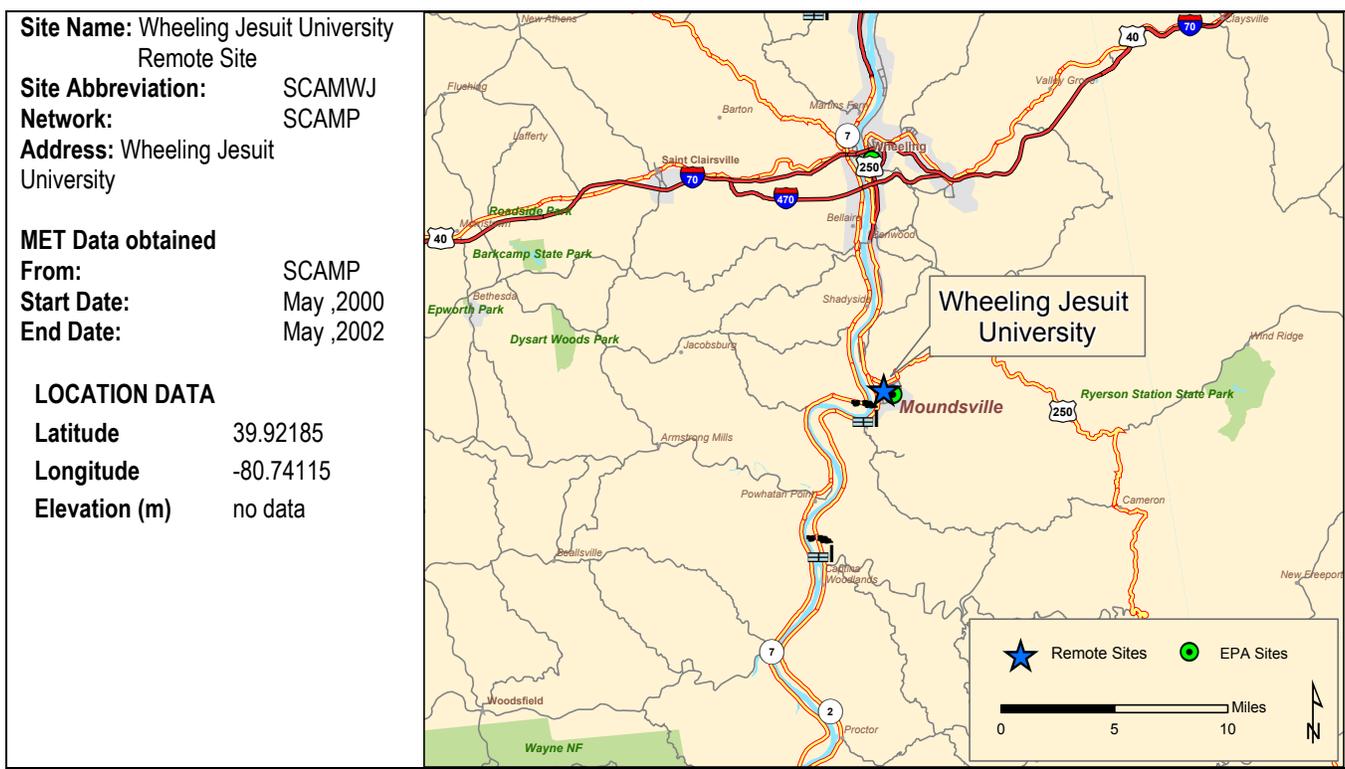
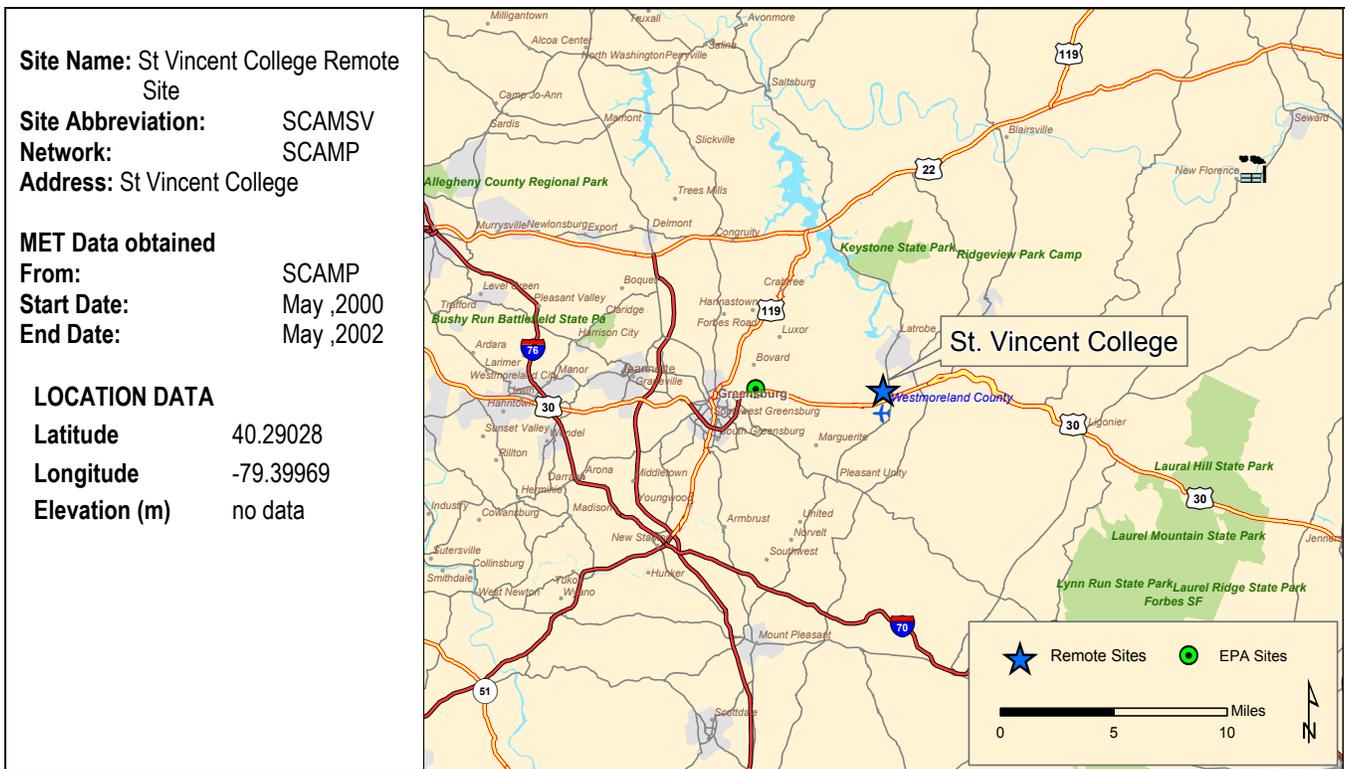
LOCATION DATA

Latitude 39.64283
Longitude -79.91633
Elevation (m) no data



B. SCAMP Monitoring Program



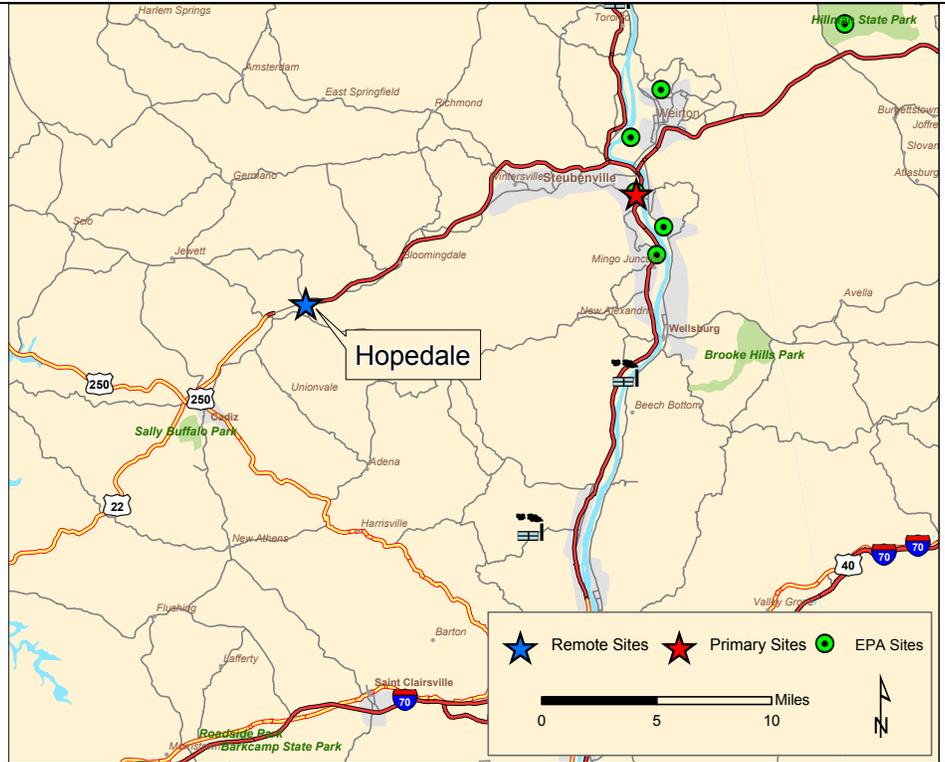


Site Name: Hopedale Remote Site
Site Abbreviation: SCAMH
Network: SCAMP
Address: Hopedale, OH - Hilltop Drive

MET Data obtained From: SCAMP
Start Date: May ,2000
End Date: May ,2002

LOCATION DATA

Latitude 40.3272
Longitude -80.90097
Elevation (m) no data



APPENDIX 3 – DATABASE TABLE DEFINITIONS

A. System Tables

Table Name	tblColumnDefinitions				
Related Tables	Name		Description		
	tblTableDefinitions		Table lists all tables for this database. It is part of the Data Dictionary which is used to construct ad hoc queries and provide an Application Programming Interface (API) for the data tables and columns.		
	tblReadLevels		Table lists all available levels of security for reading information contained in a table or built into a Node.		
	tblWriteLevels		Table lists all available levels of security for Node - Saving/Updating ('Write') functions.		
	tblAdminLevels		Table lists all available levels of security for Administrative ('Admin') functions.		
	tblBrowseLevels		Table lists all available levels of security for browsing ('browse') functions. These functions include listing of non-specified contents of any given object.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	column_ID	int (4)	N/A	Unique identifier for column record.
	FK	table_ID	int (4)	N/A	Foreign Key identifier for table record.
		column_Name	varchar (50)	N/A	Name of column.
	FK	read_Level_ID	int (4)	3	Foreign Key identifier for read level security setting.
	FK	write_Level_ID	int (4)	6	Foreign Key identifier for write level security setting.
	FK	admin_Level_ID	int (4)	6	Foreign Key identifier for Administrator level security setting.
	FK	browse_Level_ID	int (4)	4	Foreign Key identifier for browse level security setting.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all columns for all tables in this database. It is part of the Data Dictionary which is used to construct ad hoc queries.				

Table Name	tblEvents				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	event_Type_ID	int(4)	N/A	Unique identifier for event type record.
		event_Description	varchar(200)	"_"	Description of event type.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table contains predetermined event categories that provide detailed descriptive data for repetitive tasks or transactions such as data file uploads, new user registration, etc.				

Table Name	tblLogs				
Related Tables	Name		Description		
	tblTableDefinitions		Table lists all tables for this database. It is part of the Data Dictionary which is used to construct ad hoc queries and provide an Application Programming Interface (API) for the data tables and columns.		
	tblUsers		Table contains all data specifically related to each user, such as usernames, passwords, contact information and security levels.		
	tblEvents		Table contains predetermined event categories that provide detailed descriptive data for repetitive tasks or transactions such as data file uploads, new user registration, etc.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	log_ID	int(4)	N/A	Unique identifier for log entry record.
	FK	table_ID	int(4)	0	Foreign Key identifier for table record.
	FK	userID	int(4)	0	Foreign Key identifier for user record.
		event_Date	datetime	date()	Date of event.
	FK	event_Type_ID	int(4)	N/A	Foreign Key identifier for event type record.
		event_Log_Entry	text	"_"	Log entry body text.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table used to store all system logging information. Events are pre-coded and available from tblEvents.				

Table Name	tblSystemStamps				
Related Tables	Name		Description		
	tblUsers		Table contains all data specifically related to each user, such as usernames, passwords, contact information and security levels.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	system_Stamp_ID	int (4)	N/A	Unique identifier for trace log identifier.
	FK	user_ID	int (4)	0	Foreign Key identifier for user record.
		system_Stamp	timestamp	N/A	Trace log identifier.
Comments	Table used to store unique timestamp and identification data (trace log identifier) for each transaction.				

Table Name	tblTableDefinitions				
Related Tables	Name		Description		
	tblAdminLevels		Table lists all available levels of security for Administrative ('Admin') functions.		
	tblBrowseLevels		Table lists all available levels of security for browsing ('browse') functions. These functions include listing of non-specified contents of any given object.		
	tblReadLevels		Table lists all available levels of security for reading information contained in a table or built into a Node.		
	tblWriteLevels		Table lists all available levels of security for Node - Saving/Updating ('Write') functions.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	table_ID	int (4)	N/A	Unique identifier for table record.
	FK	pkColumn_ID	int (4)	N/A	Foreign Key identifier for column record for the Primary Key column for this table record.
		table_Name	varchar (50)	N/A	Name of table.
		table_Focus	varchar (100)	"_"	Descriptive comments regarding the focus, or purpose, of this table.
		table_Type	int (4)	0	Numeric representation of table type.
	FK	admin_Level_ID	int (4)	0	Foreign Key identifier for Administrator level security setting.
	FK	browse_Level_ID	int (4)	0	Foreign Key identifier for browse level security setting.
	FK	read_Level_ID	int (4)	3	Foreign Key identifier for read level security setting.
	FK	write_Level_ID	int (4)	6	Foreign Key identifier for write level security setting.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all tables for this database. It is part of the Data Dictionary which is used to construct ad hoc queries and provide an Application Programming Interface (API) for the data tables and columns.				

B. Security Tables

Table Name	tblAdminLevels				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	admin_Level_ID	int (4)	N/A	Unique identifier for level of authorized administrative functions.
		admin_Level	varchar (50)	"_"	Text to describe level of authorization to perform administrative functions.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all available levels of security for Administrative ('Admin') functions.				

Table Name	tblBrowseLevels				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	browse_Level_ID	int (4)	N/A	Unique identifier for authorization levels to browse object contents or properties.
		browse_Level	varchar (100)	"_"	Descriptive text identifying browse authorization levels.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all available levels of security for browsing ('browse') functions. These functions include listing of non-specified contents of any given object.				

Table Name	tblDataSubmitters				
Related Tables	Name		Description		
	tblUsers		Table contains all data specifically related to each user, such as usernames, passwords, contact information and security levels.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	submitter_ID	int (4)	N/A	Unique identifier for data submitter record.
		user_ID	int (4)	N/A	Foreign Key identifier for user record.
		approved	bit	N/A	Logical bit field indicating approval or request.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table used to identify approved data submitters and pending requests.				

Table Name	tblReadLevels				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	read_Level_ID	int (4)	N/A	Unique identifier for read level security setting.
		read_Level	varchar (100)	"_"	Read level description.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all available levels of security for reading information contained in a table or built into a Node.				

Table Name	tblStakeholders				
Related Tables	Name		Description		
	tblUsers		Table contains all data specifically related to each user, such as usernames, passwords, contact information and security levels.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	stakeholder_ID	int (4)	N/A	Unique identifier for stakeholder record.
	FK	user_ID	int (4)	N/A	Foreign Key identifier for user record.
		approved	bit	N/A	Logical bit assignment indicating approval or request.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table used to track stakeholder status requests and approvals.				

Table Name	tblSubscriptions				
Related Tables	Name		Description		
	tblUsers		Table contains all data specifically related to each user, such as usernames, passwords, contact information and security levels.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	subscription_ID	int (4)	N/A	Unique identifier for subscription record.
	FK	user_ID	int (4)	N/A	Foreign Key identifier for user record.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table identifies all users subscribed to email notification of announcements and newsletters.				

Table Name	tblUsers				
Related Tables	Name		Description		
	tblVCards		Table used to track all addresses. This enables multiple objects to be assigned to addresses without storing duplicate data in the database.		
	tblAdminLevels		Table lists all available levels of security for Administrative ('Admin') functions.		
	tblBrowseLevels		Table lists all available levels of security for browsing ('browse') functions. These functions include listing of non-specified contents of any given object.		
	tblReadLevels		Table lists all available levels of security for reading information contained in a table or built into a Node.		
	tblWriteLevels		Table lists all available levels of security for reading information contained in a table or built into a Node.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	user_ID	int (4)	N/A	Unique identifier for user record.
		username	varchar (50)	"_ "	Assigned username.
		password	varchar (50)	"_ "	Encrypted password hash.
		salt	varchar (50)	"_ "	Encrypted password salt.
		email	varchar (150)	"_ "	Email address.
		web_address	varchar (150)	"_ "	Website address.
	FK	vCard_ID	int (4)	0	Foreign Key identifier for vCard record.
		phone	varchar (50)	"_ "	Phone number.
		fax	varchar (50)	"_ "	Fax number.
		mobile	varchar (50)	"_ "	Mobile phone number.
		pager	varchar (50)	"_ "	Pager number.
	FK	admin_Level_ID	int (4)	0	Foreign Key identifier for Administrator level security setting.
	FK	browse_Level_ID	int (4)	0	Foreign Key identifier for browse level security setting.
	FK	read_Level_ID	int (4)	3	Foreign Key identifier for read level security setting.
	FK	write_Level_ID	int (4)	0	Foreign Key identifier for write level security setting.
		user_First	varchar (75)	"_ "	User's first name.
		user_Last	varchar (75)	"_ "	User's last name.
		user_Middle	char (1)	"_ "	User's middle initial.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table contains all data specifically related to each user, such as usernames, passwords, contact information and security levels.				

Table Name	tblWriteLevels				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	write_Level_ID	int(4)	N/A	Unique identifier for write level security settings.
		write_Level	varchar(100)	"_"	Write level description.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all available levels of security for Node - Saving/Updating ('Write') functions.				

C. Observation Tables

Table Name	tblCollectionPrinciples				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	collection_Principle_ID	int (4)	N/A	Unique identifier for collection principle record.
		collection_Principle_Code	varchar (50)	"_"	Abbreviation or code representing specific collection principle record.
		collection_Principle_Description	varchar (100)	"_"	Complete description of collection principle record.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds collection principle records which represent data class sub-categories, supplementing parameter category records.				

Table Name	tblDataClasses				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	data_Class_ID	int (4)	N/A	Unique identifier for data class record.
		data_Class	varchar (50)	"_"	Data class description.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table is used to classify the data by broadest category.				

Table Name	tblFilters				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	filter_ID	int (4)	N/A	Unique identifier for filter record.
		filter_Code	varchar (50)	"_ "	Locally assigned filter code.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table links filter (Speciation) records to observational data.				

Table Name	tblInstruments				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	instrument_ID	int (4)	N/A	Unique identifier for instrument record.
		instrument_Name	varchar (250)	"_ "	Name of instrument.
		instrument_Manufacturer	varchar (250)	"_ "	Instrument manufacturer's name.
		instrument_Serial	varchar (100)	"_ "	Serial number for instrument.
		instrument_Model	varchar (100)	"_ "	Model number for instrument.
		acquisition_Date	datetime	date ()	Date instrument was acquired.
		disposition_Date	datetime	NULL	Date instrument was disposed of.
	FK	system_Stamp_ID		N/A	Foreign Key identifier for trace log identifier.
Comments	Table stores all Instrument metadata such as serial number, model number and acquisition/disposition dates.				

Table Name	tblInstrumentLocations				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	instrument_Location_ID	int (4)	N/A	Unique identifier for instrument location record.
		instrument_Location_Code	varchar (25)	"_"	Code that describes the location of instrument.
		instrument_Location_Description	varchar (50)	"_"	Full description of instrument's location.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table used to store descriptive data regarding the collection instrument's location; e.g. stationary-surface, mobile-surface, etc.				

Table Name	tblMetaflags				
Related Tables	Name		Description		
	tblNARSTOMetaflags		Table lists all NARSTO-assigned QA/QC flags.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	metaflag_ID	int (4)	N/A	Unique identifier for metaflag record.
		metaflag	varchar (3)	"_"	Metaflag used to associate <i>metadata</i> with a data item.
		metaflat_Description	varchar (250)	"_"	Description of <i>metadata</i> .
	FK	NARSTO_metaflag_ID	int (4)	N/A	Foreign Key identifier for NARSTO standard metaflag equating to local metaflag.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table has all locally-assigned QA/QC flags and maps each local code to an equivalent NARSTO QA/QC flag.				

Table Name	tblMethods				
Related Tables	Name		Description		
	tblInstruments		Table stores all Instrument metadata such as serial number, model number and acquisition/disposition dates.		
	tblInstrumentLocations		Table used to store descriptive data regarding the collection instrument's location; e.g. stationary-surface, mobile-surface, etc.		
	tblParameters		Table holds descriptive data related to all parameters. Parameters are a broad category of data that can apply to multiple data classes.		
	tblSampleFrequencies		Table holds sample frequency, or interval, information for observation values.		
	tblSampleDurations		Table holds sample duration data which describes durations for which an observation value represents.		
	tblUnits		Table stores all unit combinations utilized for data collection and entry, along with conversion factors to easily convert values between systems.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	method_ID	int (4)	N/A	Unique identifier for method record.
	FK	instrument_ID	int (4)	N/A	Foreign Key identifier for instrument record.
	FK	instrument_Location_ID	int (4)	N/A	Foreign Key identifier for instrument location record.
	FK	parameter_ID	int (4)	N/A	Foreign Key identifier for parameter record.
	FK	sample_Frequency_ID	int (4)	0	Foreign Key identifier for sample frequency record.
	FK	sample_Duration_ID	int (4)	0	Foreign Key identifier for sample duration record.
	FK	unit_ID	int (4)	N/A	Foreign Key identifier for unit record.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds method records which link data class and parameter sub-categories into a cohesive descriptive row, describing a specific observation record.				

Table Name	tblNARSTOMetaflags				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	NARSTO_metaflag_ID	int(4)	N/A	Unique identifier for NARSTO metaflag record.
		NARSTO_metaflag	varchar(3)	"_"	NARSTO metaflag.
		NARSTO_metaflag_Description	varchar(250)	"_"	NARSTO metaflag description.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all NARSTO-assigned QA/QC flags.				

Table Name	tblNetworks				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	network_ID	int(4)	N/A	Unique identifier for study or network record.
		network_Code	char(3)	"_ _ _"	Study or network code.
		network_Description	varchar(100)	"_"	Study or network description.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all assigned Network codes for project data collection programs.				

Table Name	tblObservationFlags				
Related Tables	Name		Description		
	tblObservations		Table holds all observation readings, along with date and time values associated with each value.		
	tblMetaflags		Table has all locally-assigned QA/QC flags and maps each local code to an equivalent NARSTO QA/QC flag.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	observation_Flag_ID	int (4)	N/A	Unique identifier for observation flag record.
	FK	observation_ID	int (4)	N/A	Foreign Key identifier for observation record.
	FK	metaflag_ID	int (4)	N/A	Foreign Key identifier for metaflag record.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table associates observation records with metaflag records.				

Table Name	tblObservations				
Related Tables	Name		Description		
	tblTimeZones				
	tblDataClasses				
	tblSites				
	tblFilters				
	tblMethods				
	tblQCStatus				
	tblSystemStamps				
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	observation_ID	int (4)	N/A	Unique identifier for observation record.
		observation_Date	datetime	date ()	Date observation recorded.
		start_Time	datetime	time ()	Start time for observation value.
		end_Time	datetime	time ()	End time for observation value.
	FK	time_Zone_ID	int (4)	0	Foreign Key identifier for Time Zone record.
	FK	data_Class_ID	int (4)	0	Foreign Key identifier for data class record.
	FK	site_ID	int (4)	0	Foreign Key identifier for site record.
	FK	filter_ID	int (4)	0	Foreign Key identifier for filter record.
	FK	method_ID	int (4)	0	Foreign Key identifier for method record.
		observation_Value	float (8)	-999	Observed data value.
	FK	QC_Status_Code_ID	int (4)	0	Foreign Key identifier for Quality Control Status code.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds all observation readings, along with date and time values associated with each value.				

Table Name	tblParameterIdentifiers				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	parameter_Identifier_ID	int (4)	N/A	Unique identifier for parameter identifier record.
		parameter_Identifier_Code	varchar (25)	"_"	Code representing parameter identifier.
		parameter_Identifier_Description	varchar (50)	"_"	Parameter identifier description.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds data that identifies a specific sub-category of a parameter value; e.g. WD for Wind Direction, PM25 for particulate matter 2.5.				

Table Name	tblParameterProperties				
Related Tables	Name		Description		
	tblSystemStamps				
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	parameter_Property_ID	int (4)		Unique identifier for parameter property record.
		parameter_Property_Code	varchar (3)		Three character code for parameter property.
		parameter_Property_Description	varchar (100)		Parameter property description.
	FK	system_Stamp_ID	int (4)		Foreign Key identifier for trace log identifier.
Comments	Table holds parameter property records which represent parameter sub-categories.				

Table Name	tblParameters				
Related Tables	Name		Description		
	tblParameterIdentifiers		Table holds data that identifies a specific sub-category of a parameter value; e.g. WD for Wind Direction, PM25 for particulate matter 2.5.		
	tblParameterProperties		Table holds parameter property records which represent parameter sub-categories.		
	tblDataClasses		Table is used to classify the data by broadest category.		
	tblCollectionPrinciples		Table holds collection principle records which represent data class sub-categories, supplementing parameter category records.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	parameter_ID	int (4)	N/A	Unique identifier for parameter record.
		parameter_Source	varchar(100)	"_"	Descriptive source information for parameter record.
		parameter_Description	varchar(100)	"_"	Text describing parameter.
		max_Particle_Size	float (8)	0	Maximum particle size for parameter.
	FK	data_Class_ID	int (4)	0	Foreign Key identifier for data class record.
	FK	parameter_Identifier_ID	int (4)	0	Foreign Key identifier for parameter identifier record.
	FK	collection_Principle_ID	int (4)	0	Foreign Key identifier for collection principle record.
	FK	parameter_Property_ID	int (4)	0	Foreign Key identifier for parameter property record.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds descriptive data related to all parameters. Parameters are a broad category of data that can apply to multiple data classes.				

Table Name	tblQCStatus				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	QC_Status_Code_ID	int(4)	N/A	Unique identifier for Quality Control Status record.
		QC_Status_Code	char(2)	"_ _"	Two character code for Quality Control Status.
		QC_Status_Description	varchar(100)	"_ "	Quality Control Status description.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table lists all Quality Control values, indicating level of QA/QC handling that each observation value has been exposed to.				

Table Name	tblSampleDurations				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	sample_Duration_ID	int(4)	N/A	Unique identifier for sample duration record.
		sample_Duration_Code	varchar(50)	"_ "	Code to represent sample duration.
		sample_Duration_Text	varchar(100)	"_ "	Sample duration description.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds sample duration data which describes durations for which an observation value represents.				

Table Name	tblSampleFrequencies				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	sample_Frequency_ID	int (4)	N/A	Unique identifier for sample frequency record.
		sample_Frequency_Code	varchar (50)	"_"	Code identifying sample frequency.
		sample_Frequency_Text	varchar (100)	"_"	Frequency description.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds sample frequency, or interval, information for observation values.				

Table Name	tblSites				
Related Tables	Name		Description		
	tblNetworks		Table lists all assigned Network codes for project data collection programs.		
	tblCities		This table lists all cities included in the dataset.		
	tblSites		Table holds all site-related data, such as positioning coordinates, identification data, etc.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	site_ID	int (4)	N/A	Unique identifier for collection site record.
		site_Code	varchar (50)	"_"	Code used to identify collection site.
		site_Name	varchar (150)	"_"	Collection site name.
	FK	network_ID	int (4)	0	Foreign Key identifier for network record.
		program_Code	varchar (3)	"_"	Program code associated with collection site.
		site_Type_Code	varchar (2)	"_"	Code used to identify the type of collection site.
		longitude	float (8)	0	Decimal value equal to collection site's longitude coordinates.
		latitude	float (8)	0	Decimal value equal to collection site's latitude coordinates.
		elevation	float (8)	0	Decimal value equal to collection site's elevation (meters above sea level).
	FK	vCard_ID	int (4)	0	Foreign Key identifier for vCard record.
	FK	met_Site_ID	int (4)	0	Foreign Key identifier for site record that collects meteorology data associated with this site.
		status_ID	varchar (10)	N/A	Site status identifying text.
		start_Date	datetime	N/A	Date site started collection efforts.
		end_Date	datetime	N/A	Last date site collected data.
		site_Notes	varchar (100)	"_"	Site-specific notes.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds all site-related data, such as positioning coordinates, identification data, etc.				

D. Support Tables

Table Name	tblCities				
Related Tables	Name		Description		
	tblCounties		Table linked by county_ID to identify the name of the county where the city is located.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	city_ID	int (4)	N/A	Unique identifier for city table entry.
		city_Code	char (3)	" "	Closest airport/city code.
		city_Name	varchar (50)	" "	Name of city.
	FK	county_ID	int (4)	N/A	Foreign key reference to identify the associated county.
	FK	system_Stamp_ID	int (4)	N/A	Trace log identifier.
Comments	This table lists all cities included in the dataset.				

Table Name	tblCounties				
Related Tables	Name		Description		
	tblState		Table contains a list of all states names, abbreviations and country_ID's for each state.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	county_ID	int (4)	N/A	Unique identifier for county record.
		county_Abbreviation	varchar (50)	" "	Abbreviation for county.
		county_Name	varchar (150)	" "	Name of county
	FK	state_ID	int (4)	N/A	Foreign Key identifier for state record.
	FK	system_Stamp_ID	int (4)	N/A	Trace log identifier.
Comments	Table linked by county_ID to identify the name of the county where the city is located.				

Table Name	tblCountries				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	country_ID	int (4)	N/A	Unique identifier for country record.
		country_Code	varchar (3)	" "	Internationally recognized nation identifier.
		country_Name	varchar (50)	" "	Name of country.
	FK	system_Stamp_ID	int (4)	N/A	Trace log identifier.
Comments	Table used to list all available country names and codes for use in identifying reference locations for the data in the observations tables.				

Table Name	tblStates				
Related Tables	Name		Description		
	tblCountries		Table used to list all available country names and codes for use in identifying reference locations for the data in the observations tables.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	state_ID	int(4)	N/A	Unique identifier for state record.
		state_Abbreviation	char(2)	"_ _"	Two letter abbreviation for state.
		state_Name	varchar(75)	"_ "	Name of state.
	FK	country_ID	int(4)	0	Foreign Key identifier for country record.
	FK	system_Stamp_ID	int(4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table contains a list of all states names, abbreviations and country_ID's for each state.				

Table Name	tblThreadedNotes				
Related Tables	Name		Description		
	tblThreadedNotes		Table holds note records that are associated with any table and column value, in a discussion-list format, utilizing root, parent and child nodes.		
	tblTableDefinitions		Table lists all tables for this database. It is part of the Data Dictionary which is used to construct ad hoc queries and provide an Application Programming Interface (API) for the data tables and columns.		
	tblColumnDefinitions		Table lists all columns for all tables in this database. It is part of the Data Dictionary which is used to construct ad hoc queries.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	note_ID	int (4)	N/A	Unique identifier for threaded note record.
		entry_Date	datetime	N/A	Entry date for threaded note record.
		entry_Text	text	"_"	Body text of note entry.
	FK	parent_Note_ID	int (4)	0	Foreign Key identifier for parent note for note entry record.
	FK	table_ID	int (4)	0	Foreign Key identifier for table record associated with note entry record.
	FK	column_ID	int (4)	0	Foreign Key identifier for column record associated with note entry record.
		pkValue	int (4)	0	Primary Key value for table record identified by table_ID.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds note records that are associated with any table and column value, in a discussion-list format, utilizing root, parent and child nodes.				

Table Name	tblTimeZones				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	time_Zone_ID	int (4)	N/A	Unique identifier for Time Zone record.
		time_Zone	char (3)	"_"	Three character (maximum) representation for Time Zone; e.g. EST for Eastern Standard Time.
		gmt_Offset	int (4)	0	Integer offset from Greenwich Mean Time (GMT).
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table holds all Time Zone and GMT offset data used to convert data readings to standard (GMT) Time Zone.				

Table Name	tblUnits				
Related Tables	Name		Description		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	unit_ID	int (4)	N/A	Unique identifier for unit record.
		unit_Abbreviation	varchar (50)	"_"	Abbreviation of unit text.
		unit_Description	varchar (100)	"_"	Full description of unit.
		conversion_Factor	float	1	Multiplier to arrive at converted unit value.
	FK	conversion_Unit_ID	int (4)	N/A	Unit corresponding to converted value.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table stores all unit combinations utilized for data collection and entry, along with conversion factors to easily convert values between systems.				

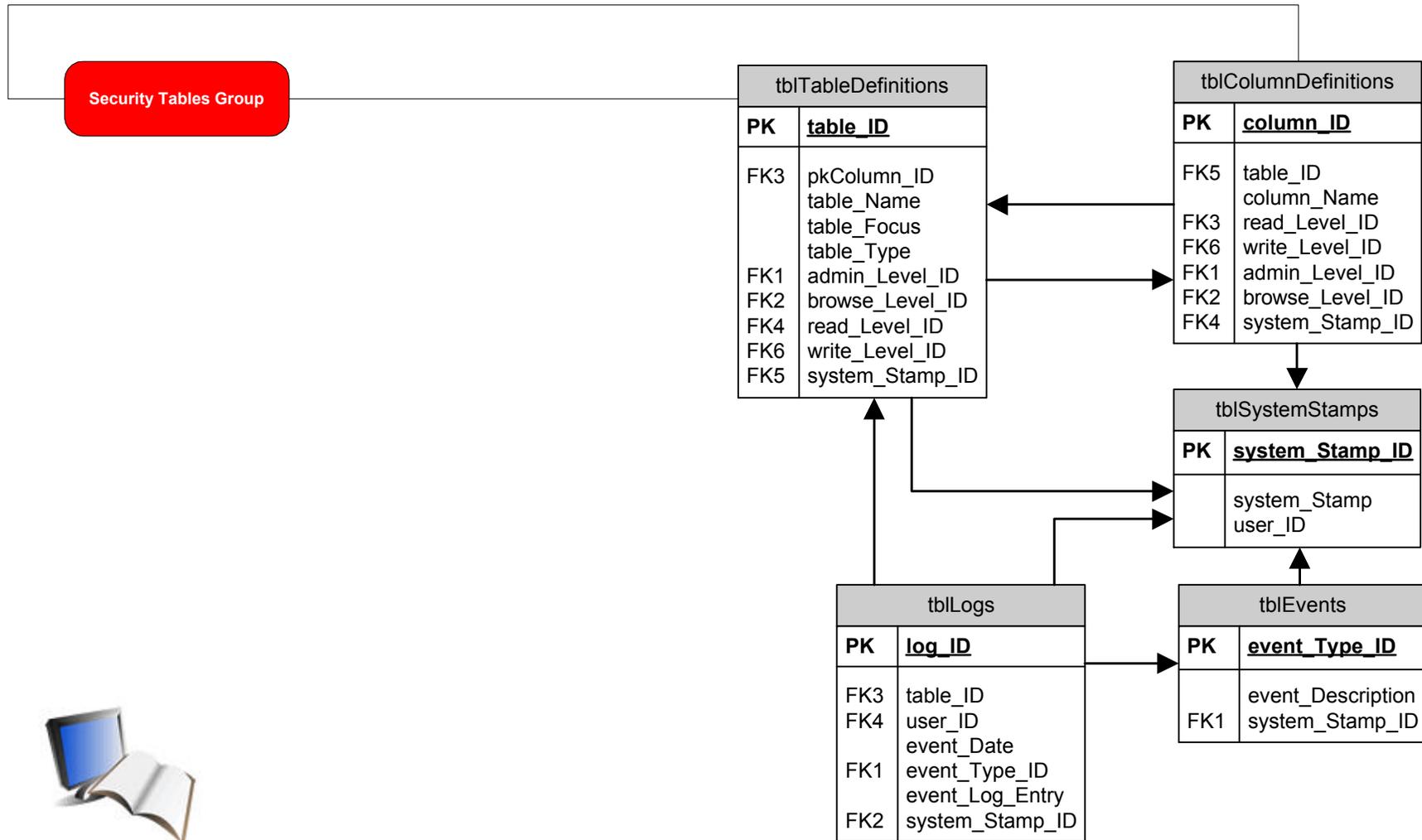
Table Name	tblVCards				
Related Tables	Name		Description		
	tblCities		This table lists all cities included in the dataset.		
	tblSystemStamps		Table used to store unique timestamp and identification data (trace log identifier) for each transaction.		
Related Views	Name		Description		
	N/A		N/A		
Columns	Key	Name	Data Type	Default Value	Comments
	PK	vCard_ID	int (4)	N/A	Unique identifier for vCard record.
		addr1	varchar (50)	"_"	Line 1 of address.
		addr2	varchar (50)	"_"	Line 2 of address.
	FK	city_ID	int (4)	0	Foreign Key identifier for city record.
		zip_code	varchar (9)	"_"	Zip code for address.
	FK	system_Stamp_ID	int (4)	N/A	Foreign Key identifier for trace log identifier.
Comments	Table used to track all addresses. This enables multiple objects to be assigned to addresses without storing duplicate data in the database.				

APPENDIX 4 – STORED PROCEDURE DEFINITIONS

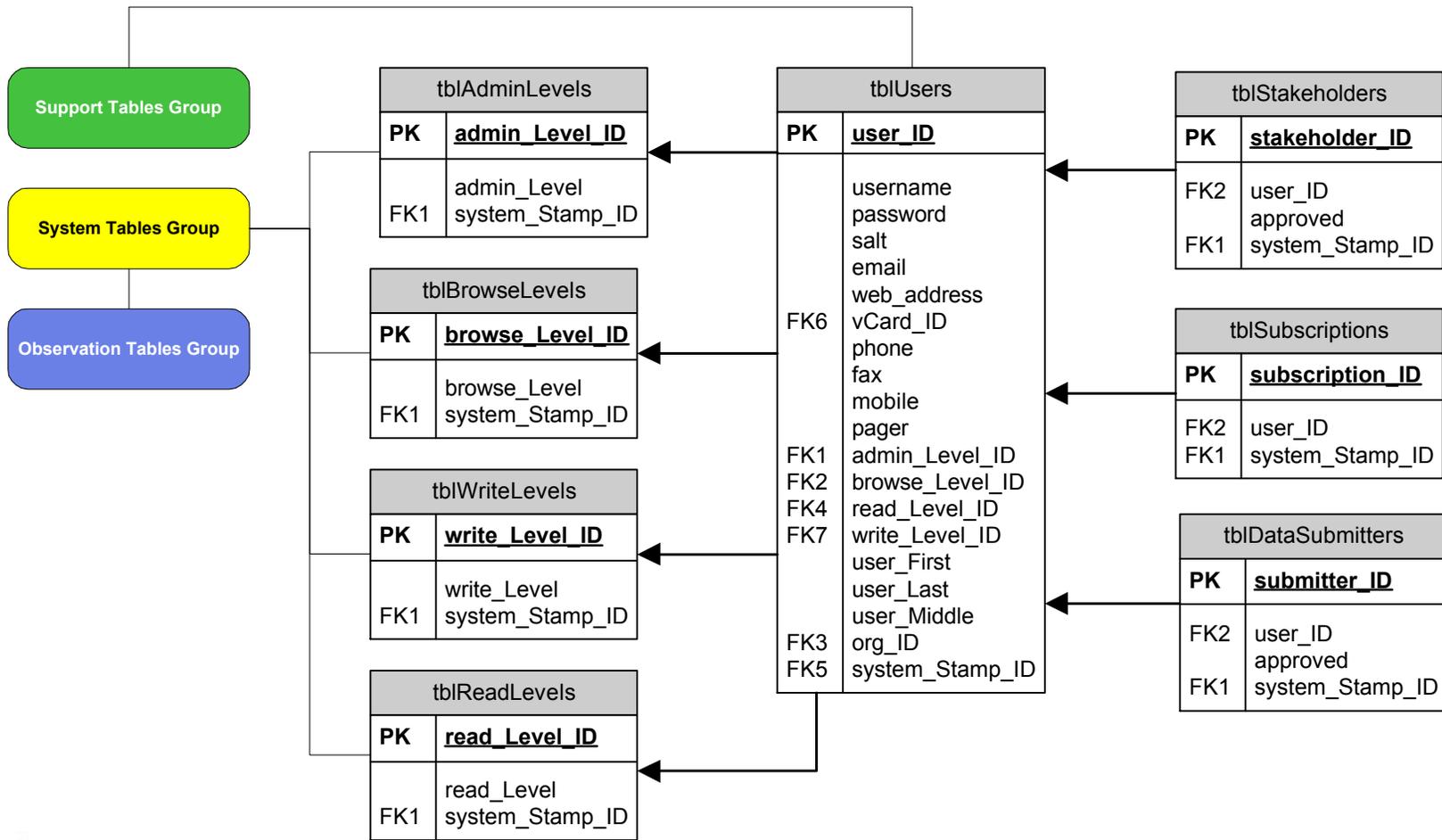
Stored Procedure Name	spNewPMUser	
Tables Referenced	tblUsers	
SQL Code	<pre>CREATE PROCEDURE spNewPMUser @userName varchar(50), @passwordHash varchar(40), @salt varchar(10), @email varchar(150), @web_address varchar(150), @vCard_ID int, @phone varchar(50), @fax varchar(50), @mobile varchar(50), @pager 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, password, 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</pre>	
Required Parameters	Name	Type
	ALL	
Comments	This stored procedure is used to insert new users into the application's users table.	

Stored Procedure Name	spUserLookup	
Tables Referenced	tblUsers	
SQL Code	<pre>CREATE PROCEDURE spUserLookup @userName varchar(50) AS SELECT password, salt FROM tblUsers WHERE username = @userName GO</pre>	
Required Parameters	Name	Type
	@userName	varchar(50)
Comments	This stored procedure is used to extract authentication data from tblUsers. It requires a valid username to locate the desired record within the table.	

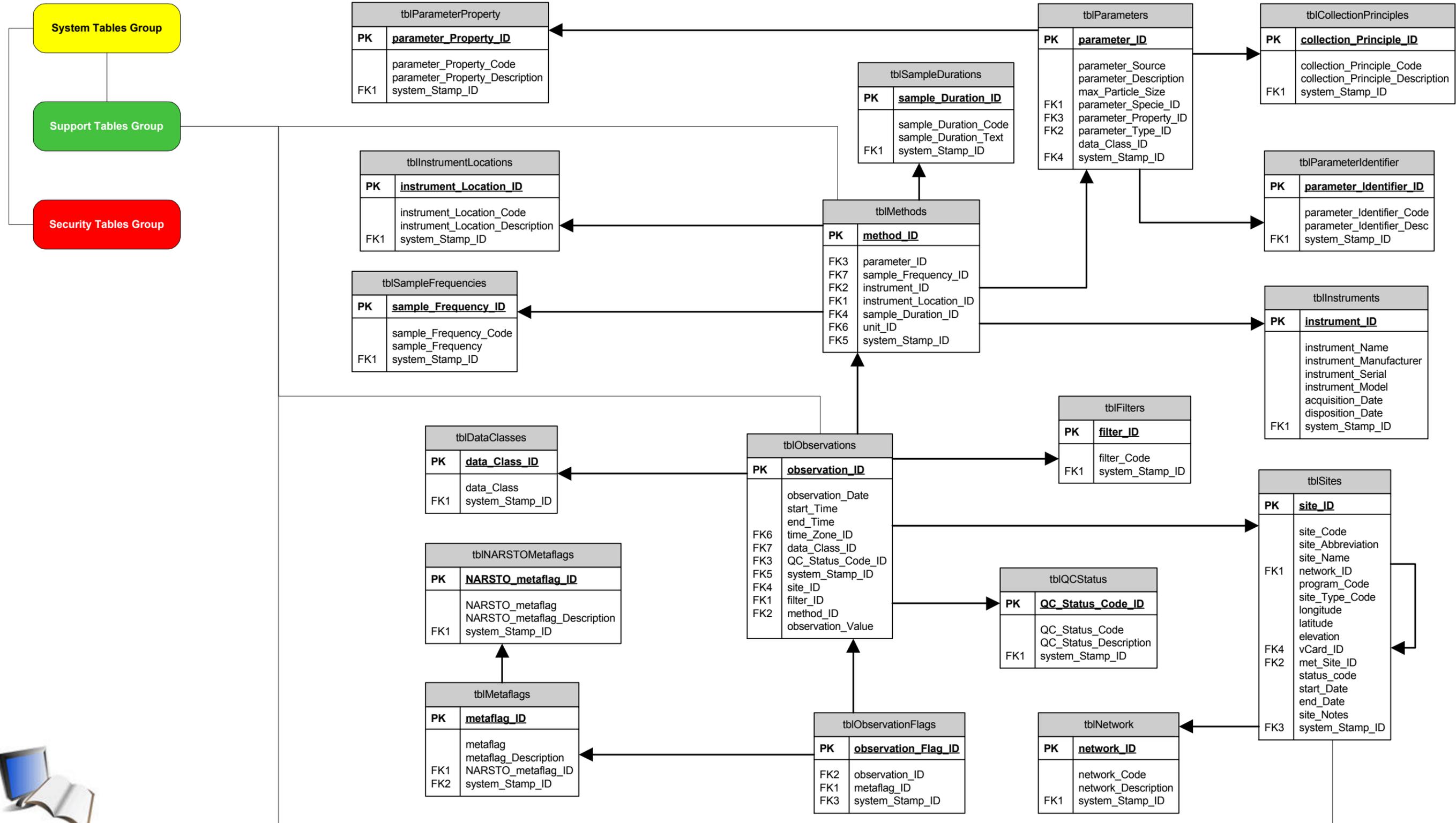
APPENDIX 5 – DATABASE RELATIONSHIP DIAGRAMS



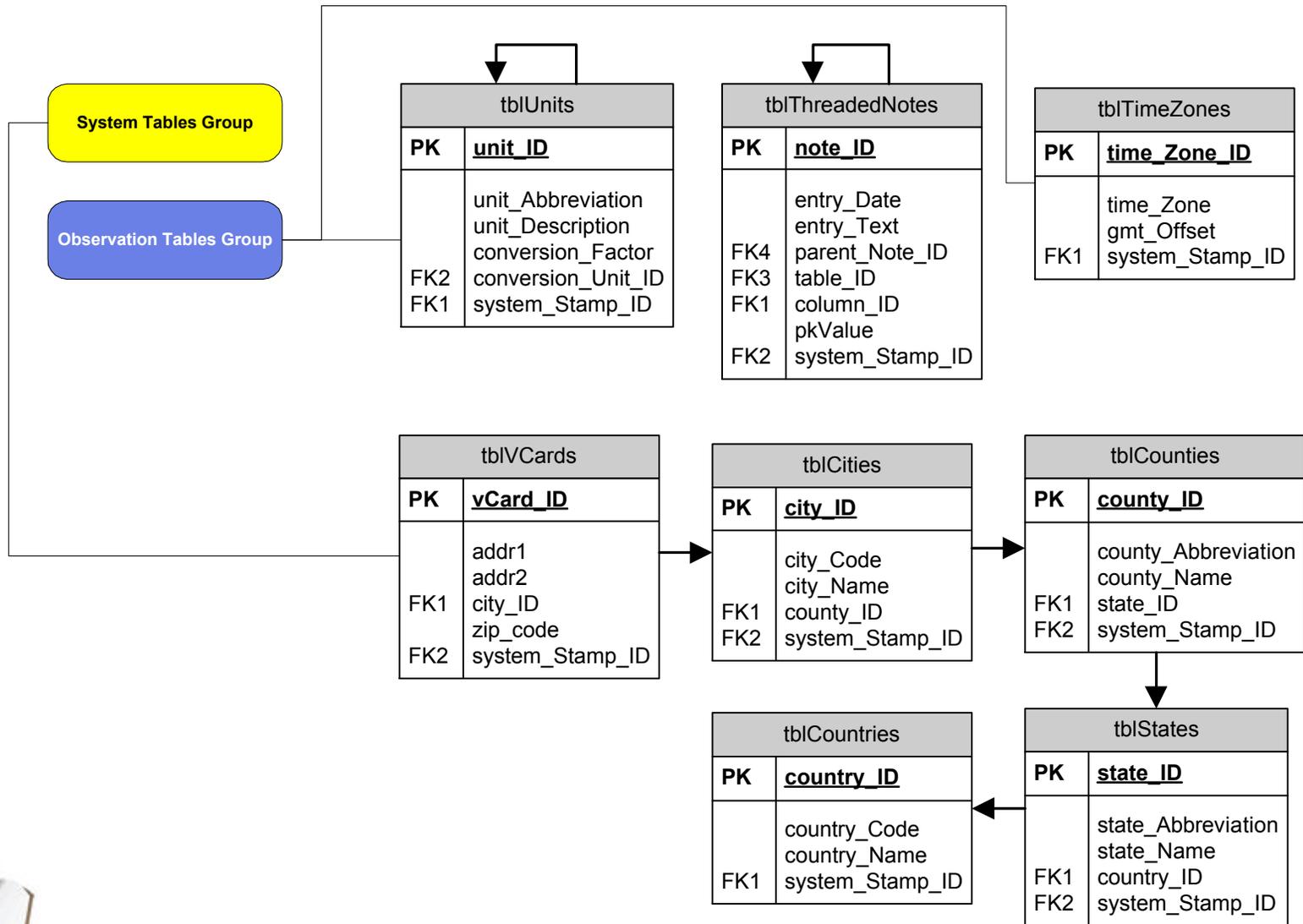
System Tables Group



Security Tables Group



Observation Tables Group



Support Tables Group

APPENDIX 6 – DATA INPUT FILE REQUIREMENTS

A. Required Input File Header Fields

FIELD NAME	DESCRIPTION
FILE_NAME	Name of file containing this header.
DATA_EXCHANGE_VERSION	Version number of established file control specification.
COMMENT	Comments related to this file.
QUALITY_CONTROL_LEVEL	Quality Control Level. (see Appendix 8 – Quality Control Definitions)
DATE_THIS_FILE_GENERATED	Date the final copy of this file was created.
DATA_SUBMITTER_USER_ID	System <i>username</i> assigned by Application Administrator(s).
ORGANIZATION_ACRONYM	Acronym or abbreviation of submitting organization's name.
ORGANIZATION_NAME	Full name of submitting organization.
STUDY_OR_NETWORK_ACRONYM	Acronym or abbreviation of study or network responsible for data collection, QA/QC and submission.

STUDY_OR_NETWORK_NAME	Name of study or network responsible for data collection, QA/QC and submission.
FILE_CONTENTS_DESCRIPTION_SHORT	Brief description of the data contained in this file.
FILE_CONTENTS_DESCRIPTION_LONG	Full description of the data contained in this file.
PRINCIPAL_INVESTIGATOR_NAME_LAST	Last name of the Principal Investigator responsible for this dataset.
PRINCIPAL_INVESTIGATOR_NAME_FIRST	First name of the Principal Investigator responsible for this dataset.
PRINCIPAL_INVESTIGATOR_NAME_TITLE	Title of the Principal Investigator responsible for this dataset.
PRINCIPAL_INVESTIGATOR_AFFILIATION	Organizational affiliation of the Principal Investigator responsible for this dataset.
CO_INVESTIGATOR_NAME_LAST	Last name of the Co-Investigator responsible for this dataset.
CO_INVESTIGATOR_NAME_FIRST	First name of the Co-Investigator responsible for this dataset.
CO_INVESTIGATOR_AFFILIATION	Organizational affiliation of the Co-Investigator responsible for this dataset.
COUNTRY_CODE	Internationally recognized character representation of a country name.

STATE_OR_PROVINCE_CODE	Two letter abbreviation for US state name.
PRINCIPAL_INVESTIGATOR_CONTACT_INFORMATION	Preferred contact information for the Principal Investigator.
DATA_USAGE_ACKNOWLEDGEMENT	Insert acknowledgement statement provided by Application Administrator.
NAME_OF_PERSON_WHO_GENERATED_THIS_FILE	Full name of individual responsible for the generation of this file.
AFFILIATION_OF_PERSON_WHO_GENERATED_THIS_FILE	Organizational affiliation of the individual responsible for the generation of this file.
DATE_OF_LAST_MODIFICATION_TO_DATA_IN_MAIN_TABLE	Date that the final changes to this file were made.
FILE_CHANGE_HISTORY_DESCRIPTION	Descriptive text detailing any relative history of this file.

B. Tabular Format Required Tags and Keywords

TAG / KEYWORD	DESCRIPTION
[bof]	Tag used to indicate start of data input file.
[begin header]	Tag used to indicate start of data input file header section.
[end header]	Tag used to indicate end of data input file header section.
[begin table]	Tag used to indicate start of data input file data table section.
[end table]	Tag used to indicate end of data input file data table section.
[eof]	Tag used to indicate end of data input file.
table_Name	Name of table section in data input file.
column_Names	Name of columns, or parameters, corresponding to data values in each table row.
reference_Table	Name of table section within the same data input file that contains reference data identified by column name.
data_Row	Keyword to indicate to the start of a new row within the data table.

C. XML Format Required Tags

TAG	DESCRIPTION
<p><xml_Input_File parameters> </xml_Input_File></p> <p>Required Parameters: name="filename.xml"</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of data input file.</p>
<p><header parameters></header></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end data input file header section.</p>
<p><DATA_EXCHANGE_VERSION> </DATA_EXCHANGE_VERSION></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of version number of established file control specification.</p>
<p><COMMENT></COMMENT></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of comments related to this file.</p>
<p><QUALITY_CONTROL_LEVEL> </QUALITY_CONTROL_LEVEL></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of Quality Control Level. (see Appendix 8 – Quality Control Definitions)</p>

<p><DATE_THIS_FILE_GENERATED> </DATE_THIS_FILE_GENERATED></p> <p>Required Parameters: format="MM/DD/YYYY" - or - format="MM/DD/YY"</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of date the final copy of this file was created.</p>
<p><DATA_SUBMITTER_USER_ID> </DATA_SUBMITTER_USER_ID></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of system username assigned by Application Administrator(s).</p>
<p><ORGANIZATION_ACRONYM> </ORGANIZATION_ACRONYM></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of acronym or abbreviation of submitting organization's name.</p>
<p><ORGANIZATION_NAME> </ORGANIZATION_NAME></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of full name of submitting organization.</p>
<p><STUDY_OR_NETWORK_ACRONYM> </STUDY_OR_NETWORK_ACRONYM></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of acronym or abbreviation of study or network responsible for data collection, QA/QC and submission.</p>

<p><STUDY_OR_NETWORK_NAME> </STUDY_OR_NETWORK_NAME></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of name of study or network responsible for data collection, QA/QC and submission.</p>
<p><FILE_CONTENTS_DESCRIPTION_SHORT> </FILE_CONTENTS_DESCRIPTION_SHORT></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of brief description of the data contained in this file.</p>
<p><FILE_CONTENTS_DESCRIPTION_LONG> </FILE_CONTENTS_DESCRIPTION_LONG></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of full description of the data contained in this file.</p>
<p><PRINCIPAL_INVESTIGATOR_NAME_LAST> </PRINCIPAL_INVESTIGATOR_NAME_LAST></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of last name of the Principal Investigator responsible for this dataset.</p>
<p><PRINCIPAL_INVESTIGATOR_NAME_FIRST> </PRINCIPAL_INVESTIGATOR_NAME_FIRST></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of first name of the Principal Investigator responsible for this dataset.</p>
<p><PRINCIPAL_INVESTIGATOR_NAME_TITLE> </PRINCIPAL_INVESTIGATOR_NAME_TITLE></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of title of the Principal Investigator responsible for this dataset.</p>

<p><PRINCIPAL_INVESTIGATOR_AFFILIATION> </PRINCIPAL_INVESTIGATOR_AFFILIATION></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of organizational affiliation of the Principal Investigator responsible for this dataset.</p>
<p><CO_INVESTIGATOR_NAME_LAST> </CO_INVESTIGATOR_NAME_LAST></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of last name of the Co-Investigator responsible for this dataset.</p>
<p><CO_INVESTIGATOR_NAME_FIRST> </CO_INVESTIGATOR_NAME_FIRST></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of first name of the Co-Investigator responsible for this dataset.</p>
<p><CO_INVESTIGATOR_AFFILIATION> </CO_INVESTIGATOR_AFFILIATION></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of organizational affiliation of the Co-Investigator responsible for this dataset.</p>
<p><COUNTRY_CODE></COUNTRY_CODE></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of internationally recognized character representation of a country name.</p>
<p><STATE_OR_PROVINCE_CODE> </STATE_OR_PROVINCE_CODE></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of two letter abbreviation for US state name.</p>

<p><PRINCIPAL_INVESTIGATOR_CONTACT_INFORMATION> </PRINCIPAL_INVESTIGATOR_CONTACT_INFORMATION></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of preferred contact information for the Principal Investigator.</p>
<p><DATA_USAGE_ACKNOWLEDGEMENT> </DATA_USAGE_ACKNOWLEDGEMENT></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of acknowledgement statement provided by Application Administrator.</p>
<p><NAME_OF_PERSON_WHO_GENERATED_THIS_FILE> </NAME_OF_PERSON_WHO_GENERATED_THIS_FILE></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of full name of individual responsible for the generation of this file.</p>
<p><AFFILIATION_OF_PERSON_WHO_GENERATED_THIS_FILE> </AFFILIATION_OF_PERSON_WHO_GENERATED_THIS_FILE></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of Organizational affiliation of the individual responsible for the generation of this file.</p>
<p><DATE_OF_LAST_MODIFICATION_TO_DATA_IN_MAIN_TABLE> </DATE_OF_LAST_MODIFICATION_TO_DATA_IN_MAIN_TABLE></p> <p>Required Parameters: format="MM/DD/YYYY" - or - format="MM/DD/YY"</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of date that the final changes to this file were made.</p>
<p><FILE_CHANGE_HISTORY_DESCRIPTION> </FILE_CHANGE_HISTORY_DESCRIPTION></p> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start and end of Descriptive text detailing any relative history of this file.</p>

<pre><table parameters></table></pre> <p>Required Parameters: name="tablename"</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate start of data input file data table section.</p>
<pre><columns parameters></columns></pre> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate the start of a group, or list of column names.</p>
<pre><column parameters></column></pre> <p>Required Parameters: reference_Table = "tablename" name="columnname"</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate the start of a column name.</p>
<pre><row></row></pre> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate the start of a new row within a table.</p>
<pre><data></data></pre> <p>Required Parameters: none</p> <p>Optional Parameters: none</p>	<p>Tags used to indicate the start and end of a data value.</p>

D. Example of Properly Formatted Tabular Data Input File

```
[bof]
[begin header]
FILE_NAME, "standard_application_format.csv"
DATA_EXCHANGE_VERSION, "Version number"
COMMENT, "Insert general comments"
QUALITY_CONTROL_LEVEL, "Quality Control level"
DATE_THIS_FILE_GENERATED, "Date"
DATA_SUBMITTER_USER_ID, "Assigned username"
ORGANIZATION_ACRONYM, "Assigned organization acronym"
ORGANIZATION_NAME, "Name of submitting organization"
STUDY_OR_NETWORK_ACRONYM, "Network Acronym"
STUDY_OR_NETWORK_NAME, "Network name"
FILE_CONTENTS_DESCRIPTION_SHORT, "Brief description of file contents"
FILE_CONTENTS_DESCRIPTION_LONG, "Detailed description of file contents"
PRINCIPAL_INVESTIGATOR_NAME_LAST, "Last Name"
PRINCIPAL_INVESTIGATOR_NAME_FIRST, "First Name"
PRINCIPAL_INVESTIGATOR_NAME_TITLE, "Title"
PRINCIPAL_INVESTIGATOR_AFFILIATION, "Organization Affiliation"
CO_INVESTIGATOR_NAME_LAST, "Last Name"
CO_INVESTIGATOR_NAME_FIRST, "First Name"
CO_INVESTIGATOR_AFFILIATION, "Organization Affiliation"
COUNTRY_CODE, "Country Code"
STATE_OR_PROVINCE_CODE, "State Code"
PRINCIPAL_INVESTIGATOR_CONTACT_INFORMATION, "Contact Information"
DATA_USAGE_ACKNOWLEDGEMENT, "Data use acknowledgement"
NAME_OF_PERSON_WHO_GENERATED_THIS_FILE, "Full Name"
AFFILIATION_OF_PERSON_WHO_GENERATED_THIS_FILE, "Organization Affiliation"
DATE_OF_LAST_MODIFICATION_TO_DATA_IN_MAIN_TABLE, "Date"
FILE_CHANGE_HISTORY_DESCRIPTION, "Change history description of this file"
[end header]
[begin table]
table_Name, "sites"
column_Names, "site_ID", "site_Code", "site_Abbreviation", "site_Name"
reference_Table, "N/A", "N/A", "N/A"
data_Row, "1", "HB", "HB", "Holbrook"
data_Row, "2", "LW", "LW", "Lawrenceville"
data_Row, "3", "STB", "STB", "Steubenville"
data_Row, "4", "NETL", "NETL", "NETL Supersite"
data_Row, "5", "PGHSPS", "PGHSPS", "Pittsburgh Supersite"
[end table]
[begin table]
table_Name, "data"
column_Names, "row_ID", "site_ID", "NOX"
reference_Table, "N/A", "sites", "N/A"
data_Row, "1", "4", "-999"
data_Row, "2", "4", "-999"
data_Row, "3", "1", "-999"
data_Row, "4", "3", "-999"
data_Row, "5", "2", "-999"
[end table]
[eof]
```

The preceding example would represent the following two data tables, plus header data:

sites

site_ID	site_Code	site_Abbreviation	site_Name
1	HB	HB	Holbrook
2	LW	LW	Lawrenceville
3	STB	STB	Steubenville
4	NETL	NETL	NETL Supersite
5	PGHSPS	PGHSPS	Pittsburgh Supersite

data

row_ID	site_ID	NOX
1	4	-999
2	4	-999
3	1	-999
4	3	-999
5	2	-999

E. Example of Properly Formatted XML Data Input File

```
<xml_Input_File>
<header>
<FILE_NAME>
standard_application_format.xml
</FILE_NAME>
<DATA_EXCHANGE_VERSION>
Version number
</DATA_EXCHANGE_VERSION>
<COMMENT>
Insert general comments
</COMMENT>
<QUALITY_CONTROL_LEVEL>
Quality Control level
</QUALITY_CONTROL_LEVEL>
<DATE_THIS_FILE_GENERATED>
Date
</DATE_THIS_FILE_GENERATED>
<DATA_SUBMITTER_USER_ID>
Assigned username
</DATA_SUBMITTER_USER_ID>
<ORGANIZATION_ACRONYM>
Assigned organization acronym
</ORGANIZATION_ACRONYM>
<ORGANIZATION_NAME>
Name of submitting organization
</ORGANIZATION_NAME>
<STUDY_OR_NETWORK_ACRONYM>
Network Acronym
</STUDY_OR_NETWORK_ACRONYM>
<STUDY_OR_NETWORK_NAME>
Network name
</STUDY_OR_NETWORK_NAME>
<FILE_CONTENTS_DESCRIPTION_SHORT>
Brief description of file contents
</FILE_CONTENTS_DESCRIPTION_SHORT>
<FILE_CONTENTS_DESCRIPTION_LONG>
Detailed description of file contents
</FILE_CONTENTS_DESCRIPTION_LONG>
<PRINCIPAL_INVESTIGATOR_NAME_LAST>
Last Name of Principal Investigator responsible for this dataset
</PRINCIPAL_INVESTIGATOR_NAME_LAST>
<PRINCIPAL_INVESTIGATOR_NAME_FIRST>
First Name of Principal Investigator responsible for this dataset
</PRINCIPAL_INVESTIGATOR_NAME_FIRST>
<PRINCIPAL_INVESTIGATOR_NAME_TITLE>
Title of Principal Investigator responsible for this dataset
</PRINCIPAL_INVESTIGATOR_NAME_TITLE>
<PRINCIPAL_INVESTIGATOR_AFFILIATION>
Organization Affiliation of Principal Investigator responsible for this dataset
</PRINCIPAL_INVESTIGATOR_AFFILIATION>
<CO_INVESTIGATOR_NAME_LAST>
Last Name of Co-Investigator responsible for this dataset
</CO_INVESTIGATOR_NAME_LAST>
<CO_INVESTIGATOR_NAME_FIRST>
```

First Name of Co-Investigator responsible for this dataset
 </CO_INVESTIGATOR_NAME_FIRST>
 <CO_INVESTIGATOR_AFFILIATION>
Organization Affiliation of Co-Investigator responsible for this dataset
 </CO_INVESTIGATOR_AFFILIATION>
 <COUNTRY_CODE>
Country Code
 </COUNTRY_CODE>
 <STATE_OR_PROVINCE_CODE>
State Code
 </STATE_OR_PROVINCE_CODE>
 <PRINCIPAL_INVESTIGATOR_CONTACT_INFORMATION>
Contact information for Principle Investigator responsible for this dataset.
 </PRINCIPAL_INVESTIGATOR_CONTACT_INFORMATION>
 <DATA_USAGE_ACKNOWLEDGEMENT>
Data use acknowledgement
 </DATA_USAGE_ACKNOWLEDGEMENT>
 <NAME_OF_PERSON_WHO_GENERATED_THIS_FILE>
Name of person who generated this file
 </NAME_OF_PERSON_WHO_GENERATED_THIS_FILE>
 <AFFILIATION_OF_PERSON_WHO_GENERATED_THIS_FILE>
Organization Affiliation of person who generated this file
 </AFFILIATION_OF_PERSON_WHO_GENERATED_THIS_FILE>
 <DATE_OF_LAST_MODIFICATION_TO_DATA_IN_MAIN_TABLE>
Date
 </DATE_OF_LAST_MODIFICATION_TO_DATA_IN_MAIN_TABLE>
 <FILE_CHANGE_HISTORY_DESCRIPTION>
Change history description of this file
 </FILE_CHANGE_HISTORY_DESCRIPTION>
 </header>
 <table name="sites">
 <columns>
 <column reference_Table="N/A">
 site_ID
 </column>
 <column reference_Table="N/A">
 site_Name
 </column>
 <column reference_Table="N/A">
 site_Abbreviation
 </column>
 </columns>
 <row>
 <data>
 1
 </data>
 <data>
 Holbrook
 </data>
 <data>
 HB
 </data>
 </row>
 <row>
 <data>
 2
 </data>

```

<data>
Lawrenceville
</data>
<data>
LW
</data>
</row>
<row>
<data>
3
</data>
<data>
Steubenville
</data>
<data>
STB
</data>
</row>
<row>
<data>
4
</data>
<data>
NETL Supersite
</data>
<data>
NETL
</data>
</row>
<row>
<data>
5
</data>
<data>
Pittsburgh Supersite
</data>
<data>
PGHSPS
</data>
</row>
</table>
<table name="data">
<columns>
  <column reference_Table="N/A">
    row_ID
  </column>
  <column reference_Table="sites">
    site_ID
  </column>
  <column reference_Table="N/A">
    NOX
  </column>
</columns>
<row>
<data>
1
</data>

```

```
<data>
4
</data>
<data>
-999
</data>
</row>
<row>
<data>
2
</data>
<data>
4
</data>
<data>
-999
</data>
<data>
3
</data>
<data>
1
</data>
<data>
-999
</data>
<data>
4
</data>
<data>
3
</data>
<data>
-999
</data>
<data>
5
</data>
<data>
2
</data>
<data>
-999
</data>
</row>
</table>
</xml_Input_File>
```

The preceding example would represent the following two data tables, plus header data:

sites

site_ID	site_Code	site_Abbreviation	site_Name
1	HB	HB	Holbrook
2	LW	LW	Lawrenceville
3	STB	STB	Steubenville
4	NETL	NETL	NETL Supersite
5	PGHSPS	PGHSPS	Pittsburgh Supersite

data

row_ID	site_ID	NOX
1	4	-999
2	4	-999
3	1	-999
4	3	-999
5	2	-999

APPENDIX 7 – QUALITY CONTROL DEFINITIONS

VALUE	DESCRIPTION
0	A reasonably complete dataset of unspecified quality and subjected to minimum processing in the field and/or in the laboratory by data collection staff. Metadata is assigned to at least 75% of the values in the dataset.
1	A complete dataset of specified quality that has been subjected to quality assurance and quality control checks and data management procedures. Metadata is assigned to at least 90% of all values in the dataset.
2	A complete, externally consistent dataset of specified quality that consists of research products that have undergone interpretative and diagnostic analyses by the submitting agency, organization or group. Metadata is assigned to all data (100%), with researcher comments assigned to values within the dataset.
3	Data that have received intense scrutiny through analysis or use in modeling. Metadata assigned to all values (100%) in the dataset, along with researcher comments and notes regarding past analyses.
-9	Data submitted with no indication of the quality control level. No metadata present.

APPENDIX 8 – Default Security Levels

BROWSE LEVEL	DESCRIPTION
0	None. Browsing of nodes denied.
1	unused
2	unused
3	Normal User. Granted limited permission to browse specific nodes only.
4	Stakeholder User.
5	Site User. Granted permission to browse all site-related nodes.
6	Project Team User. Granted permission to browse all nodes.

READ LEVEL	DESCRIPTION
0	None. Reading of nodes denied.
1	unused
2	unused
3	Normal User. Granted limited permission to read specific nodes only.
4	Stakeholder User.
5	Site User. Granted permission to read all site-related nodes.
6	Project Team User. Granted permission to read data from all nodes.

WRITE LEVEL	DESCRIPTION
0	None. Saving of data is denied.
1	unused
2	unused
3	Normal User. Save permissions limited to comments area.
4	Stakeholder User. Save permissions limited to assigned areas.
5	Site User. Save permissions allowed on all site-related data.
6	Project Team User. Granted permission to write data to all nodes.

ADMIN LEVEL	DESCRIPTION
0	None. No administration permissions.
1	unused
2	unused
3	Normal User.
4	Stakeholder User.
5	Site User.
6	Project Team User. Access granted to all Extranet applications.