# Architecture and Functionality of the PHAiRS (Public Health Applications in Remote Sensing) Web Services

William B. Hudspeth, Karl K. Benedict, and Amelia M. Budge

Earth Data Analysis Center, University of New Mexico, Albuquerque, NM 87131, USA – (bhudspeth, kbene, abudge)@edac.unm.edu

**Abstract - In the public health community, an increasing number of existing disease and syndromic surveillance decision support systems (DSS) utilize geospatial technologies to providing focused mapping and analysis capabilities on the desktop. Under a cooperative agreement funded by NASA, the Earth Data Analysis Center at the University of New Mexico is developing a suite of capabilities for the Public Health Applications in Remote Sensing (PHAiRS) project. These capabilities will allow public health officials to access, visualize, and analyze data output from a regional dust forecasting model. They are designed to be integrated with existing DSSs as published web services. This paper discusses the current status of work on the PHAiRS web services, and provides an overview of the architecture, programs and technology used.**

**Keywords:** web services, public health, DREAM dust model, decision support systems

## 1. INTRODUCTION

Environmentally induced health risks to populations with respiratory illnesses are a growing concern globally. Of particular concern are dust and smoke events carrying $PM_{2.5}$ (respirable) and $PM_{10}$ (inhalable) particle sizes. Public health officials and authorities worldwide are making decisions concerning health risks to humans caused by these events based on the best available information. In some cases they rely upon reports received at clinics, hospitals, and other care facilities. Others may be accessing databases containing information on syndromes and outbreaks in local and regional areas. Some assess environmental conditions at a global scale such as dust storms originating in the Sahara (Westphal et al., 1987 and 1988; Goudie et al., 2001) or in China. Decision support systems (DSS) that provide early detection and analysis of these events enhance the ability of officials to warn populations at risk.

Many syndromic surveillance systems have been developed in recent years that provide electronic access to information on a variety of diseases and syndromes. A few, such as the Rapid Syndrome Validation Project (RSVP) (Sandia National Laboratories), the Syndrome Reporting Information System (SYRIS) (ARES Corporation), and BioSense (Centers for Disease Control and Prevention) (Bradley 2004), have enhanced their tools with mapping and visualization capabilities. Evolving visualization and mapping technologies can improve these tools.

The Public Health Applications in Remote Sensing (PHAiRS) project, jointly executed by the University of New Mexico's Earth Data Analysis Center and the University of Arizona's Department of Atmospheric Sciences, is developing an application framework to enhance the capabilities of existing public health DSSs such as RSVP and SYRIS. Funded under NASA's REASoN program, the project will provide enhanced web-based visualization and analytical tools to public health decision makers; and, assimilate NASA data into the Dust Regional Atmospheric Model (DREAM) (Nickovic et al., 2001) to provide enhanced dust forecast capabilities.

It is anticipated that, ultimately, these enhancements to public health decision support systems will be accomplished through the development of three related technological frameworks. First, the development of technologies that streamline the ingestion of new environmental data into the system, both as model inputs for DREAM and as observed, *in-situ* particulate concentration data that are used for verification and validation of DREAM forecasts. Second, map data, image services, and analytical functions that may be directly accessed from and integrated into DSS user interfaces. And third, through the development of free-standing, web-based, interactive mapping environment that allows users to explore and analyze environmental and aggregated public health data. This paper describes progress made in the first and second technological frameworks cited above.

## 2. ARCHITECTURE OF PHAiRS DATA DELIVERY SYSTEM

In order to address the technological frameworks described above, recent work on the PHAiRS project has focused on a number of tasks. First, in order to streamline the ingestion of new environmental data, a number of shell scripts have been written that: (1) automate the download of meteorological data from the NCEP and NOAA NOMADS websites; (2) automate the collection and archiving of EPA AirNow particulate data; (3) automate the daily run of the DREAM model and the generation of dust concentration forecasts for the PHAiRS modeling area (Morain et al. 2005; Hudspeth et al. 2005); and (4) post-process of DREAM output data, and archive of resultant raster images in the GRASS GIS system.

Second, the construction of a technological framework for allowing map data, image services, and analytical functions to be directly accessed from and integrated into DSS user interfaces has involved the development of web-based services that allows both project developers and end-users to access a suite of capabilities that extends from identifying and accessing dust concentration data, to statistical evaluation of DREAM output performance, to visualization of both modeled and observed dust concentration in the study area. A services-oriented architecture (SOA) is the foundation for information delivery capabilities that are published for integrating into public health DSSs. As seen in Figure 1, this architecture consists of multiple components, each of which plays one of three roles: (1) data acquisition and storage; (2) data processing and product generation; and (3) product delivery. Each of these steps is discussed in more detail below.

## 2.1 Data ingest and storage

Data storage and provision components of the SOA represent elements of the architecture that store geospatial and related attribute data, and external data resources. These components include the following (right side of Figure 1): (1) GRASS GIS storage; (2) PostGIS storage (Ryden, 2005); (3) external Open Geospatial Consortium (OGC) service providers; and (4) external providers of non-OGC services.

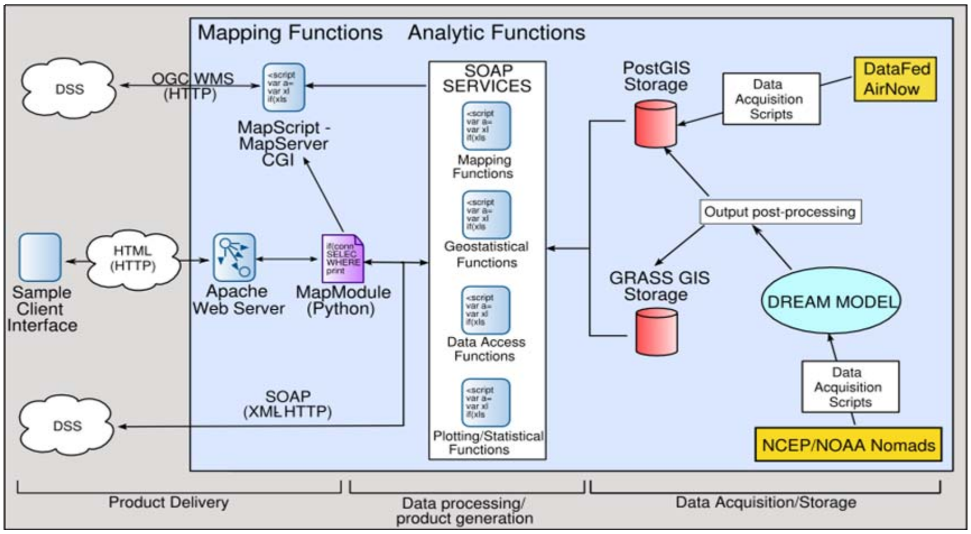


Figure 1. Schematic of PHAiRS Architecture

GRASS GIS and PostGIS contribute data storage, management, and analysis capabilities, while the external data providers contribute already-processed data, either as data services, or as mapping services that provide images of data. GRASS GIS provides an environment for storing, analyzing, and delivering raster data through visualization and analysis services built upon the rich function library in GRASS. PostGIS provides complementary storage, management, and delivery of vector data.

At present, two classes of data download are automated, and are obtained from external data providers that host both OGC and non-OGC services through a variety of protocols. The first class of automated downloads for the PHAiRS project uses the Environmental Protection Agency (EPA) AIRNow OGC Web Coverage Services (WCS) (Evans, 2005), hosted by the DATAFED project. PM2.5 and PM10 particulate data for all recording stations within the PHAiRS project area are harvested daily. As each day's data are downloaded, they are imported into a PostgreSQL/PostGIS database that stores hourly records that include the station ID, latitude and longitude coordinates, the UTC time, and the concentration value for the particle class of interest. The second class of data download is a parallel collection and archive of DREAM initialization and boundary condition meteorological data from the HTTP-based download services hosted by the National Oceanic and Atmospheric Administration's (NOAA) National Operational Model Archive & Distribution System (NOMADS), and the National Weather Service's, FTP-based, operational forecast delivery system. Also used are products obtained from the Land Processes Distributed Active Archive Center (DAAC) via the Earth Observing System (EOS) Data Gateway. These data are simply stored on the data file system in individual subdirectories for each day's DREAM model run.

## 2.2 Data Processing and Product Generation

The central functions of the system are provided by components that process data stored in the previously described management system. The raw, archived data discussed above are processed and managed by a single BASH shell script that coordinates daily DREAM model runs, post-processing of DREAM outputs, and storage of resultant information back to the database management system. In turn, it is this output information that can be used to generate products for public health officials that are deployable within existing DSSs.

The processing flow begins with the ingest of the daily NCEP meteorological data, and its processing by the DREAM model to generate a collection of binary-formatted data files that encode dust concentration values for each of the 49 hours modeled by a daily DREAM run. Moreover, each hour is represented by four different files, each containing the ground level concentration values for one of the four particle size bins output by DREAM. These binary files are in turn written to Arc-ASCII text files, and imported into the GRASS GIS database as an array of three-dimensional raster files where each raster represents a unique UTC time and particle size category. In addition to the somewhat arbitrarily-defined particle bin sizes output by DREAM, the import process uses the two smallest particle size classes to build GRASS rasters that approximate PM2.5 and PM10 concentrations for each hour. As a final step, each raster produced in the GRASS archive is recorded in a PostgreSQL/PostGIS table as a unique record.

### 2.3 Product Delivery

A technological framework that allows both developers and end-users to access and assess the products generated in the previously discussed activities is represented by the Simple Object Access Protocol (SOAP) services depicted in the center of Figure 1. Each service provides specialized products based upon data stored within, and accessed by, the data storage and processing components.

Data processing and product generation is supported through automated processes that are initiated through W3C SOAP (Gudgin, et al. 2003) requests to the project server that executes one of four functions: (1) data access functions; (2) geostatistical functions; (3) mapping functions; and (4) statistical and plot generation functions. These services are the primary means through which external DSSs (such as RSVP and SYRIS) gain access to, and embed within their systems, the products of the PHAiRS project.

SOAP services provide a more sophisticated service interface for delivering complex data structures and binary products (i.e. PDF files, lists of WMS and image URLs, and PNG files). This increased capacity facilitates delivering high-end capabilities to public health DSS applications without developers writing the often complex code necessary to generate the returned products. Using SOAP as the standard service protocol also provides maximum flexibility to DSS developers since SOAP is widely supported by numerous programming languages and host platforms as a standard web service model. A more detailed discussion of available SOAP services follows below.

## 3. FUNCTIONALITY OF PHAiRS WEB SERVICES

Web service functionality for the PHAiRS project falls into three main categories. They include: (1) data discovery and download; (2) statistical verification and validation functions; (3) map image generation, and (4) geostatistical functions. While the development of much of the existing functionality of the PHAiRS project web services is currently oriented toward verification and validation of the DREAM model outputs, many of the capabilities are directly relevant to decision-support systems that are dependent upon the web services described here.

### 3.1 Data Discovery and Download

The PHAiRS web service architecture allows users to search for and download both EPA AirNow PM2.5 and PM10 particulate data, as well as DREAM model output values for specific locations. At present, the EPA AirNow data values are not speciated, and represents a composite measure of geologically-derived, as well as anthropogenically-produced particles. Users have the option of downloading solely PM2.5 or PM10 EPA AirNow data for a defined date range, or a single day. Similarly, SOAP service functions allow the user to download both EPA and DREAM dust concentration values for a single station, or for all stations within the modeling domain, as well as to download data for a single day, a 48-hour period corresponding to a DREAM model run, or a date range specified by the user.

### 3.2 Statistical Verification Procedures

In order to verify and validate the performance of consecutive versions of the DREAM dust model, we have designed a number of web services that calculate measures of central tendency as well as measures of variability for both observed and modeled dust concentration values. These include the mean and standard deviation. Another set of statistics provide various measures of association between these two variables. These include the mean bias, mean error, normalized mean bias, normalized mean error, fractional bias, fractional error, index of agreement, the correlation coefficient (R), and the centered root mean square (RMS). These statistics can be obtained for a single station for a 48-hour DREAM run, or for a date range specified by the user.

### 3.3 Map Image Generation

Publishing PHAiRS data as OGC WMS for delivering mapped data to public health DSSs follows a model developed for RSVP. This model proved effective in facilitating the deployment of dynamic map content into external applications. Developers of public health DSS applications do not need to be expert in online geospatial application development. Instead, they can embed maps developed by the PHAiRS system in their interfaces. An important advance in the PHAiRS WMS capabilities was achieved by implementing the WMS Time specification for the EPA AIRNow historical data and modeled DREAM outputs. This achievement allows for efficient management and delivery of time-stamped measurements of particulate densities and corresponding DREAM model outputs.

Thus, requests for single map images, as seen in Figure 2, typically involve a Web Map Service (WMS) request to the Minnesota Mapserver CGI. Requests are formatted as a simple HTTP URL that indicates the map layers of interest, as well as a time stamp argument required to display either or both DREAM output rasters and classified AirNow PM2.5 and PM10 data.

### 3.4 Geostatistical Functions

Geostatistical SOAP services provide the functionality that is demonstrated in the sample client interface developed for the project. For example, one of the geostatistical functions in this interface permits users to select, on screen, a geographic extent represented by an irregular polygon, such as a political boundary, which converts that selection to an appropriate SOAP client call to the geostatistical and graphics generation SOAP service. A density plot of the value of interest (e.g. $PM_{2.5}$ concentration) is returned to the user. This functionality uses the R statistical programming language and its ability to access raster data stored in GRASS to perform geostatistical analyses. While only an initial step, this illustrates a more general capability to provide statistical summary functions of raster data over irregular areas to public health professionals through the SOA.

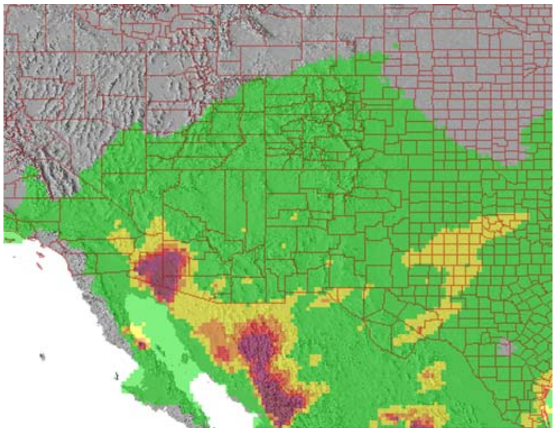

Figure 2. Map Image Service

## 4. CONCLUSIONS

Accomplishments discussed in this paper toward developing the web-based mapping services to enhance existing public health DSSs represent the first three years of a five-year project. This progress reflects both a rapid development cycle and closely linked integration of new technologies into the products of the PHAiRS project. Visualization and mapping capabilities of the sample client interface have evolved to provide a Python-based, modular, multifunctional client that provides access to the multiple data and image services based upon two core standards: the W3C SOAP specification, and the OGC WMS specification. Creating services to support these protocols also provides a streamlined means for public health DSSs to integrate the enhanced DREAM model outputs and related environmental data products into their application frameworks.

## ACKNOWLEDGEMENTS

The authors thank all of the members of the PHAiRS project team, including S. Nickovic, D. Yin, B. Barbaris, T. Budge, S. Morain, P. Shaw, and W. Sprigg.

This work is sponsored under NASA CA# NNS04AA19A.

## REFERENCES

C.A. Bradley, H. Rolka, D. Walker, and J. Loonsk, "BioSense: implementation of a national early event detection and situational awareness system", in *Morbidity and Mortality Weekly Report: Syndromic Surveillance, Reports from a National Conference*, Dept. of Health and Human Services, Centers for Disease Control and Prevention, vol. 54, p.p.11-19, 2004.

J.D. Evans (ed.), 2005. Web Coverage Service (WCS), version 1.0.0 (corrigendum). Document OGC 05-076. Open Geospatial Consortium, Inc., USA.

A.S. Goudie, and N.J. Middleton, "Saharan dust storms: nature and consequences", Earth Sci. Rev., vol. 56, p.p. 179-204, 2001.

M. Gudgin, M. Hadley, N. Mendelsohn, J. Moreau, and H.F. Nielsen, 2003. "SOAP version 1.2 part 1: messaging framework.", World Wide Web Consortium. http://www.w3.org/TR/soap12-part1.

W.B. Hudspeth, S. Nickovic, D. Yin, B. Chandy, B. Barbaris, A. Budge, T. Budge, S. Baros, K. Benedict, C. Bales, C. Catrall, S. Morain, G. Sanchez, W. Sprigg, and K. Thome, "PHAiRS – A Public Health Decision Support System: Initial Results," Proceedings of the 31st International Symposium on Remote Sensing of Environment, Saint Petersburg, Russian Federation, June, 2005.

S. Morain, A. Budge, T. Budge, S. Baros, K. Benedict, W. Hudspeth, C. Bales, G. Sanchez, W. Sprigg, D. Yin, B. Barbaris, B. Chandy, K. Thome, C. Cattrall, S. Nickovic, S. Caskey, J. Speer, and J. Bradbury, "Modeling Atmospheric Dust for a Public Health Decision Support System," Proceedings of the 31st International Symposium on Remote Sensing of Environment, Saint Petersburg, Russian Federation, June, 2005.

S. Nickovic, G. Kallos, A. Papadopoulos, and O. Kakaliagou, "A model for prediction of desert dust cycle in the atmosphere, " J. Geophys. Res., vol. 106, p.p. 18113-18130, 2001.

K. Ryden (ed.), 2005. OpenGIS implementation specification for geographic information – simple feature access – part 1: common architecture. Document OGC 05-126, version 1.1.0, Open Geospatial Consortium, Inc., USA.

K. Ryden (ed.), 2005. OpenGIS implementation specification for geographic information – simple feature access – part 2: SQL option. Document OGC 05-134, version 1.1.0, Open Geospatial Consortium, Inc., USA.

D.L. Westphal, O.B. Toon, and N. Carlson, "A two-dimensional investigation of the dynamics and microphysics of Saharan dust storms," J. Geophys. Res., vol. 92, p.p. 3027-3049, 1987.

D.L. Westphal, O.B. Toon, and N. Carlson, "A case study of mobilization and transport of Saharan dust," J. Atmos. Sci., vol. 45, p.p. 2145-2175, 1988.