CONTRACT NO. 93-328
FINAL REPORT
JUNE 1996

Development of a Pilot Air Quality and Meteorological Information System for the Greater San Joaquin Valley

CALIFORNIA ENVIRONMENTAL PROTECTION AGENCY

AIR RESOURCES BOARD
Research Division

Development of a Pilot Air Quality and Meteorological Information System for the Greater San Joaquin Valley

Final Report
Contract No. 93-328

Prepared for:

California Air Resources Board
Research Division
2020 L Street
Sacramento, CA 95814

Prepared by:

Dr. Len Fisk, Professor
Department of Computer Science
California State University, Chico
Chico, CA 95929

June 1996

Disclaimer

The statements and conclusions in this report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

Acknowledgments

Throughout the life of this project, a number of students at CSU, Chico contributed significantly to the progress made. I would like to acknowledge the excellent and thoroughly professional work provided by Kevin Freeman, Cheuk Chau, Nick Trown, Ryan Roffelsen, Terry Little, and Ron Pike.

This report was submitted in fulfillment of CARB contract number 93-328, entitled "Development of a Pilot Meteorological and Air Quality Information System for the Greater San Joaquin Valley," by California State University, Chico under the sponsorship of the California Air Resources Board. The work on this contract was completed as of June 30, 1996.

Abstract

This project was undertaken to address problems in sharing current meteorological and air quality data for the San Joaquin Valley Air Basin. The project assessed the information needs of various agencies and groups in the San Joaquin Valley, including the San Joaquin Valley Unified Air Pollution Control District, the Air Resources Board, the U.S. Forest Service, and others. The project also assessed the availability of various sources of real-time meteorological and air quality data. The needs defined by the study were then used to guide the development of a pilot Air Quality and Meteorological Information System (AQMIS) for the region. The pilot system developed for this project consists of two Internet-connected PCs, both running Windows NT and a body of software developed to acquire and disseminate hourly, real-time, meteorological and air quality data. The project team used Microsoft SQL Server NT, Microsoft Visual C++ and PERL for NT to develop the software necessary to achieve its objectives. The software products include: (1) a scheduler and "data grabber" which acquires real-time meteorological and air quality data on an automated schedule via both "FTP" on the Internet and via modem telephone connections to data loggers and other data sources; (2) translation scripts, written in PERL, that convert the many different data formats that exist at the multiple data sources to the standard, "normalized" data format of the SQL Server relational database, and (3) a "run time" meteorological and air quality map builder, written in C and translated to Visual C++, that operates in concert with Microsoft's web server for Windows NT to build both color maps and data tables in response to queries that can be posed via any Internet web browser that is HTML 3 compatible. The pilot AQMIS provides an engine that permits ARB to collect real-time data from a great many disparate data sources and disseminate those data in either graphic or tabular form to any person or organization that has a "web browser." The system, as designed, can be expanded to include a much larger territory.

Table of Contents

Introduction

Organization of the report

This Air Quality and Meteorological Information System (AQMIS) project culminated in the development of a pilot system for the acquisition and dissemination of data for the greater San Joaquin valley. This project to develop an AQMIS does not easily fit within the confines of a traditional empirical research project. Consequently, the "**Introduction**" section is used to discuss the initial reasons for embarking on the project, the planned project stages, and briefly, the actual development stages for AQMIS; the "**System Development**" section is used to detail the hardware and software environment and the tools used to build the AQMIS software suite; the "**Results**" section is used to describe the final products and their relationship to the initial requirements; the "**Summary, Conclusions and Recommendations**" section contains a brief summary of the project team's conclusions about the capabilities of the AQMIS software suite; and its recommendations for both short-term and long-term follow-on efforts that were defined as a result of the AQMIS pilot project.

Project purpose

The AQMIS pilot project focused on current statewide problems in the acquisition, storage, sharing and analysis of real-time meteorological and air quality data. The ultimate objective was to develop a pilot program to deal with data from the San Joaquin Valley Air Basin and adjacent geographic areas which influence and are influenced by air circulation, pollutants and other factors. Specifically, the areas covered include the San Joaquin Valley Air Basin; Calaveras, Tuolumne and Mariposa Counties; the Sacramento metropolitan area; and the San Francisco Bay Area. The pilot program focused on air quality and meteorological information and was ultimately embodied in this Air Quality and Meteorological Information System (AQMIS).

The need for sharing real-time meteorological and air quality information among the various agencies and groups has existed for some time. However, the events which led directly to this AQMIS project were related to the impacts of planned burns of forests and agricultural fields. Specifically, requests from Mariposa County Supervisor Arthur Baggett and State Senator Dan McCorquodale led to the development of this pilot project, which will enable neighboring areas to better plan their burns.

A number of agencies, including the Air Resources Board, the San Joaquin Valley Unified Air Pollution Control District, the California Department of Forestry, the U. S. Forest Service and others have developed data collection and distribution networks that do little to encourage the dissemination of data outside the confines of the developing agency. The task of making these varied data available in a standard format for general use was made more difficult by their varied availability, format, and method of collection

and distribution. The advent of advanced computer analysis and modeling techniques that can make good use of large bodies of data suggested that the development of a centralized system with data gathering and distribution capability would augment our capability of using a large body of current, up-to-the-minute meteorological and air-quality data to make improved nowcasts and predictions of pollution levels and the development of more effective plans for air quality improvements.

Although the AQMIS project did not focus primarily on the development of models and modeling techniques, it does provide immediate access to a wealth of data that can be used in the development and testing of such models. As such, AQMIS may be viewed as a platform from which both researchers and operational forecasters and policy makers may develop more accurate and reliable models of air pollution transport and meteorological influences. The system is a tool which offers valuable "real-time" information to operational forecasters, operational planners and policy makers, and environmental disaster recovery managers.

Summary of Proposed Study Phases

The Project Phases were defined as follows:

Phase Description

I. Prepare Preliminary Requirements
   - A. Identify known sources and potential users of the pilot San Joaquin Valley Air Basin Air Quality and Meteorological Information System (AQMIS).
   - B. Interview knowledgeable staff of likely data providers and users to determine meteorological and air quality data acquisition and dissemination needs for AQMIS.
   - C. Document the detailed requirements for AQMIS, including the data polling needs for acquisition and short- and long-term retention of all data and software to convert, summarize and analyze these data.

II. Develop detailed requirements for the pilot AQMIS.
   - A. Work with ARB to determine the detailed requirements for phase 4.

III. Design and implement the pilot system.
   - A. Prepare specifications for hardware and off-the-shelf software, match the specifications to the detailed requirements and prepare detailed performance specifications for all software to be developed by the project team.
   - B. Acquire hardware and off-the-shelf software.
   - C. Develop and test data acquisition and conversion software.
   - D. Develop and test data dissemination software.

IV. Train appropriate staff in the use of the system.

V. Evaluate the efficacy of the pilot system and suggest further actions to be taken.

Bill Wilson of the ARB's Meteorology Section was assigned to be the ARB liaison person and a project steering committee was constituted to assist in making decisions at

the various stages of developing the pilot AQMIS project. Leon Dolislager acted as project manager throughout the life of the contract. Progress reports were provided to Mr. Dolislager and Mr. Wilson at all phases of development, and the AQMIS Steering Committee was called upon to participate in key decisions regarding the project.

An overview of the project's history

The initial phases of the project were marked by a number of delays. The initial requirements analysis for the system was to be completed during the summer of 1994. Unfortunately, a protracted delay in the approval of the contract prevented the project team from doing this work in the summer of 1994, and delayed the major portion of this work until the summer of 1995.

A second delay was occasioned by a very long shipping time for the first computer for project development work. Ordered in February of 1995, the machine did not arrive until the end of June, due, we surmised, to problems with the Pentium processor that backlogged the shipping of Pentium-based products at Hewlett-Packard.

Dramatic changes in the development environment have occurred throughout the life of the contract. Some of these changes were occasioned by major changes in the availability of data and the means by which data could be obtained. For example, as the systems of various agencies became more automated, more data became available in summary form via File Transfer Protocol (FTP) over the Internet, and the need to access data via modem became less important. Still other changes became necessary in order to comply with an informal, emerging standard for database servers and web servers, and the increasing availability of Internet access.

System Development

Preparing preliminary requirements

As the project team began its work during the fall and winter of 1994, it began the preliminary definition of project requirements. During the initial stages of the study, the team met with the steering committee that was formed for the purpose of directing the development work, and met with a number of organizations and individuals that could either supply or utilize the data to be gathered by the AQMIS system.

Determining data sources

In August of 1994, the project team immediately began to contact known data sources and potential users of the AQMIS and to interview them to determine meteorological and air quality data acquisition and dissemination needs. As these needs began to be defined, a pattern of change began to emerge: while the initial assumption was that the data sources and needs would remain relatively static during the time of the study, it became increasingly apparent that the acquisition and dissemination of data was shifting dramatically from the use of modems to the use of the Internet.

Another change in plans followed the observation that the reliability of data sources with regard to the availability of data was far less than had been supposed at the outset. The initial assumption was that a set of reliable sources of data could be used throughout the life of the project. This somewhat naive assumption was replaced by the more realistic observation that the availability of data fluctuated almost monthly and that many data sources, like the San Joaquin Valley Air Pollution Control District, were in the process of modifying the means of obtaining data from individual "data logger" access via modem to Internet access of data summaries via "File Transfer Protocol" (FTP). Still other sources that were promised early in the cycle became unavailable when it came time to obtain the data to generate translation scripts to add them to the real-time AQMIS database.

In short, the foundations upon which the initial planning was conducted were rapidly shifting and required a paradigm shift in the manner with which the building of the pilot AQMIS was conducted. The project team's immediate response to the shifting requirements for data acquisition and dissemination was to propose to build a more open-ended and flexible set of data acquisition and dissemination tools, and to focus more closely on the development of Internet capabilities that allowed the FTP transfer of data files and the dissemination of air quality and meteorological information via World Wide Web documents (i.e., via HTML pages) accessed via "web-browsers." An altered plan of attack, with an altered set of objectives was presented and approved at a meeting of the Steering Committee in January of 1995.

Hardware platforms

At the outset of the study, the project team had planned to acquire a single workstation to carry out basic data acquisition and dissemination tasks. This plan was amended to include a second workstation as the sophistication of the processing requirements escalated during the early stages of the project. The first incarnation of the plan to purchase a second processor required the purchase of an HP 9000, to be administered in partnership with the Department of Water Resources because they had staff members experienced in maintaining such machines, and a PC running Windows NT, to be maintained at the ARB Meteorological section. When preliminary efforts to arrive at a mutually satisfactory plan failed, it was decided to purchase two PCs, one a dual processor Pentium running Windows NT Server, the other a single processor Pentium running Windows NT Workstation. It was determined that the dual processor Pentium would provide processing power comparable to that of a UNIX workstation, but would

allow easier local maintenance by staff who had a relatively less sophisticated level of training.

The choice of PCs and Windows NT proved to be a wise one because the platform proved to integrate well with the MARTA meteorological and satellite picture acquisition system already in service in the Meteorology section at ARB.

Software development environment

The initial choices of software development tools were dictated by the need to be consistent with standards already developed in other groups at ARB. Because the data acquisition engine needed to be capable of multi-tasking, the project development team selected the newest version of Microsoft Visual C++. Visual C++ has an excellent graphic development interface, permits multi-tasking, is object-oriented, and compiles as "native" code that executes very quickly on the PC.

The project team discovered that the sources of data for AQMIS would be in constant flux. Consequently, the development of the "scripts" with which to access and translate the data of each source was restricted to a non-compiled "scripting language." Because a great many scripts that translate among different meteorological data format have already been developed in PERL, and because an academic and research version of PERL had been developed for the NT environment, PERL was adopted as the scripting language of choice. PERL has the advantage of permitting relatively easy modification of instructions compared to C++ and other compiled languages.

The model for a solution of data dissemination problems that came the closest to those posed by the present study was the research data system maintained by the Modeling group at ARB. Because this group used Visual Foxpro for their database server, this system was purchased for the present project. The HTML server chosen for the project was the new Netscape HTML Server.

All of the chosen software was acquired, installed, and used for software development through the development phase of the present project until April, 1996.

Last minute changes in development software

In April, 1996, the project team discovered that an effort to develop a web page for ARB was proceeding in parallel with the present project. Because the team was concerned that the maintenance of the system be as simple and generic as possible, the team, with the assistance of several members of the ARB staff, decided to acquire a new database server and new HTML server. Specifically, the team decided to switch to the Microsoft HTML server and Microsoft SQL Server for Windows NT. Although this introduced some confusion at the last minute, the advantages in simplicity of maintenance were deemed to have sufficient merit to justify a great deal of last minute code modification.

The final hardware and software environment

The final development environment consisted of the following hardware and software:

- one H-P dual processor Vectra (network name = Pokey) with dual 90 MHz. Pentium processors, 32 Megabytes of RAM, one Ethernet card, two - 1.2 Gigabyte Hard Disks, one 250 Megabyte streaming tape drive, and one 3.5" floppy drive, running Windows NT 3.51 Server, with mirrored drives;
- one Gateway 2000 Pentium 130 with 32 Megabytes of RAM, one Ethernet card, 2 - 28.8 modems, two - 2 Gigabyte drives, and one - 3.5" floppy drive, running Windows NT 3.5;
- one - 20 workstation license for NT clients;
- one copy of Microsoft Visual C++;
- one copy of Microsoft SQL Server for NT (client hookups already licensed within ARB);
- one copy of Microsoft HTML server for NT; and
- one "freeware" copy of PERL for NT.

Specific applications of the software development environment

The AQMIS comprises two computers with some specialized hardware capabilities and an integrated suite of software. Some of the software has been purchased for the project, and much has been written expressly for the AQMIS. The AQMIS project staff has developed three major programs/resources:

- the "data grabber",
- the AQMIS database, and
- the "map presenter," plus
- a number of ancillary programs which work in concert with these major programs.

The data grabber development environment

The "data grabber" program was written in Visual C++ for Windows NT and makes maximum use of that language's capability of "spawning" multiple "threads" to make the task of retrieving real-time data from multiple sources a concurrent, multiple-tasked process. The "data grabber" has three major components: (1) the scheduler, (2) data acquisition scripts, and (3) data conversion scripts.

The scheduler and data acquisition scripts

The heart of the C++ "data grabber" program is a scheduler, which, at prescribed times every hour, initiates, or "spawns," a process to obtain and translate data from each source, either via Internet FTP transfer process or via modem data exchange with individual data loggers. The multiple threaded nature of these processes allows them to execute concurrently, thereby allowing many such processes to occur at the same time. This maximizes the "throughput" of the system, and permits a maximum capacity that is not limited to the geographic area defined for this study.

The first script to be "called" by each thread is a "data acquisition script". Each data source generally must have a unique PERL script that permits the system to access that particular computer or data logger. In the case of an agency with multiple data loggers to be accessed (San Joaquin Valley Unified Air Pollution Control District, for example), a separate script must exist for each logger, and the access of each logger is a separately scheduled event for the scheduler. Each site for summarized data must also have an individually tailored script.

Although the scheduler can accommodate the addition or deletion of new data sites without being recompiled, a new script must be tailored for each new site. The modification of an access protocol at any site will also require a new script.

"data grabber" and data conversion scripts

As soon as the "data grabber" has succeeded in obtaining new data for the hour, the new data must be converted to a form that is consistent with all of the data being inserted into the AQMIS database. Because there is so little standardization among sources, and because each new source will most probably have idiosyncratic data format choices, it was decided to use PERL scripts, launched by the C++ "threads" to accomplish this task as well. As is the case for the data acquisition scripts, these scripts must also be tailored for each new source of data added to the system.

AQMIS database

Initially the AQMIS database was designed for Microsoft Visual Foxpro for Windows NT. This development tool was chosen because it offered a very flexible graphic design and development interface, and because it was being used by the only similar system being used in-house at ARB -- the modeling group database being developed and maintained by Rich Hackney. This initial choice was overturned in April of 1996, and a last minute conversion to Microsoft SQL Server for NT was made. This change occurred in response to a new standard imposed by an HTML-to-database interface developed to represent ARB on the web by yet another ARB development team.

The new database management system is capable of accommodating all of the design work the AQMIS team had done before. Because the AQMIS data were all rendered in "third normal form," and because all of the software that accessed or updated the database was developed in a highly modular fashion, the conversion was not as disruptive as it might have been. Although the conversion of the HTML server and the database management system during the final months of the project was disruptive, the long term benefits to the maintenance of the system were sufficient to warrant the change.

"Map Presenter"

Early in the project, the development team determined that the Internet would become the major pathway for data communication before the project ended. During the two years that the AQMIS pilot project has been developed, the wisdom of this early decision has become increasingly apparent. The "Map Presenter" was designed to make full use of this new technology, and provides Internet access to the AQMIS database in two different forms: (1) a tabular display of data specific to a measure, to a set of locations and to a span of time; and (2) a graphic, map display of air quality and meteorological data that can be varied by type of display, map scale, map center, measure and geographic map features.

The initial version of this program was begun before the project had obtained a development machine, so it was written in the C language for a UNIX HTML server running on an HP 9000-735 with HPUX. As soon as an NT development machine became available, the C code developed for the UNIX C compiler was "ported" to the NT environment and translated and compiled by the Visual C++ compiler. The first "port" of the software was designed to interface with the Netscape HTML server for Windows NT that was used during the major portion of the project. A second "port," or "translation" was required when the project shifted to the use of the Microsoft HTML server for NT. As was the case for the database management system, the shift to the Microsoft HTML server reflected a desire to standardize the software to permit the sharing of maintenance expertise and responsibility across groups within ARB. The project team decided that the value of such a shift outweighed the inconvenience.

Results

The project began with a fixed set of objectives. Although the project team's perspective of these objectives was much altered during the early stages of the project (see Appendix B for a full report of the changes in objectives that were called for by the project team and approved by AQMIS Steering Committee), the initial objectives can still serve as an effective means for organizing a discussion of the assessment of whether the project met its stated goals.

Objective 1. Prepare a Preliminary AQMIS Requirements Document.

Objective 1.a. Identify all sources and potential users of the pilot San Joaquin Valley Air Basin Air Quality and Meteorological Information System (AQMIS).

Objective 1.b. Interview knowledgeable staff to determine meteorological and air quality data acquisition and dissemination needs for AQMIS.

Objectives 1, 1.a. and 1.b. have been effectively met by the project. During the initial months of the contract, a great many potential sources of real-time weather

data were identified. These sources were identified by the AQMIS Steering Committee, by the California Air Pollution Control Officers Association, by contacts with the members of the Interagency Air and Smoke Council, by contacts with meteorologists working with agencies and private companies throughout the San Joaquin Valley and environs.

It soon became apparent that the availability of data was an ever shifting thing. Data that were promised for project use from such agencies as Water Resources and other air resources groups earlier in the interview cycle were withdrawn when it became clear that these data could be disseminated to all of the air quality districts. The team's response was to build a system that permitted a maximum of flexibility in its ability to adapt to ever changing data availability. The core elements of the AQMIS system remain unchanged, while additional data sources can be easily incorporated with the addition of new data acquisition and conversion scripts to the system.

Objective 1.c. Document the detailed requirements for AQMIS, including the data polling needs for acquisition and short- and long-term retention of all data and software to convert, summarize and analyze these data.

Because the detailed data acquisition and conversion needs for individual data sources were so mercurial, it was determined that a one-time summary would have served little purpose. In truth, the current set of PERL scripts that accomplish these tasks are as objective a source of documentation as can be compiled.

The detailed data summary and analysis needs included a great many factors, including analysis of upper air data. Because the most fundamental needs expressed were for simple graphical and tabular summaries of surface air quality and meteorological data for specific areas, time spans and measures, this was the chosen focus of the project. The inclusion of upper air modeling was ruled to be beyond the scope permitted by the budget of the AQMIS pilot study.

The measures included in the analysis are defined within the primary HTML "map presenter" screen, which can be accessed via the Internet and an HTML 3.0 web browser (the URL address is: http://www.aqmis.arb.ca.gov).

Objective 2. Develop detailed requirements for the pilot AQMIS.

The requirements for the AQMIS pilot project underwent major change as the project development team delved further into the problem. In January of 1995, a major refinement of requirements was approved by the AQMIS Steering Committee. The major thrust was to provide an HTML, Internet connection with

the system's users, and augment this with a modem-based Bulletin Board System (BBS) capability, using the standards supplied by the Modeling group at ARB. The HTML component has been completely developed, although many of the standards have been supplanted by those adopted by another group within ARB (see "**Last minute changes in development software**", above).

The one component that was not adequately developed was the modem-linked BBS. The development team lacked both the hardware resources and the personnel to develop both the BBS and the HTML access components in the pilot project. As the Internet emerged as a universal data access tool during the course of the project, it was decided to develop this portion of the project rather than the BBS portion. As BBS capabilities are offered by other systems on the Local Area Network (LAN) at ARB, it is quite feasible that a BBS capability could be developed with relative ease, once the existing BBS has access to the AQMIS server on the LAN. In the meantime, data "downloads" are provided for via the HTML server's FTP capability, and the "Map Presenter's" ability to provide tabular listings of data.

Objective 3. Work with ARB to determine the detailed requirements for phase 4, adding or subtracting products and stages as required.

The development of the pilot system required a great many false starts and directional changes in response to shifting requirements, shifting data availability, and an ever changing technology. The final list of AQMIS capabilities reflects these changes and was arrived at in close collaboration with ARB.

Objective 4. Implement the pilot system.

Objective 4.a. Prepare specifications for hardware and off-the-shelf software, and prepare detailed performance specifications for all software to be developed by the project team.

Objective 4.b. Acquire Hardware and off-the-shelf software.

Objectives 4., 4.a., and 4.b. were met as the performance specifications of the off-the-shelf software was defined in stages by the project team. The specifications of both hardware and software reflected the following concerns:

- the need to adhere to informal and emerging ARB standards;
- the need for high performance at a low cost;
- the need to maintain the AQMIS system with as little involvement of technical professionals as possible;
- the need for the system to respond flexibly and frequently to changes in data sources and formats;

- the need for a data dissemination interface that is readily available and does not require specialized software; and
- the need for a system which is modular enough and "scaleable" enough to permit the wholesale shifting of system modules among different LAN servers and a "scaling up" of services to include the whole of the State of California.

The "off-the shelf" software chosen for the project serves all of these needs.

Objective 4.c. Develop and test data acquisition and conversion software.

The data acquisition and conversion software consists of two different categories of programs: (1) the scheduler, and (2) the "scripts" which obtain the data and convert it to the proper format for the AQMIS database. Of the two categories, the operation of the scheduler is the more critical, because it is compiled in C++ and cannot easily be amended, and because the logic of its operation is more complex and "multi-threaded." By their very nature, PERL scripts can be readily modified, and will be continuously modified as the data sources and formats change.

Both the scheduler and a set of sample scripts that control both modem access of data loggers and Internet FTP transfers of data were tested continuously for many months. This testing will continue as new scripts are added to the system's repertoire.

At the present time, the scheduler is capable of running without failure for weeks at a time, and its failures are generally affected by external events such as power surges and/or failures.

Objective 4.d. Develop and test data dissemination software.

The compiled C++ code that links with the HTML server to provide "on-the-fly" construction of maps of meteorological and air quality data via the "web" has existed in one form or another since it was first written for this project more than a year ago. Since that time the code has undergone continual change, first to "port" it from C and UNIX to C++ and NT, then to accommodate a new HTML server, and, finally, to accommodate a different database server. Although almost all of the obvious "bugs" have been identified and eradicated, some final capabilities have not been tested under heavy usage. One such capability is the computation of wind charts.

To date, it appears that all elements of the HTML user interface are functioning quite well.

Objective 4.e. Train appropriate staff in the use of the system.

To date, only two individuals have been partially trained in the maintenance of the AQMIS pilot system. The primary reason for the last minute conversions to a different database server and a different web server was to conform with another system currently in use within ARB, and thereby increase the base of expertise applicable to the maintenance of AQMIS.

The task of providing an interface that would allow a relatively unskilled person to prepare the scripts necessary to access disparate systems and convert data of disparate formats to the standard format used by AQMIS is beyond the current technology. Therefore, it will be necessary to retain the part-time services of a person trained in the modification of PERL scripts and knowledgeable in the interface of these scripts with the scheduler and the portion of the database that contains the scheduling information. Much more can be done to make the task of amending the system to deal with new data sources and formats a friendlier one, but the portion of the task that deals with modifying the PERL scripts cannot be simplified at the present time.

Objective 4.f. Evaluate the efficacy of the pilot system and suggest further actions to be taken.

The project team's evaluation of the pilot system and suggestions for further actions appear in the "Summary, Conclusions and Recommendations" section.

Summary, Conclusions, and Recommendations

This project was undertaken to address problems in sharing current meteorological and air quality data for the San Joaquin Valley Air Basin. The project assessed the information needs of various agencies and groups in the San Joaquin Valley, including the San Joaquin Valley Air Pollution Control District, the Air Resources Board, the U.S. Forest Service, and others. The project also assessed the availability of various sources of real-time meteorological and air quality data. The needs defined by the study were then used to guide the development of a pilot Air Quality and Meteorological Information System (AQMIS) for the region. The pilot system developed for this project consists of two Internet-connected PCs, both running Windows NT and a body of software developed to acquire and disseminate hourly, real-time, meteorological and air quality data. The project team used Microsoft SQL Server NT, Microsoft Visual C++ and PERL for NT to develop the software necessary to achieve its objectives. The software products include: (1) a scheduler and "data grabber" which work together to acquire real-time meteorological and air quality data on an automated schedule via both "FTP" on the Internet and via modem telephone connections to data loggers and other data sources; (2) translation scripts, written in PERL, that convert the many different data formats that exist at the multiple data sources to the standard, "normalized" data format of the SQL Server relational database, and (3) a "run time" meteorological and air quality map builder, written in C and translated to Visual C++, that operates in concert with Microsoft's web server for Windows NT to build both color maps and data tables in response to queries that can be posed via any Internet web browser that is HTML 3 compatible. The pilot AQMIS provides an engine that permits ARB to collect real-time data from a great many disparate data sources and disseminate those data in either graphic or tabular form to any person or organization that has a "web browser." The system, as designed, can be expanded to include a much larger territory.

Data Sources

At present, the AQMIS system is acquiring real-time data from a number of data-loggers maintained by the San Joaquin Valley Unified Air Pollution Control District and also accesses real-time data available via the MARTA system services leased by the Meteorology Section at ARB.

The San Joaquin Valley Unified Air Pollution Control District data loggers are accessed via PERL modem scripts that are run by the scheduler and data grabber. As both modems are used concurrently by multiple processes, the task is carried out in a relatively efficient manner. However, the San Joaquin Valley Air Pollution Control District has announced its intention of making summary data available via "FTP" on the Internet in the near future. The increasing availability of summary data will radically reduce the time required by the AQMIS to access San Joaquin data, and will free system resources. The net effect will be to permit the expansion of the system to incorporate data from a greater number of sources and sites.

The MARTA data are acquired via an "FTP" process on ARB's Local Area Network (LAN). The FTP process is capable of moving a great deal more data than a modem during a set period of time. Because the "FTP" process is standard throughout the Internet, it will be possible to acquire data from sources outside ARB's LAN.

Accessing additional data via the "web"

The project team has prepared PERL translation scripts for a number of additional data sources available on the Internet in order to prove the feasibility of using such sources. It will, for example, be a simple matter to add "day-old" data from the California Irrigation Management Information System (CIMIS) if it is determined that this source would be of use to the users of AQMIS data.

As the team tested the "data grabber," it proved to be capable of "grabbing," via FTP, as many as forty large data files per hour. With such data acquisition capabilities, the AQMIS developed in this project is fully capable of maintaining data for the whole of the State of California. The only apparent limitation to the acquisition of additional data is the availability of the data and the ability of the ARB maintenance team to develop PERL data access and translation scripts.

The format chosen by the project team is one that is universally available at a very low cost; "web browsers" like Netscape Navigator and Mosaic have become readily available to any user, and inexpensive, modem-based web access now costs less than $20 per month. Although users who are limited to the speeds imposed by a 28.8K modem will find the interface more frustrating than users who have a high-speed data link, the team's expectation is that the user-friendly format of the web page and the quality of data presentations will prove to be attractive and useful to air quality and meteorological data users throughout the state.

The importance of providing data as well as "consuming" it

As more districts that acquire their own real-time data begin to use the AQMIS, it is hoped that they will be more likely to make their data available to others as well. The administrators of the ARB web site have the option of password protecting some or all of the information provided by AQMIS. Although it would be advantageous to permit all users open access to all data initially in order to popularize data usage, it might ultimately be useful to limit data access to those users who provide real-time data to the AQMIS. Such policies can be instituted only after ARB has acquired experience in offering the services over a long span of time.

The AQMIS project has demonstrated that an automated system can collect and disseminate real-time meteorological and air quality data in an extremely user-friendly way. Furthermore, it has been shown that such a system can be built so as to require only

a small measure of technical expertise for maintenance in its long-term technical maintenance.

Recommendations for continued maintenance and development

Several suggestions for future development can be made.

New PERL scripts must be developed as part of a continuing maintenance process. The data sources and formats will be certain to change. When such changes occur, AQMIS will most certainly require the assistance of a person well versed in amending PERL scripts and including new sites in the scheduler's hourly rounds. The project team estimates that this need should be no more than 3 to 4 workdays per month, and can be carried out by a trained student assistant.

FTP access to real-time AQMIS data tables must be provided. The new database and web server software has a built-in capability to allow web page-based SQL queries to "FTP" data via the Internet. Such a capability would allow any user, perhaps limited by password protections, to extract any portion of the on-line data to his or her machine. The recent development of this software capability and the even more recent acquisition of this software (April, 1996) has prevented the project team from including this capability in the designed interface to AQMIS. We recommend that this capability be developed by a student assistant during the next fiscal year.

The performance of the system must be tuned as its usage increases. The system, as developed, will require some performance tuning. This will entail some shifting of software modules between the workstation machine (Gumby) and server machine (Pokey). The software was developed in a completely modular fashion so as to permit such shifting of software.

The area served by the system must be "scaled up" to cover a greater portion of California. The project team believes that the AQMIS software and hardware will prove robust enough to permit the gradual "scaling up" of services to include larger regions in the database. We recommend the this "scaling up" be planned to occur during the next two years, with the final objective being the inclusion of the whole of the State of California.

The maintenance process must address the many problems introduced by the lack of standards. One problem encountered by the development team is that the many different sources of data use different three-letter names to describe data collection sites. There are many homonyms and synonyms among the three letter codes that describe station locations. The AQMIS software sidesteps the issue by requiring two three-

character sequences -- one to identify the station and the other to identify the agency that supplied the information -- thereby allowing the synonyms and homonyms to coexist on the same system. This is a messy solution to a problem that requires the definition of a true standard.

Other problems stem from the lack of clear standards. The existence of multiple data formats is one clear example -- if only one format existed, it would be unnecessary to create a new translation script for every new data source added to the system.

Glossary of Terms, Abbreviations and Symbols

AQMIS	- Air Quality and Meteorological Information System
ARB	- California Air Resources Board
BBS	- Bulletin Board System: a system whereby users can "log on" to a server and either post or read messages from other users
C	- the language most commonly used for software development in the software industry -- the precursor of C++
C++	- an object-oriented language that is standard in the software industry
compiler	- a program that translates high level language code to machine code so that the machine is capable of executing the code
database server	- a LAN-connected computer that allows "client" computers to access a common database that it maintains
Foxpro	- a Microsoft database software package that runs on PCs and Macintoshes
FTP	- File Transfer Protocol: a specific kind of file transfer that is standard in UNIX and on the Internet
Gumby	- The name given the 120 MHz. Gateway computer on the LAN for the College of Engineering, Computer Science and Technology at CSU, Chico.
HP 9000-735	- a particular model in the HP 9000 line of UNIX workstations
HPUX	- Hewlett-Packard's version of UNIX
HTML	- Hypertext Markup Language: a standard "markup" or "formatting" language that is

	used to provide formatted displays via the "World-Wide Web"
Internet	- a conglomeration of thousands of computers that are interconnected according to a standard initially developed for research use with UNIX machines
interpreter	- a program that "interprets, in a line-by-line manner, the high level language code to machine code so that the machine can execute the code -- this is a much slower process than executing compiled code
LAN	- Local Area Network: a collection of computers that are interconnected in the same building
MARTA	- a particular vendor that provides meteorological software and data
MHz	- Megahertz: a measure of a processor's clock speed in millions of cycles per second
Microsoft SQL Server for Windows NT	- a database server that provides SQL access capability via the "Open Database Connectivity" (ODBC) standard
multi-tasking	- a computer that "multi-tasks" can carry out multiple operations simultaneously, much as mere humans can walk, chew gum and carry on intelligent conversations all at the same time
multiple threaded	- a particular way of implementing multi-tasking
PERL	- a scripting language, common in the UNIX world, which provides a simpler development environment than a compiled language like C++
Pokey	- The name given the dual processor HP computer on the LAN for the College of Engineering, Computer Science and Technology at CSU, Chico.

RAM	- Random Access Memory
server [NT]	- the NT version, and machine on which it is loaded, that provides services for the NT workstation
SQL	- Structured Query Language: a non-procedural, "fourth generation" database query language
UNIX	- the most commonly used operating system for machines larger than the PC
URL	- Uniform Resource Locator: in practical usage, the URL is the address of a "web site" or page
web server	- an Internet-connected computer that runs specific software that enables it to provide HTML pages for viewing on the "World-Wide Web"
Windows NT	- a Microsoft operating system that is intended to compete with UNIX while offering the standard Microsoft Windows interface
workstation [NT]	- the NT version, and machine on which it is loaded, that is connected to the NT server

Appendix A: Using AQMIS

If you wish to use AQMIS, (1) you need to be Internet-connected, and (2) you should be using a reasonably-up-to-date browser like Netscape 2.0 or higher, or its Microsoft equivalent.

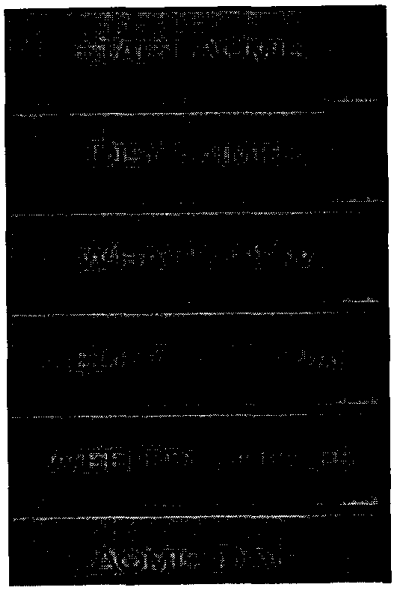
The "URL" or web page address at which AQMIS resides is, at present, *http://pokey.ecst.csuchico.edu.* Entering this code will connect you directly to the AQMIS server.

The entry (or "index") page should look like this:

California Environmental Protection Agency

AIR RESOURCES BOARD
The AQMIS Project

Welcome to The AQMIS Project. We are working hard to improve the state of our system, adding state wide data and tuning our interface.



NEW! Please submit comments or suggestions though our new Comments Page

INTRODUCTION

The goal of the project is to provide California Air Resource Board with a HTML front end to their data.

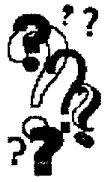
The ARB has approximately 350-400 gathering sites throughout California that gather all sorts of information. Everything from ozone quantities to wind direction. What this project will do is allow maps of California (or sections of California) to be built automatically with the information the user requests to see.

START AQMIS

The outline sections entitled "Initial Map Selection" and "Map Refinement" are "hypertext links" to pages that detail the mechanics of obtaining graphic and map output from AQMIS. The "Initial Map Selection" link will take you to the following two pages. The screen must be "scrolled" via the scroll bar shown at the right of the monitor screen to read the full page.

California Environmental Protection Agency

AIR RESOURCES BOARD
The AQMIS Project



Help & Tutorial

The AQMIS system consists of two basic parts for the actual display of data. Each of these has several sub-parts, providing the varied options. Finally there is a third part to assist in interpreting the data. Below is a table of contents for pages describing the function of each of these parts.

- Initial Map Selection
  - Map details
  - Zoom Factor
  - Map Size
  - Center On Coordinates
  - Center On Site
- Map Refinment
  - Map details
  - Data Representation Method
  - Data Interval
- Site Table

The outline sections entitled "Initial Map Selection" and "Map Refinement" are "hypertext links" to pages that detail the mechanics of obtaining graphic and map output from AQMIS. The "Initial Map Selection" link will take you to the following two pages. The screen must be "scrolled" via the scroll bar shown at the right of the monitor screen to read the full page.

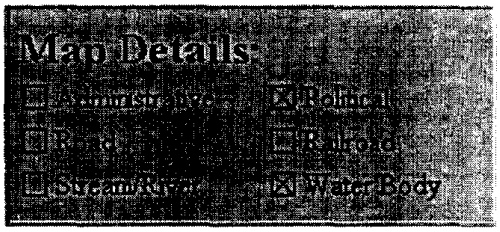
California Environmental Protection Agency

AIR RESOURCES BOARD
The AQMIS Project

Initial Map Selection

This is the first page presented by the AQMIS Map builder. In this process you will select which portion of California you want data displayed for and the amount of detail on the map.

Map Details

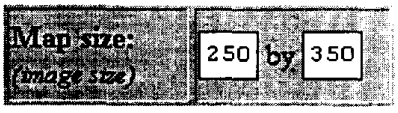


The Map Details section shown in the adjacent figure allows for the selection of any combination of six types of landmarks. The default selection is shown in the picture and is the recomended level of detail for a map of the entire state of California. With a zoom setting of about 3.0 the use of Roads and Streams makes interpreting the map easier.

Zoom Factor

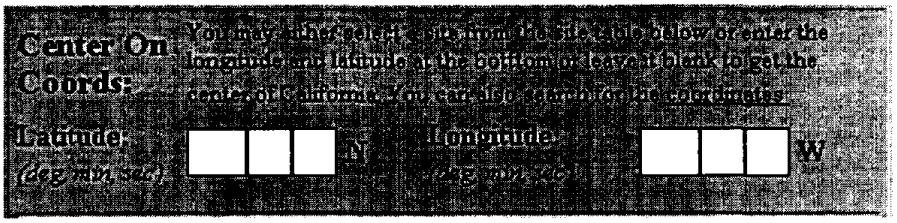
This option determines the extents of the viewable area in the map image. The default value of one will display the entire map of California. A setting of 3 to 4 is a good value for viewing a region.

Map Size



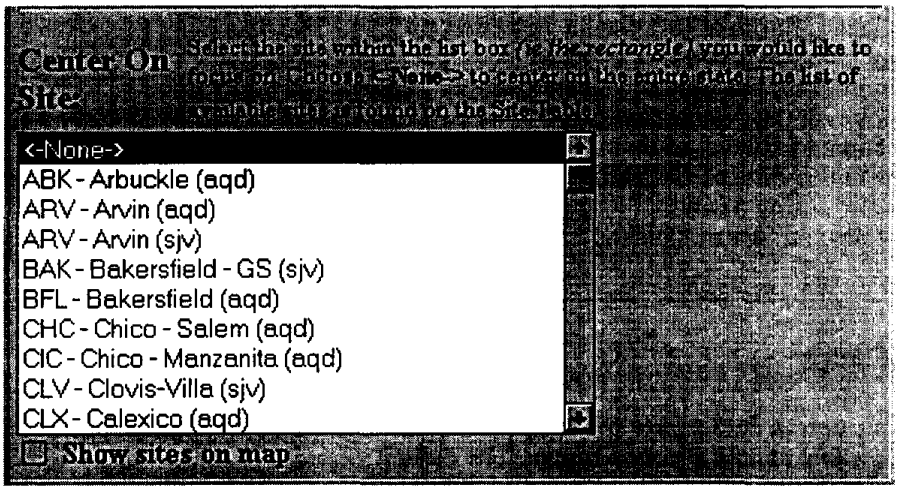
This option allows for sizing of the map. The default size, 250 by 350 pixles, is optimized for a screen resolution of 800x600 pixels. For larger screen resolutions or image downloading the map size can be increased.

Center On Coordinates



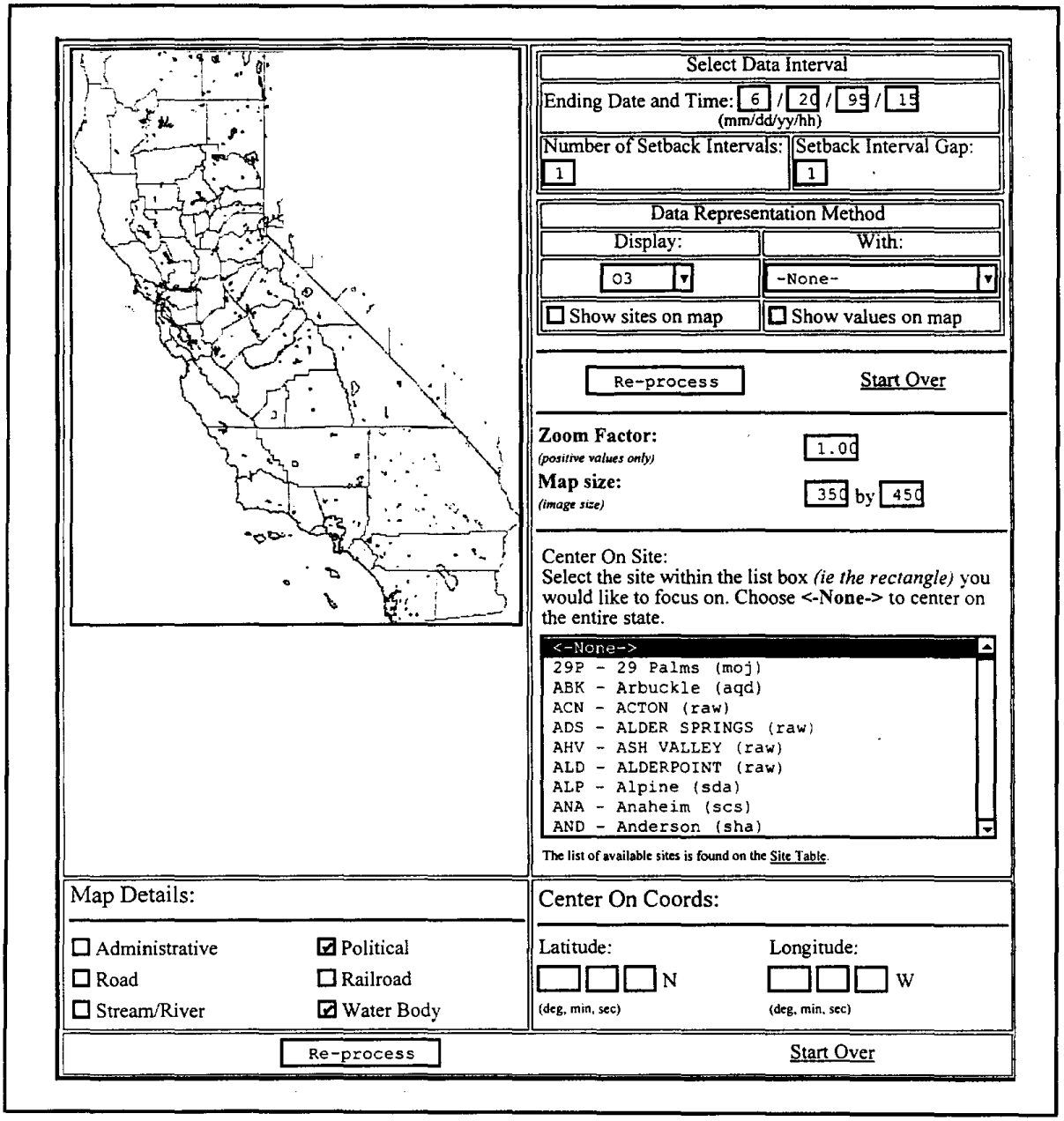
This section chooses where the map center will be. With this option you can enter a latitude and longitude that will become the center of the map image. If left blank, as in the default setting, the map image will be centered on the entire state. The provided link connects to the U.S. Gazetteer, a searchable Geographic Name Server, that will give the latitude and longitudie for any U.S. city.

Center On Site



This section provides a list of sites in the database from which data is collected. Selecting one of the items, other than `<-None->`, will center the map image on the site selected. This option is exclusive with the previous option. If an option other than `<-None->` is selected the map will be centered on that site's latitude and longitude, regardless of the previous sections selection. The provided link connects to a table generated from the current database, listing all sites in it, thier latitude and longitiude, and elevation.

Let us presume that you wish to browse current air quality and/or meteorological data collected by AQMIS. From the tutorial page, you would back track to the entry (or "index") page using either the toolbar or menu "Back" command. Once there, you would select the "Start AQMIS" button by pointing and clicking on it. The page you would immediately see is shown below (the screen must be scrolled to see the page in its entirety):



Once you have selected a map size (please experiment with this, as the default is actually quite small if you have a high resolution monitor), center point, zoom (default is all of California), and select the overlays you wish to be shown, you will need to press the "Re-process" button.

You will probably encounter a security "dialog" window at this point. The version of this dialog box used in Netscape is as follows: ***Any information you submit is insecure and could be observed by a third party while in transit. If you are submitting passwords, credit card numbers, or other information you would like to keep private, it would be safer for you to cancel the submission.*** The purpose of this dialog window is to warn you that the configuration information you are about to send to the AQMIS is not entirely secure. This is a precaution that is built into all browsers to protect users in financial transactions, and is not particularly useful here, so you will probably prefer to click on the small box at the bottom of this dialog to turn the feature off. You can also turn this feature off in the "preferences" section of your menu bar (under "tools" in Netscape).

Once you have passed beyond the initial map selection page, you are ready to make repeated queries using the map refinement page. Each query will require you to reset various options shown on the screen, and click on the "reprocess" button. At any point you can review past screens by using the "Back" menu or toolbar option, or return to the most recent queries by using the "Forward" menu or toolbar option.

A copy of the map refinement screen is shown on the following page.

At this point you may choose to vary both the data you select to display and the manner in which they are displayed. One can choose: (1) to amend the map details that have already been set in the initial map selection page, and/or (2) to select a different meteorological or air quality measure and the manner in which it is displayed, and/or (3) the data intervals from which these data are compiled and displayed.

To select a new center point for the map, you can simply point and click on a new map location, and the map will be redrawn with the "clicked" point as its center. The other settings are amended in the same manner that they were set initially.

The meteorological and air quality measures range from wind and temperature to ozone and carbon monoxide. Although the wind is necessarily displayed in only certain ways, most measures permit a number of choices in the manner in which they can be displayed, including graphs over time and tables, as well as some variations of map displays (some of them relatively experimental). If you wish to transfer tables of data to your own machine for more detailed analysis via the familiar Internet "FTP" method, this is provided as one of the display alternatives.

The data intervals from which the data are drawn are chosen by selecting an ending date and hour, then determining the size (in hours) and number of set-back intervals you wish to have included in the analysis. If, for example, you wish to see the 3:00 p.m. temperatures for the past week, you would select the most recent "15:00" measure (military time), set the size of the setback interval to 24 hours, and select "7" as the number of intervals desired.

After resetting these parameters, point and click the "Re-process" button, wait for the result, and repeat for as many analyses as you wish.

Focused at MOD - Modesto (aqd) (37 38 21 N, 120 59 45 W)