

Husar, R.; Poirot, R. DataFed and Fastnet: Tools for Agile Air Quality Analysis; *Environmental Manager* 2005, September, 39-41

DATAFED AND FASTNET: TOOLS FOR AGILE AIR QUALITY ANALYSIS

Rudolf B. Husar and Rich L. Poirot

The instantaneous 'horizontal' diffusion of information via the Internet now permits the delivery of the right information to the right people at the right place and time. .

INTRODUCTION

Traditionally, air quality analysis was a slow, deliberate investigative process occurring months or years after the monitoring data had been collected. Satellites, real-time pollution detection and the World Wide Web have changed all that. Analysts can now observe air pollution events as they unfold. They can 'congregate' through the Internet in *ad hoc* virtual work-groups to share their observations and collectively create the insights needed to elucidate the observed phenomena. Air quality analysis has become much more agile and responsive to the needs of air quality managers, the public and the scientific community.

The era of near real-time air quality analysis began in the late 1990s, with the availability of real-time satellite data over the Internet. High-resolution color satellite images were uniquely suited for early detection, and tracking of extreme natural or anthropogenic aerosol events. In April 1998, for example, a group of analysts keenly followed and documented on the Web, in real-time the trans-continental transport and impact of Asian dust from the Gobi desert on the air quality over the Western US. (Husar, et al., 2001, <http://capita.wustl.edu/Asia-FarEast>). Soon after, in May 1998, another well-documented incursion of Central American forest fire smoke caused record PM_{2.5} concentrations over much of the Eastern US (e.g. Peppler, et al., 2000; <http://capita.wustl.edu/Central-America>).

The high value of qualitative real-time air quality information to the public is well demonstrated through EPA's successful AIRNOW program (Weyland and Dye, 2006).

However, during extreme air quality events like the above-mentioned, air quality managers need more extensive 'just in time analysis', not just qualitative air quality information. In fact, in the 1998 Asian Dust Event, local air quality managers in Oregon and Washington used the real-time analysis to issue health advisories. Soon after the Central American Smoke Event, the federal EPA granted some states 'exceptional event' exemptions from ozone standard violations. These responsive air quality management actions were largely facilitated by the agile event-analyses provided by the ad hoc community of scientist and managers collaborating through the Internet.

In recent years, air quality management has also changed. The old command and control style is giving way to a more participatory approach that includes all the key stakeholders from multiple jurisdictions and application of scientific and 'weight of evidence' approaches. The air quality regulations now emphasize short-term monitoring while at the same time long-term air quality goals are set to glide toward 'natural background' levels over the next decades. In response to these and other development, EPA has undertaken a major redesign of the monitoring system that provides the data input for air quality management. The new National Ambient Air Monitoring Strategy (NAAMS), through its multi-tier integrated monitoring system, is geared to provide more relevant and timely data for these complex management needs. All these changes in management style place considerable burden on the information system that supports them.

Fortunately, both the air quality monitoring and data dissemination technologies have also advanced considerably since the 1990s. Recent developments offer outstanding opportunities to fulfill the information needs for the new agile air quality management approach. The data from surface-based air pollution monitoring networks now provides routinely high-grade, spatio-temporal and chemical patterns throughout the US for PM₂₅ and ozone. Satellite sensors with global coverage and kilometer-scale spatial resolution now provide real-time snapshots which depict the pattern of haze, smoke and dust in stunning detail. The 'terabytes' of data from these surface and remote sensors can now be stored, processed and delivered in near-real time. The instantaneous 'horizontal' diffusion of information via the Internet now permits, in

principle, the delivery of the right information to the right people at the right place and time. Standardized computer-computer communication languages and Service-Oriented Architectures (SOA) now facilitate the flexible processing of raw data into high-grade ‘actionable’ knowledge. Last but not least, the World Wide Web has opened the way to generous sharing of data and tools leading to faster knowledge creation through collaborative analysis and virtual workgroups. Nevertheless, air quality analysts face significant hurdles.

The new developments introduced a new set of problems. The “data deluge” problem is especially acute for analysts interest in aerosol pollution, since aerosols are so inherently complex and since there are so many different kinds of relevant data – from extensive, new surface-based monitoring networks, meteorological and aerosol forecast models, satellite imagery and associated data products, etc.

In this brief report we present *DataFed*, an infrastructure for real-time integration and web-based delivery of distributed monitoring data. Next we highlight *FASTNET*, a recent application project built on the DataFed infrastructure, which utilizes real-time and historical monitoring data for the collaborative study of major aerosol events.

DATAFED: DATA FEDERATION

The federated data system, DataFed, (<http://datafed.net>) aims to support air quality management and science by more effective use of relevant data. Building on the emerging pattern of the Internet itself, DataFed assumes that datasets and new data processing services will continue to emerge spontaneously and autonomously on the Internet, as shown schematically in Figure 1. Example data providers include the AIRNOW project, modeling centers and the NASA Distributed Active Archive Centers (DAAC).

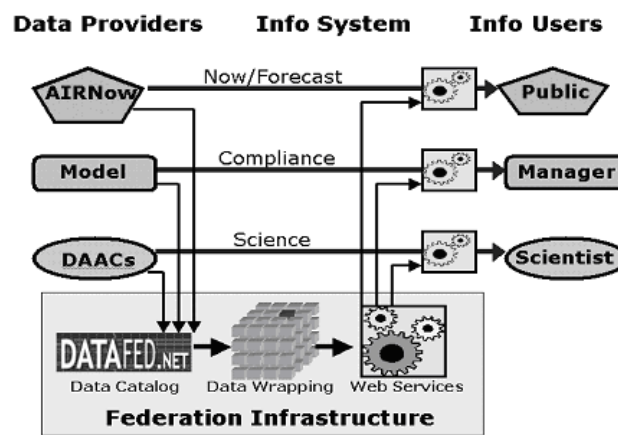


Fig 1 Schematic air quality information system: providers, users and data federation infrastructure.

DataFed is not a centrally planned and maintained data system but a facility to harness the emerging resources by powerful dynamic data integration technologies and through a collaborative federation philosophy.

The key roles of the federation infrastructure are to (1) facilitate registration of the distributed data in a user-accessible catalog; (2) ensure data interoperability based on physical dimensions of space and time; (3) provide a set of basic tools for data exploration and analysis. The federated datasets can be queried, by simply specifying a latitude-longitude window for spatial views, time range for time views, etc. This universal access is accomplished by ‘wrapping’ the heterogeneous data, a process that turns data access into a standardized web service, callable through well-defined Internet protocols.

The result of this ‘wrapping’ process is an array of homogeneous, virtual datasets that can be queried by spatial and temporal attributes and processed into higher-grade data products. The Service Oriented Architecture (SOA) of DataFed is used to build web-applications by connecting the web service components (e.g. services for data access, transformation, fusion, rendering, etc.) in Lego-like assembly. The generic web-tools created in this fashion include catalogs for data discovery, browsers for spatial-temporal exploration, multi-view consoles, animators, multi-layer overlays, etc.(Figure 2).



Fig 2. Main software programs of DataFed-FASTNET: Catalog for finding and selecting data, Viewer for exploration, and an array of Consoles for data analysis and presentation.

A good illustration of the federated approach is the realtime AIRNOW dataset described in a companion paper in this issue (Wayland and Dye, 2005). The AIRNOW data are collected from the States, aggregated by the federal EPA and used for informing the public (Figure 1) through the AIRNOW website. In addition, the hourly real-time O₃ and PM_{2.5} data are also made accessible to DataFed where they are translated on the fly into uniform format. Through the DataFed web interface, any user can access and display the AIRNOW data as time series and spatial maps, perform spatial-temporal filtering and aggregation, generate spatial and temporal overlays with other data layer and incorporate these user-generated data views into their own web pages. As of early 2005, over 100 distributed air quality-relevant datasets have been ‘wrapped’ into the federated virtual database. About a dozen satellite and surface datasets are delivered within a day of the observations and two model outputs provide PM forecasts.

THE FASTNET PROJECT

FASTNET (Fast Aerosol Sensing and Tools for Natural Event Tracking) (Poirot, et al, 2005) is a data acquisition and analysis facility for improving the efficiency as air quality analysts, with particular emphasis on detailed real-time and post-analysis of major aerosol events. Natural aerosol events from forest fire smoke and windblown dust are particularly interesting events, due to their large emission rates over short periods of time, continental and global-scale impacts, and unpredictable sporadic occurrence.

For the FASTNET project, 14 specific data sets are highlighted which include various surface-based aerosol data (EPA fine mass, speciated aerosol composition from EPA and IMPROVE network), hourly surface meteorology and visibility data, aerosol forecast model results, and various satellite data and images. Many of these data are available in near-real-time, while others (for example the IMPROVE filter-based aerosol chemistry data and associated back trajectories) are available with a time lag of about 1 year. Analysts access the desired data through the DataFed Data Catalog (Figure 2.) The selected data are automatically loaded into a web-based data browser designed for easy exploration of the spatiotemporal pattern. Semantic data homogenization assures that all datasets can be properly overlaid in space and time views (Figure 2).

For illustration of the FASTNET analysis concept, tools and application methods, we use the April 2003 Kansas smoke event. For several days every spring, the rangeland grass in Kansas-Oklahoma is burned resulting in major smoke plumes that cover multi-state areas of the Midwest. Figure 3 shows the location of fires over Kansas derived from the MODIS satellite sensor, the noon aerosol optical thickness derived from the SeaWiFS satellite sensor and the spatial pattern of organic fine particle mass concentration derived from speciated aerosol data collected through several surface monitoring networks.

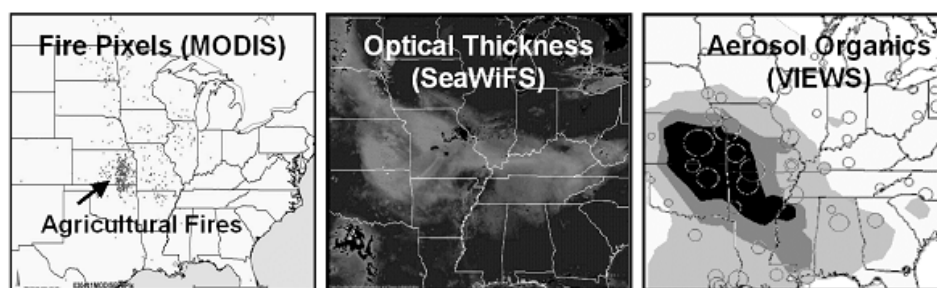


Fig 3 Multi-sensory characterization of agricultural smoke over the Midwest on April 12, 2003: a. Fire pixels over Kansas from the MODIS sensor; b. Vertical aerosol optical thickness derived from the SeaWiFS satellite; c. Concentration pattern of organics from speciated chemical samplers (EPA)-

The spatio-temporal sparseness of routine surface monitoring networks prevents the full characterization of aerosol events but complementary satellite observations can fill in many of the missing pieces. Auxiliary data from the 1200 station NOAA surface visibility network (not

shown) is particularly useful for the study of fine-scale aerosol pattern at the surface. During the event and the days-years after, the data from these real-time monitoring systems were analyzed by several groups using the FASTNET tools and methods. Based on such exploratory analyses, the federal EPA is considering exceedance waivers for remote sites that were impacted by the agricultural smoke. The local managers are also using the dynamic analysis tools to evaluate alternative grass burning schedules and approaches. Collectively, these types of analyses are used in the new, more flexible ‘weight of evidence’ approach to air quality compliance analysis.

SUMMARY AND DISCUSSION

Recent developments in surface and satellite sensing along with new information technologies now allow real-time, ‘just-in-time’ data analysis for the characterization and partial explanation of the of major air pollution events as well as more in-depth post-analysis, integration and fusion can also be performed using the federated historical data resources and tools. By making available many spatio-temporal data sources through a single web interface and in a consistent format, the DataFed tools allow anyone to view, process, overlay, and display many types of data to gain insight to atmospheric physical and chemical processes. A goal of the current effort is to encourage use of these tools by a broad community of air pollution researchers and analysts, so that a growing group of empowered analysts may soon enhance the rate at which our collective knowledge of air pollution evolves. In recent years, such agile analyses have provided occasional real-time support to the air quality managers and to the public but much more could be done. The current challenge is to incorporate such support into the air quality management process in a more regular and robust way.

ACKNOWLEDGEMENTS

DataFed is a community-supported effort. While the data integration web services infrastructure was initially supported by specific information technology grants from NSF and NASA, the data resources are contributed by the autonomous providers. This material is based upon work supported partially by the National Science Foundation under Grant NO. 0113868. Any opinions, findings, and conclusions or recommendations expressed in this material are those

of the authors and do not necessarily reflect the views of the National Science Foundation. The application of the federated data and tools is in the hands of users as part of specific projects. Just like the way the quality data improves by passing it through many hands, the analysis tools will also be improve with use and feedback from data analysts. A partial list is at <http://datafed.net/projects>. At this time the DataFed-FASTNET user community is small but substantial efforts are under way to encourage and facilitate broader participation through larger organizations such as the Earth Science Information Partners (ESIP) Federation (NASA, NOAA, EPA main member agencies) and the Regional Planning Organizations (RPOs) for regional haze management.

REFERENCES

Husar, R.B., et al. The Asian Dust Events of April 1998; *J. Geophys. Res. Atmos.* **2001**, 106, 18317-18330. See also The Asian Dust Events of April 1998 at <http://capita.wustl.edu/Asia-FarEast/>.

Peppler, R.A., et al. RM Southern Great Plains Site Observations for the Smoke Pall Associated with the 1998 Central American Fires; *Bull. Am. Meteor. Soc.*, **2000**, 81, 2563-2591. See also A Resource Catalog and Discussion Forum on Smoke from Central American Fires at <http://capita.wustl.edu/Central-America/>.

Wayland, R.A.; Dye, T.S. AIRNOW: America's Resource for Real-Time and Forecasted Air Quality Information; *Environmental Manager* 2005, *September*, 19-27.

Poirot, R. L. et al. Aerosol and Haze observations during 2004 with the FASTNET Distributed Monitoring System. Presented at the A&WMA Specialty Convergence on Regional and Global Perspectives on Haze: Cause, Consequences and Controversies, Asheville, NC, 2004; Paper # 93.

About the Authors:

Rudolf B. Husar is a Professor of Mechanical Engineering and director of the Center for Air Pollution Impact and Trend Analysis at Washington University. Richard L. Poirot is an Air Quality Analyst at the State of Vermont Department of Natural Resources.

Contacts: rhusar@me.wustl.edu and richard.poirot@state.vt.us