CYBERINFRASTRUCTURE FOR AIR QUALITY MANAGEMENT

The following describes a potential future government procurement requirement. EPA would be interested in input regarding the following questions:

- *Is the effort, as described here, likely to be successful in supporting a transition to a more integrated, service-oriented architecture for air quality information systems in the U.S. and elsewhere?*
- *If not, why? What are obstacles or challenges are not addressed adequately?*
- *What are the key areas of expertise or tasks that should be emphasized?*
- *What might be accomplished in the first year? at what cost?*
- *What might be accomplished over multiple years? at what cost?*

SUMMARY

EPA is considering the need to issue a solicitation inviting proposals to assist EPA in planning, developing, and maintaining a cyberinfrastructure for air quality management, linking and drawing upon existing data systems where possible. The cyberinfrastructure is envisioned as a service-oriented, open-source, web-based network of air quality and pollutant emissions data repositories or providers and existing and new data analysis tools for use by the air quality management and research communities. The successful proposal will bring together a team with expertise in several areas: air quality modeling, data analysis, and forecasting; emissions inventory development and assessment; wildfire smoke analysis; use of satellite and ground-based remote sensing information; scientific visualization; communication of air quality data to decision-makers and the public; and service-oriented architecture and software development. The team will assess current tools and repositories for integration, synthesis, and publication of air quality and related data, identify appropriate standards and best practices for an air quality data cyberinfrastructure, and otherwise support EPA in the production of a plan for the transformation of the current set of data repositories and software tools into a service-oriented, open-source air quality data system of systems. The team will then work to implement this plan, working in concert with the broader air quality management and science community to link tools and repositories, fill in gaps between existing tools and repositories, make tools and data available to the wider community, and educate and train the community about them and their use. The effort will include implementation and maintenance of the cyberinfrastructure plan, seeking and facilitating community input on the cyberinfrastructure plan and implementation, software development, and outreach to educate and train the community. The resulting cyberinfrastructure for air quality management is expected to be consistent with and contribute to the U.S. Integrated Earth Observing System (IEOS)[1] and the Global Earth Observations System of Systems (GEOSS).[2]

[1] http://www.usgeo.gov/
[2] http://www.earthobservations.org/

BACKGROUND

Effective air quality management requires the acquisition and analysis of a variety of types of data, such as atmospheric concentrations of various species from ground- and satellite-based instruments, information about emissions and emissions-generating activities, meteorological parameters, and other environmental characteristics. Moreover, numeric models provide essential information, as many important quantities cannot be measured at the needed times and places or with the needed quality. These data are collected and compiled by various government agencies, as well as academic and private sector scientists and engineers, at the local, state, regional, national, and international scales. To better understand the linkages between individual air quality pollutants or between problems at different spatial scales, it is necessary to share, integrate, and analyze data from ambient monitors, satellites, and numeric models. However, managing data from diverse sources is often difficult and resource-intensive. Fusing data requires technical skill and computational resources, but also needs scientific judgment and insight.

The World Wide Web and emerging information technologies make it possible to improve the efficiency of efforts to share, integrate, and analyze data from diverse sources. In addition to more effectively using existing data to generate insights, such technology will be useful in generating new information to fill gaps. The need for this integration of air quality-related information has been articulated in the Strategic Plan for the U.S. Integrated Earth Observation System (IEOS), the U.S. Group on Earth Observations' (USGEO) Near Term Opportunity Plan for Air Quality Assessment and Forecasting,[3] and the Strategic Plan of the international Group on Earth Observations (GEO), which calls for the development of a integrated Global Earth Observations System of Systems (GEOSS). Several recent documents have helped to begin to describe the architecture and needed attributes of such a system of systems, including the GEO Work Plan for 2009 – 2011,[4] the GEO Architecture Implementation Pilot−Phase 2,[5] and a recent report prepared for EPA's Office of Air and Radiation.[6]

To meet these needs, EPA is interested in helping to develop a cyberinfrastructure for air quality management. Cyberinfrastructure, as defined by the National Science Foundation's Blue Ribbon Advisory Panel, refers to the hardware, software, services, personnel, and organizations necessary to build new types of distributed and collaborative knowledge environments to support science and engineering.[7] EPA is currently supporting, directly or indirectly, a variety of individual projects that would contribute to a cyberinfrastructure for air quality management; other agencies at the state, federal, and international levels are also supporting elements that would contribute. However, these tools are typically independent, end-to-end applications built for specific operational or

[3] http://www.usgeo.gov/docs/nto/Air_Quality_NTO_2006-0925.pdf
[4] http://www.earthobservations.org/documents/work%20plan/geo_wp0911_rev1_090113.pdf
[5] http://portal.opengeospatial.org/files/?artifact_id=28934
[6] See section 1.8: http://wiki.esipfed.org/index.php/Air_Quality_Data_Systems_Assessment
[7] National Science Foundation (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure. Arlington, Virginia. http://www.cise.nsf.gov/evnt/reports/toc.htm

research purposes. Furthermore, many of the tools developed at EPA and some EPA datasets are not generally available to the wider community outside EPA. Moreover, there are significant gaps among these components of the cyberinfrastructure which prevent information from being fully used for effective air quality management. The envisioned cyberinfrastructure will be a more versatile, service-oriented, integrated, and open system of systems for air quality data discovery, access, manipulation, fusion, and visualization. This will allow analysts, decision-makers, and researchers inside and outside the EPA to achieve more value from existing and future air quality data sets. This solicitation is expected to lead to the formation of a team of experts that can facilitate linkages between the separate ongoing efforts across the air quality community and both foster and directly carry out new activities that will develop and extend the cyberinfrastructure as part of GEOSS for the benefit of all.

Described here are three broad examples of some of the intended uses of the air quality cyberinfrastructure. These challenging capabilities are included here as illustrations. The ultimate goal is a cyberinfrastructure with the flexibility to meet these and other scientific and decision-support needs. As described below, the successful proposal will describe a plan for development of these capabilities, building upon currently existing elements.

- **Air Quality Forecasting and Public Information**
  The cyberinfrastructure will be used to inform air quality forecasting efforts by improving access to and analysis of near real-time air quality and meteorological observations from surface sites and satellites, along with air quality and meteorological model forecasts. Currently, satellite air quality data is only available after significant time lags, and tools for handling both satellite and model data are needed. The cyberinfrastructure will assist the fusion of ambient air quality measurements (which are the basis of current public information systems) with the developing satellite and model data. Important existing elements of the cyberinfrastructure for these purposes are AIRNow and IDEA (see Table 1, below).

- **Air Quality Data Analysis, Modeling, and Assessment**
  The cyberinfrastructure will be used to facilitate the analysis, modeling, and assessment of past air quality trends and episodes for purposes of air quality research and management. One area of analysis will involve air quality model evaluation through model to observations comparisons. Tools are needed to access, visualize, and compare model outputs and observations from surface sites, aircraft, and satellites, in various standard formats, handling different spatial and temporal scales and performing standard tests and diagnostics. Another area of analysis will involve the "re-analysis" or "fusion" of multiple types of observations and/or model estimates (available *via* the cyberinfrastructure) to create an integrated data set that best describes the state of the atmosphere at a given point and time. Re-analysis datasets are essential for many research and policy efforts, such as exposure and public health assessments, control measure evaluation or accountability assessments, trends analysis, exceptional event determinations, quantification of long-range transport, and source apportionment analysis. Similar to model evaluation, the production of these datasets will require fusing observations and model output with multiple temporal and spatial domains, using various standard formats. Some examples of existing elements

of a cyberinfrastructure related to these functions include Datafed, VIEWS, GIOVANNI, RSIG, and AMET (see Table 1).

- **Emissions Inventory Development and Evaluation**
  The cyberinfrastructure will be used to access, expand, compare, and evaluate inventories of emissions. Emissions data presents specific challenges for data fusion, visualization, and evaluation. Current emissions inventories have large shortcomings; fusion with models and observations (of air quality as well as emissions related activity, such as fire activity or land cover) are needed to constrain these inventories. Some existing elements of a cyberinfrastructure in this area are EIS, NEISGEI, GEIA, and EDGAR (see Table 1).

The successful proposal will describe how the contractor's team will build upon current efforts at EPA and elsewhere to enable the three functions described above. This potential solicitation follows the end of a previous cooperative agreement between EPA and a team led by Washington University, St. Louis, which explored the development of an air quality cyberinfrastructure. This initial effort supported the development of a portal for the Networked Environmental Information System for Global Emissions Inventories (NEISGEI), an initiative to develop a distributed emissions database using the World Wide Web to provide access and facilitate analysis of emissions data collected at the local, regional, and global scales, [8] as well as contributing to the development of DataFed, a web-based system that supports data sharing and processing for collaborative air quality management and atmospheric science research.[9] The potential solicitation seeks to extend the concepts behind DataFed and NEISGEI to connect other existing systems and tools as nodes in a broader network of air quality information sources and processing tools.

This solicitation also seeks to build upon the EPA's Advanced Monitoring Initiative (AMI),[10] which funded 17 air quality-related pilot projects in FY06 and FY07 to assist in the implementation of IEOS and GEOSS. A number of AMI projects have focused on data fusion for various research and decision-making needs; the cyberinfrastructure will integrate and build upon these efforts. The funding for the first year of this agreement is expected to come from AMI funds for FY09.

Together with a number of partners from the U.S. and abroad, EPA is participating in an air quality response to Phase 2 of GEO's Architecture Implementation Pilot (AIP-2). AIP-2 is a collaborative effort to define what is needed from GEOSS in the societal benefit areas defined by GEO, and describe the cyberinfrastructure needed to make GEOSS successful. The initial Call For Participation for AIP-2[11] offered a description of the envisioned architecture of GEOSS as of summer 2008. The ongoing AIP-2 effort is refining this architecture plan. Air quality participants are building an Air Quality Community Portal and Catalog, which will help the air quality community interface with

[8] http://www.neisgei.org/
[9] http://www.datafed.net/
[10] http://www.epa.gov/geoss/factsheet/ami06.pdf
[11] http://portal.opengeospatial.org/files/?artifact_id=28934

GEOSS. This will be a significant step in creating the needed cyberinfrastructure, which this solicitation seeks to build upon.

Table 1 lists a number of elements of the existing air quality information system of systems within the United States and beyond that the proposed cyberinfrastructure would be expected to connect and support. Many of these elements were discussed at the Air Quality Data Summit hosted by EPA, February 12-13, 2008,[12] which primarily focused on sources, processing centers, and repositories for ambient air quality data relevant to national, state, and local air quality management in the United States. EPA has continued the evaluation and planning process initiated at this Summit,[13] and commissioned a report to survey of a number of the existing elements of the proposed air quality cyberinfrastructure, offer a description of the preferred future, and make specific recommendations for EPA actions that can readily fill significant gaps in the cyberinfrastructure.[14]

In addition to the efforts under GEO AIP-2, there are a number of tasks identified in the GEO 2009 – 2011 Work Plan that the proposed cyberinfrastructure for air quality management should seek to contribute to or build upon:

HE-09-02a	Aerosol Impacts on Health and Environment: Research, Monitoring and Prediction
HE-09-02b	Air Quality Observations, Forecasting, and Public Information
HE-09-02e	Global Monitoring Plan for Atmospheric Mercury
DA-09-02e	Atmospheric Model Evaluation Network

[12] http://wiki.esipfed.org/index.php/Air_Quality_Data_Summit
[13] http://wiki.esipfed.org/index.php/Community_Air_Quality_Data_System_Workspace
[14] http://wiki.esipfed.org/index.php/Air_Quality_Data_Systems_Assessment

Table 1. Some existing elements of a cyberinfrastructure for air quality management and research.

Data Repositories	
Observations	
AIRNow & AIRNow Gateway	http://airnow.gov, http://www.airnowtech.org
CASTNET	http://www.epa.gov/castnet
AIRQuest Data Warehouse	http://airquest.epa.gov
AQS & AQS Data Mart	http://www.epa.gov/ttn/airs/airsaqs/
NARSTO Data Archive	http://cdiac.ornl.gov/programs/NARSTO/
NASA Atmospheric Science Data Center	http://eosweb.larc.nasa.gov
NILU EBAS	http://www3.nilu.no/EBAS/
NADP	http://nadp.sws.uiuc.edu/
IADN	http://www.msc.ec.gc.ca/iadn/
Emissions	
EIS & NEI	http://www.epa.gov/ttn/chief/net/
EDGAR	http://www.mnp.nl/edgar/
Modeling	
CMAS Data Clearinghouse	http://www.cmascenter.org/
NCDC & NOMADS	http://nomads.ncdc.noaa.gov/data.php
AEROCOM	http://nansen.ipsl.jussieu.fr/AEROCOM/
HTAP	http://www.htap.org, http://htap.icg.fz-juelich.de/data
Data (and Analysis) Portals, Secondary Compilations	
DataFed	http://datafed.net
HEI Air Quality Database	http://hei.aer.com
VIEWS	http://vista.cira.colostate.edu/views/
NEISGEI	http://www.neisgei.org
GEIA	http://www.geiacenter.org/
GIOVANNI	http://giovanni.gsfc.nasa.gov/
GeoWeb	http://gds.rtpnc.epa.gov
RSIG	http://portal.epa.gov/rsig
IDEA	http://idea.ssec.wisc.edu/
GEO Architecture Implementation Pilot, Phase 2	https://sites.google.com/site/geosspilot2/air-quality-and-health-working-group
Data Analysis and Modeling Tools	
CMAQ	http://www.cmascenter.org/
EMF	http://www.ie.unc.edu/cempd/projects/emf/
AMET	http://www.epa.gov/ipbpages/current/v21/1190.htm
WRF-Chem	http://ruc.fsl.noaa.gov/wrf/WG11/
VERDI	http://www.verdi-tool.org/

SCOPE OF WORK

The primary purpose of this solicitation is to create a mechanism by which the existing elements of an air quality information system of systems can be assessed, linked, standardized, and extended to create a stable cyberinfrastructure. The successful contractor will not assume responsibility for the development and maintenance of all of the elements of the air quality cyberinfrastructure: generally, the various elements that comprise the system of systems will remain the responsibility of their respective developers and hosts. The contractor's team will, however, perform a number of functions, including

- assisting in the creation of a plan for the needed cyberinfrastructure
- developing standards, best practices, and tools
- building and maintaining new components of the cyberinfrastructre, where appropriate and efficient
- linking the existing and newly developed systems together

The successful proposal will present a team that has demonstrated capability in developing, maintaining, applying, and extending existing air quality cyberinfrastructure elements. As detailed below, the team will be expected to:

- Document functions, strengths, and weaknesses of existing air quality information systems or tools and the relationships between them, building upon recent, publicly available assessments of existing systems
- Identify or develop appropriate standards for nomenclature, formatting, documentation, archiving, exchanging, and processing air quality related data
- Network, , enhance, and develop air quality information systems to support air quality forecasting and public information, air quality assessment or "re-analysis", air quality model evaluation and intercomparison, emissions inventory development and evaluation, and fire and smoke management
- Connect the air quality cyberinfrastructure to other existing data networks and systems supported by EPA and the U.S. government
- Seek out and document feedback and input on the existing air quality data systems and the planned cyberinfrastructure from the research and policy support communities
- Provide logistical support for meetings and other outreach activities related to the air quality cyberinfrastructure, including activities organized under GEO and USGEO and related committees

The contractor shall perform the work specified in the work assignments in the following performance areas. The successful proposal will describe the general approach to these work areas and suggest work products that could be completed in the first year.

Work Area 1. Wiki-based Cyberinfrastructure Assessment and Planning Support

The contractor's team will be expected to develop and maintain a wiki to be used by the air quality community as the main forum for the development of an assessment of the current elements and tools for air quality data management. On the wiki, the contractor's team will lead efforts to:

- Identify functions, strengths, weaknesses of, and relationships between existing air quality information systems or demonstrations, including data systems or pilot projects developed by EPA, NOAA, NASA, USFS, States, and Regional Planning Organizations and the contributions of pilot projects funded under EPA's AMI.
- Identify or develop appropriate standards for nomenclature, formatting, documentation and metadata, archiving, exchanging, interoperability, visualization, and data processing for the cyberinfrastructure. To the extent practical, these standards will be compatible with GEO and USGEO standards, and harmonized with the community infrastructure being developed in the GEO AIP.
- Identify limitations and gaps in the current cyberinfrastructure which prevent air quality information from being fully and efficiently used for research and decision support. Identify opportunities for future development to close these gaps consistent with EPA's enterprise architecture plan, the Strategic Plan for the U.S. Integrated Earth Observation System, the U.S. Group on Earth Observation's Near Term Opportunity Plan for Air Quality Assessment and Forecasting, the international GEO 10-Year Implementation Plan, evolving GEO Architecture, and other relevant consensus recommendations.
- Provide guidelines for data providers on how to make data more available and accessible via emerging web service standards, web search tools (social networking, knowledge sharing), and other technologies.
- Provide guidelines for data analysts developing online and offline analytical or visualization tools to facilitate the generalization of their tools to other data sources and applications.

The evaluation and planning work described above is expected to build upon current, public assessments of existing air quality information systems.

Work Area 2. Cyberinfrastructure Network Development and Maintenance

The contractor's team will work to develop the air quality cyberinfrastructure by establishing and/or supporting appropriate data exchange connections between existing elements of the system of systems. Many existing elements for the needed cyberinfrastructure have been developed by EPA, other government agencies, and by academic institutions, in the United States and abroad. In the previous work area, the contractor's team will facilitate/lead both identification of opportunities for future development of the cyberinfrastructure and the definition of standards for establishing service-oriented linkages between these existing elements to enable processing, analysis, and visualization of air-quality data for public information, research, and decision support. It is hoped that the organizations responsible for the existing elements will be largely responsible for implementing these connection standards, each contributing to the formation of the cyberinfrastructure.

However, the contractor's team will be responsible for establishing some of the initial, high-priority connections to build a core cyberinfrastructure. Priorities for this Work Area are discussed below and will be further informed the by existing assessments,[14] activities (*e.g.*, AIP-2), and feedback received in Work Area 1, in consultation with the Project Officer. In some cases, the contractor's team may need to take over the hosting of some elements, so that they can be publicly available and maintained and updated for operational use. Transitioning these elements into a sustainable cyberinfrastructure may provide an opportunity for efficient standardization.

A focus of this effort will be to ensure the connectivity and accessibility of the relevant data repositories of EPA, NASA, NOAA, Regional Planning Organizations, and state and local air quality agencies through a broader cyberinfrastructure. A high initial priority will be to contribute to GEO Sub Task DA-09-02e Atmospheric Model Evaluation Network. The initial efforts under this task have focused on connecting the various data centers in the United States and elsewhere that are currently contributing to the assessment work of the TF HTAP, thereby creating an international backbone for an air quality data network, as well as contributing to the work of the TF HTAP itself. These data centers, listed in Table 1, are HTAP, Datafed, Giovanni, AEROCOM, and NILU EBAS.

Another priority for this effort is to ensure compatibility and connections to AIRNow (including AIRNow International and AIRNow Gateway) and related systems for air quality forecasting and public information and other information systems being developed under GEO and WMO. Over time, the connections should be extended within the United States to regional, state, and local agencies and globally, by reaching out to data sources in Canada, Mexico, Europe, Asia, and elsewhere.

A third priority for this effort is to maintain and expand development of a distributed network of emissions information begun in NEISGEI by working with states, regional planning organizations,[15] and global emission inventory efforts, including GEIA and EDGAR.

Work Area 3. Analytical and Visualization Tool Development

Where appropriate and efficient, the contractor's team will develop a set of processing, fusion, visualization, and analytical tools for air quality data to support the three capabilities described earlier:

- Air Quality Forecasting and Public Information
- Air Quality Data Analysis, Modeling, and Assessment
- Emissions Inventory Development and Evaluation

The software should have a service-oriented architecture and to the extent practical should be written with open source code and built around current standards, especially as those are identified and developed under Work Area 1. The contractor's team will be expected to build upon existing efforts and elements of the cyberinfrastructure (see Table 1), working with those who develop and maintain that software. In particular, the tools should

[15] http://www.epa.gov/air/visibility/regional.html

- Build upon existing model evaluation and visualization tools, including EPA's Air Model Evaluation Tool (AMET), Visualization Environment for Rich Data Interpretation (VERDI), and Remote Sensing Information Gateway (RSIG).
- Be compatible with the CMAQ and WRF-Chem models and the formats of the HTAP intercomparison and other international modeling standards
- Review and build upon the data fusion tools developed by the FY06-07 AMI projects (which would be documented in the wiki described above in Work Area 1) which aimed to develop methods or tools to integrate multiple types of observations and/or model estimates to create a more rich description of air quality for different applications.
- Use observational data from ground-based and space-based instruments drawn from across the air quality data network
- Enable comparison and compilation of model-ready emissions inventories from available data sets (including re-projection and re-gridding tools) and comparison of emissions data to observations or activity data.

Work Area 4: Outreach and Training

The contractor, in consultation with the Project Officer, will be responsible for organizing an annual meeting of the contractor's team with the Project Officer, other EPA staff, and other stakeholders from the air quality management and research community to discuss the needs and opportunities related to the development and maintenance of the air quality cyberinfrastructure.

The contractor's team will be expected to help organize meetings and other outreach efforts to educate and communicate with the broader air quality management and research community concerning the emerging cyberinfrastructure. The contractor's team should leverage existing organizations and forums within the air quality community to raise awareness, provide training, and gather feedback concerning the elements of the air quality information system of systems.

The outreach and training work described here will also assist in building and activating an Air Quality Community of Practice (CoP). In the GEO framework, a CoP has an important role in harvesting feedback from a discipline / community on the architecture and implementation of GEOSS and related activities. The CoP should be an international group, and the contractor's team should seek international participation in the activities described in this Work Area.

For planning purposes, the contractor's team will be expected to organize at least one workshop per year focused on this project and to participate in at least two other meetings per year to communicate about the cyberinfrastructure to the broader community.

AGENCY CONTACT:

Terry Keating, U.S. EPA Office of Air and Radiation, 1200 Pennsylvania Ave., NW (MC 6103A), Washington, DC 20460. Phone: (202)564-1174; Fax (202)564-1554, or email keating.terry@epa.gov.