Data Management Challenges in Developing a Network of Distributed North American Emissions Databases

Stefan Falke
Washington University in St. Louis,
One Brookings Dr., St. Louis, MO 63130
stefan@me.wustl.edu

Gregory Stella
Alpine Geophysics, LLC
387 Pollard Mine Road, Burnsville, NC 28714
gms@alpinegeophysics.com

Terry Keating
U.S. EPA - Office of Air & Radiation
1200 Pennsylvania Ave. NW, Mail Code 6103A (Room 5442), Washington DC 20460
keating.terry@epa.gov

Brooke Hemming
U.S. EPA – Office of Research & Development,
Mail Code B-243-01, RTP, NC 27711
hemming.brooke@epa.gov

ABSTRACT

In an age of international air quality agreements and annexes, an approach is needed to integrate emissions data from multiple inventories to support public outreach, emission trends reporting, control strategy application studies, benefit analyses, and estimation of air quality in large regional areas. The Commission for Environmental Cooperation (CEC) has been working closely with North American federal environmental agencies to gather the latest emissions data which can be used to create an integrated picture of emission inventories. This paper presents results of a CEC sponsored study of approaches for the comparability of techniques and methodologies for data gathering and analysis, data management, and electronic data communications for promoting access to publicly available environmental information held by public authorities of each of the three participating countries.

Integration of North American emissions databases faces numerous challenges because the data are distributed among multiple servers and are heterogeneous in format. We outline challenges faced in developing a network of distributed emissions databases and offer some solutions for addressing these challenges. A key solution in addressing data management issues is the application of new Web services technologies. A prototype web browser interface for accessing, exploring, and visualizing distributed emissions data sources using web services is presented.

INTRODUCTION

The inability to share and integrate environmental information among agencies and organizations has lead to inefficiencies and ineffectiveness in environmental management and policy. A recent U.S. General Accounting Office (GAO) report attributed a lack of effective collaboration among the agencies involved in the management of forest fires to the heterogeneous information systems employed by the numerous agencies and their inability to adopt new information technology to address these issues[1]. While the report singles out forest fire management, its critique of effective information systems extends to most other environmental areas and is reflected in new efforts among federal and international organizations to address similar issues. The National Science Foundation[2], U.S. EPA[3], NASA[4], and other federal and international organizations have initiated efforts to address these challenges in data integration among diverse data sources.

A common goal among these programs is to make environmental data easier to access in forms relevant for a variety of end applications. In general, these programs are centered within specific sub domains of environmental science. A step-wise method, beginning with domain-specific applications and gradually expanding in scope by connecting with other parallel efforts could lead to a multi-domain information network. Initially, every community, whether air quality management in general or emissions inventories in particular, will define its set of standards and protocols for unifying its data and its infrastructure.

Within the emissions inventory community, a real need is seen for a way to integrate emissions data from multiple inventories in order to support public outreach, emission trends reporting, control strategy application studies, benefit analyses, and estimation of air quality in large regional areas. However, without consistent emission data sets within the domain of study, results of these applications can be speculative. Recognizing the value in sharing emissions information, the three environment ministers of North America who comprise the Commission for Environmental Cooperation (CEC) Council agreed, in 2001, that their governments would work towards the development of a shared North American emissions inventory for criteria air pollutants and greenhouse gases. Working towards this goal of a shared inventory, the CEC began a pilot project in 2003 to explore the availability of emissions data for the electricity generating sector and the feasibility of sharing that data electronically between the three countries. The availability of emissions data has been summarized[5] and a demonstration of a distributed database of this emissions information has been prepared[6]. This paper presents results of the CEC study and the challenges faced in developing a network of distributed emissions.

One vision for a future integrated North American emissions inventory is provided by the Networked Environmental Information Systems for Global Emissions Inventories (NEISGEI), a U.S. EPA initiated effort to create a web-based global air emissions inventory network that provides a catalog of distributed emission inventory data, tools for processing and analyzing the data, means for registering new data, and an environment for collaboration among air quality researchers, policy-makers, and the interested public.[7]

The development of an integrated network of emissions data, tools, and community is being pursued incrementally as several pilot projects (including the CEC project described in this paper) are underway with the hope of laying the foundation for the network. The initial projects were designed to incrementally develop the information technology required to support user access to existing local, regional and global pollutant inventories, and to provide interoperable tools for merging and manipulating these heterogeneous data-sets for modeling and policy analysis. A design goal in the infrastructure development is that interfaces will be built on top of existing inventories to make the inventory data accessible by network, while the data remain distributed and under the control of the institutions responsible for developing the emission inventories.

The successes of these initial projects, as well as advances made in related distributed information system research and development will be leveraged to begin building operational networks of distributed emissions data and analysis tools among a consortium of users and inventory developers. Other efforts with relevance to integrated emissions inventories include EPA's Environmental Information Exchange Network[8], efforts by the Regional Planning Organizations in support of the Regional Haze Rule[9], and broader interagency efforts such as Geospatial One-Stop.[10]

DISTRIBUTED, YET INTEGRATED

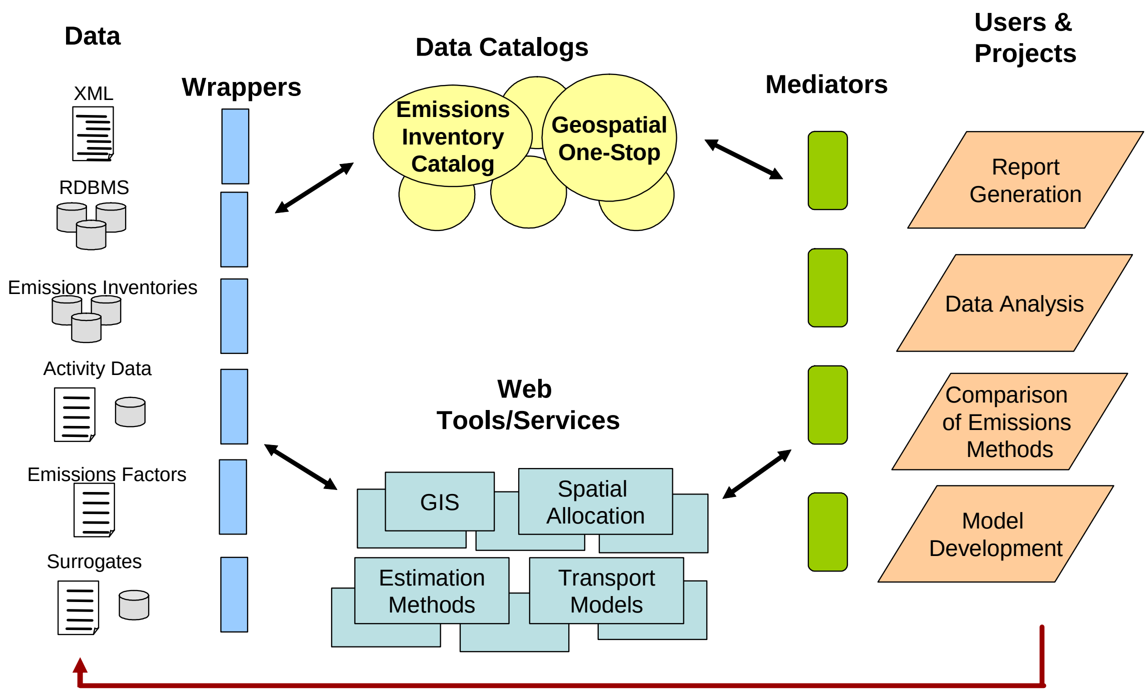
The overarching challenge in developing an integrated emissions inventory is how to integrate data that are distributed among many sources without requiring strict data format standards or introducing a new data repository to centrally store and maintain the data. The guiding principles of an integrated emissions inventory follow those of distributed databases and computing. The design objectives are to create a network of data and tools that is characterized by the following attributes.

- Distributed. The data sources remain distributed and in the control of their providers. The data are dynamically accessed through the internet rather than through a central repository.
- Non-intrusive. Data providers are more likely to participate if joining an integrated network does not impose new or additional burden on them.
- Transparent. The distributed data should appear to originate from a single database to the end user. One stop shopping and one interface to multiple data sets are desired without required special software or download on the user's computer.
- Flexible/Extendable. An emissions network should be designed with the ability to easily incorporate new data and tools from new nodes joining the network so that they can be integrated with existing data and tools.

Benefits of a Distributed, yet Integrated Inventory

A long term goal of a North American inventory is to be universally available to all who want to access its information, of high-resolution and source and facility specific, comprehensive with respect to pollutants and sources, well documented, and based on comparable methodologies and factors. An envisioned end state of a distributed emissions inventory is depicted in Figure 1. Distributed data sources (emissions estimates, activity data, surrogates, etc) in a variety of formats (relational database management systems, text files, etc.) are available through the Internet and registered in one or more data catalogs. These data can be uniformly accessed with the aid of data wrappers (translators) and connected with web tools and services to support a variety of end applications. Mediators are used to find and combine the appropriate mix of data and services to fulfill a user's task.

Figure 1. Conceptual Diagram of a Distributed Emissions Inventory



Current online inventory management systems are designed to receive, store, process, display and output emissions data[11,12]. These systems are also being modified to receive, store, process, and display combinations of the activity data and emissions calculation methods used to estimate emissions inventories. Continued work is underway to provide inventory display capabilities which will include

GIS functionality, tabular and flat file data formats, graphs and charts, and the ability to capture these displays in user-defined report formats.

The development of a distributed emissions inventory network will not replace operational systems for querying and downloading emissions inventory data but will be joined to these systems from two perspectives. First, the existing emissions inventory databases are the originating source for emissions data within the integrated network. Second, the display and analysis capabilities available in the integrated network will supplement what is available in a particular inventory management system. A distributed system will open the emissions inventories to a broader set of users and will increase availability, review, and ultimately the quality of emissions inventories.

PROTOTYPE INTEGRATED INVENTORY TOOL

One of the objectives of the CEC study was to examine the feasibility of employing new information technology to build a web tool that could dynamically integrate heterogeneous emissions data from Canada, Mexico, and the U.S. The construction of a prototype tool required two primary inputs, 1) emissions data and 2) software and infrastructure to build web tools. Emissions data from online and offline sources were identified and used as the data sources. The infrastructure and web components used to build the tool were taken from DataFed.net.

Data Used

A report that identified and summarized available North American emissions inventories was generated as part of the CEC project[5]. Table 1 lists the available online emissions inventory data used as part of the prototype integrated inventory tool. These are publicly available, on-line accessible emissions data. Other data resources are available, including Mexican emissions data that were electronic format from the Big Bend Regional Aerosol & Visibility Observational Study (BRAVO) emissions inventory.

Table 1. Online emissions inventories accessed as part of this project.

Data Source	Time Coverage	Pollutants	Reporting Level
NEI (US)	1985-1999 (criteria) 1996-1999 (HAPs)	NOx, SO2, CO, PM, VOC, HAPs	Boiler
eGrid (US)	1996-2000	NOx, SO2, CO2, Mercury	Boiler & Generator
Clean Air Markets (US)	1980, 1985, 1988-1999	NOx, SO2, CO2	Generator
NPRI (Canada)	1994-2001	HAPs (Criteria starting in 2002)	Facility

Emissions inventories are based on different underlying data models. Each inventory uses a uniquely defined set of field names. However, many of these field names are similar to (or their content is similar to) fields in another country's inventory. In mapping between datasets, some of the key relationships among the inventories were captured. These mappings provide a set of connections that can subsequently were applied to automated query and integration of data from multiple inventories.

DataFed.net

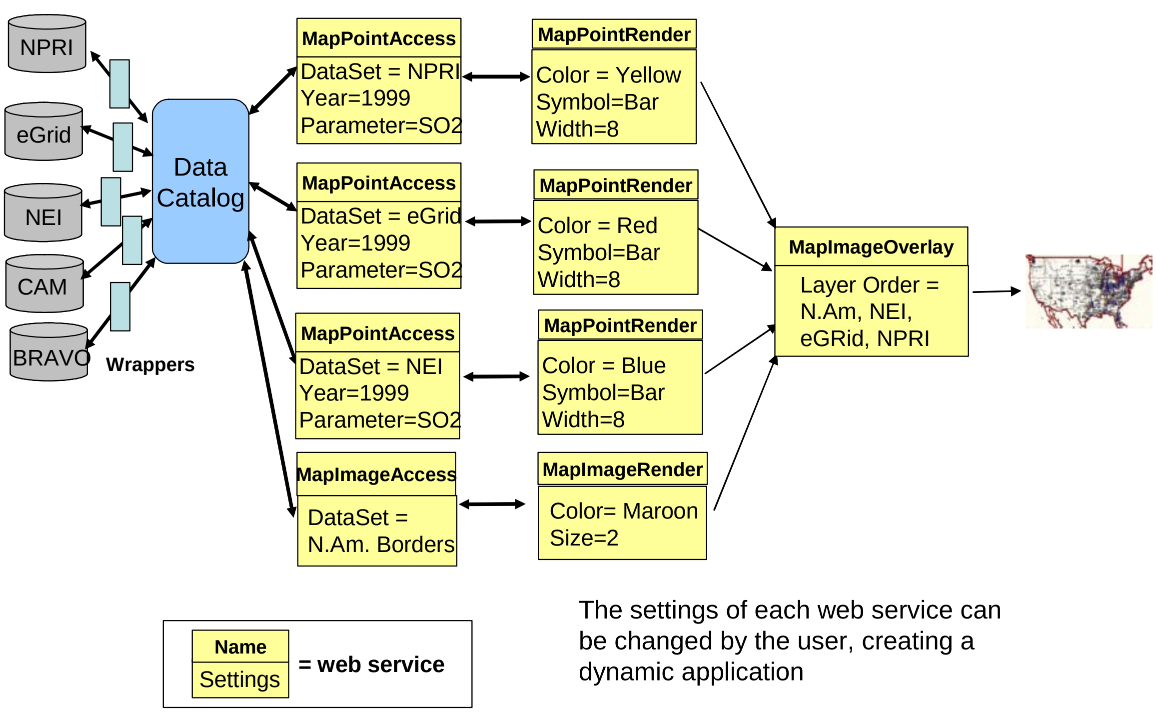
DataFed.net is a web-based infrastructure that supports data sharing and processing for collaborative air quality management and atmospheric science research (www.datafed.net). The emissions data were registered in the DataFed.net catalog where the registered data access instructions can be interpreted for browsing and visualizing the data. DataFed.net provides mediator software for creating data "views," including maps, time series, and tables, of data that are distributed among multiple web servers. The views are each created using web service thereby allowing them to be used and reused in custom applications with standard web programming languages.

Approach

The approach used in building a prototype integration emissions tool initially focused on accessing relevant data sets and acquiring or developing the necessary information technology. The data were registered in the DataFed.net data catalog. The data registration process includes describing the data's spatial, temporal and parameter properties as well as access instructions for retrieving the data from its source. The data access instructions are described in order to allow queries to be created and run based on space, time, and parameter conditions. The query results can then be used to display, compare or otherwise process the data. The data registration and access descriptions are referred to as a data wrapper." The wrapper is specific to a data type (e.g. a lat/lon point data set in a relational database). Emissions datasets that could not be dynamically accessed and for which wrappers could not be created were manually downloaded and stored in a relational database on a local server.

Web services are software used over the Web interfaces. They are self-contained and use XML-based standards for describing themselves and communicating with other web resources, thereby allowing them to be reused in a variety of independent applications. Because they are designed to be independent of any particular database platform, they are ideally suited for building a distributed database and tools network. Web services from DataFed.net were used to build the prototype tool.

Figure 2. Example of web services used in accessing, rendering and creating a map view.



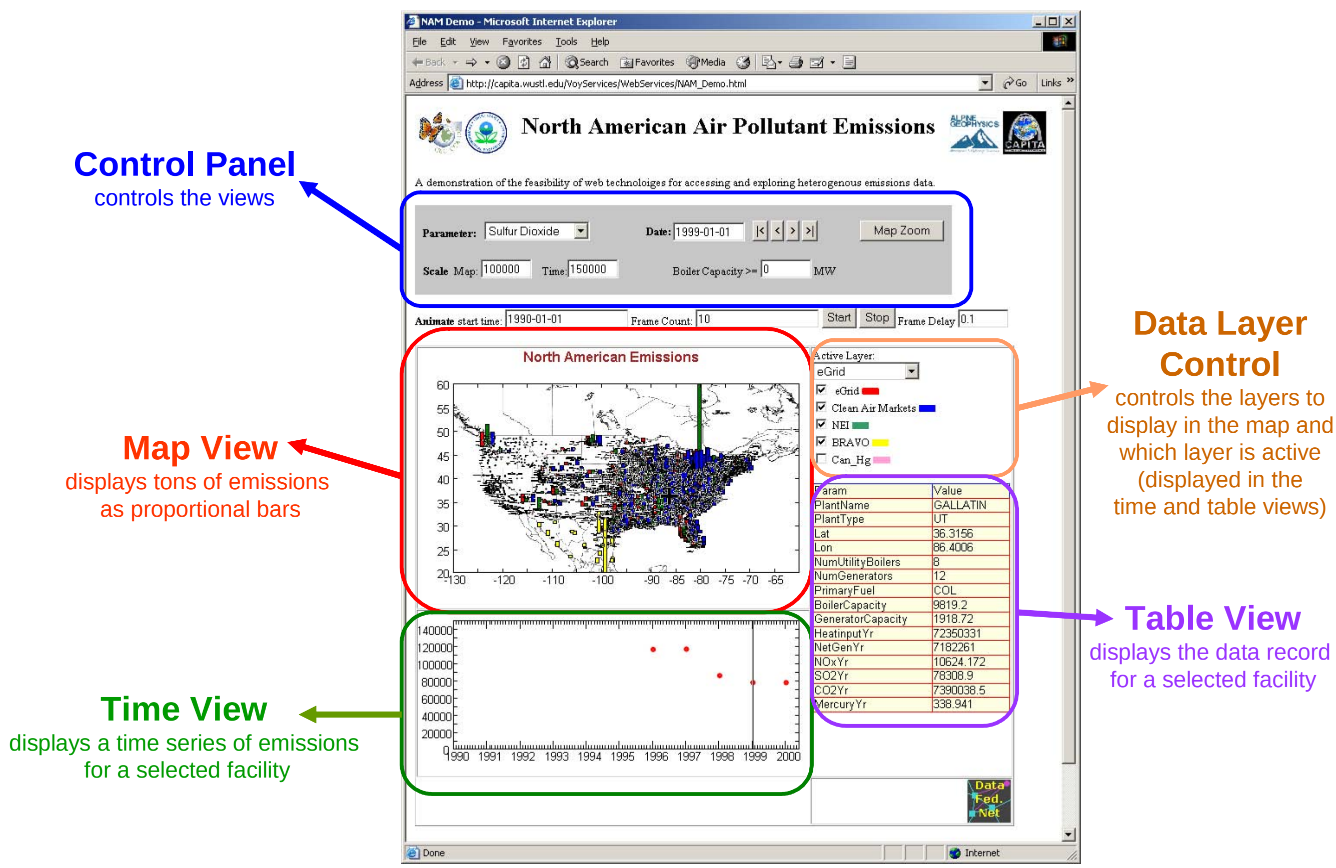
DataFed.net includes web services for accessing and displaying data in map, time series, and table views. Figure 2 illustrates an example of using web services to generate a map view. Emissions data are registered in a data catalog along with their access instructions. EGU emissions data are available a specific latitude/longitude coordinate points and are therefore registered as point datasets. DataFed.net includes a point access service with settings for such variables as dataset name, time period, and parameter. The executed MapPointAccess web service executes a query to the dataset's data source and returns a dataset. This dataset is that passed through the MapPointRender service where the display variables are set. The data along with its rendering settings are used to create a map which is subsequently combined with other maps using the MapImageOverlay service to arrive at the final map view.

The settings of each DataFed.net web service can be changed by the user, creating a dynamic application. The individual views created with web services can be embedded in a standard HTML web page. Javascript is used to allow the views to be dynamically updated. Controls, in the form of text boxes, select lists, and buttons, allow the user to manipulate the views. The prototype tool for exploring integrated emissions data is essentially an HTML page with a map, time series, and table views that linked together and controlled to enable emissions data exploration, querying and analysis.

Result

The multi-dimensional nature of emissions data are the center piece of the prototype integrated emissions inventory tool. Its multidimensionality (plant, year, pollutant, fuel type, boiler capacity, etc.) is displayed in multiple “view” in the tool. Figure 3 is a screen shot of the tool and identifies the tool’s components. The tool consists of three views (map, time, and table) and two controls for manipulating the views.

Figure 3. Components of the Integrated Emissions Inventory Web Tool.



The *map view* displays EGU plant locations and their associated tons of emissions as proportional bars. Multiple emission inventory data can be superimposed on the map by selecting the datasets in the *data layer control*. The pollutant and year of the emissions displayed in the *map view* can be adjusted in the *control panel*.

The *time view* displays a time series of tons of emissions for a particular facility. The facility is selected by clicking on a facility in the *map view*. If multiple datasets are displayed in the map view, the active layer in the *data layer control* is used to designate which dataset to display in the *time series*.

The *table view* displays characteristic data of the currently selected facility in the active layer dataset.

In addition to setting the pollutant and year, the *control panel* also allows the user to adjust the scales of the proportional bars in the *map view* and the y-axis scale in the *time view*. An example query capability is included that filters the emissions displayed in the *map view* based on a threshold boiler capacity value.

The prototype tool is available for testing and comment at http://capita.wustl.edu/NAmEN.

The prototype tool provides a simple to user interface for exploring and visualizing heterogeneous emissions inventory data. At the initial demonstration of the integrated emissions tool, potential users mentioned the ability for the tool to save the time and effort involved in finding, collecting and formatting international emissions data generated outside of current programs. Additionally, the tool's usefulness in exploring these data and comparing emissions from multiple sources was commended.

The presented prototype tool is but one of many that could be assembled. Reusable views can be reconfigured, reconnected or rearranged to create other web tools to serve different end user needs.

CHALLENGES AND OPPORTUNITIES IN BUILDING DISTRIBUTED INVENTORIES

The feasibility of a future integrated North American emissions inventory has been demonstrated as a web tool for browsing heterogeneous emissions data sources. From a general distributed database technology perspective, we are at a point where distributed database concepts can be applied to actual implementations. However, when focusing on the development of a distributed air emissions inventory, the CEC pilot project encountered numerous technological and organization challenges in dynamically accessing the currently available emissions inventories.

A summary of identified barriers to distributed emissions inventories is provided in Table 2. The data actor column in Table 1 indicates a stage in the access and use of emissions data. Data providers are the organizations that store and maintain the emissions inventory database. Mediators are at the intermediate stage and provide the necessary processing and tools for translating data into uniform access. The users are the end customers of the data who apply the data into their specific applications. Each stage has its own set of requirements for a successful distributed data network.

Table 2. Distributed Emission Inventory Barriers

Data Actor	Distributed Network User Issues / Requirements
Provider	Technology implementation Server and database security Bandwidth limitations Data misuse
Mediator	Consistent and stable access to data provider Database mapping
User	System performance and responsiveness Easy to use interface

Organizational Challenges

The challenges in attaining distributed data and tool networks encompass both technological and organizational aspects. It is generally argued that technological innovation has reached a state where distributed data networks could be meaningfully implemented.[13] Assuming that argument to be true, organizational and cultural barriers are the most significant to such a network being deployed on a substantial scale. The institutional history of government organizations tends to be defined by information systems and applications designed for specific purposes. In many cases, these systems are developed and maintained by contractors with narrowly defined contracts. Changing these contracts to reflect a distributed data sharing approach is a non-trivial and potentially costly endeavor.

In building a consortium of data providers, it is imperative that the providers can see a benefit, or return on investment, from joining the network. Joining for the sake of the community at large is not sufficient. Data providers must stand to gain from sharing their data. Some potential benefits include increased exposure and use of their data, access to other data sources, access to tools that add value to their data, and easier methods for collaborating with other organizations.

Another critical issue is maintaining appropriate acknowledgement and recognition for the data provider's information. Even though the data remain physically within the purview of the data provider in a distributed network, the front end to that data can be located anywhere on the network. A third party interface to an organization's data can potentially give the impression that the data are being served from the third party and in the process lose the credit due to the data provider. Ensuring credit for contributions is a priority in the design of distributed data networks.

An additional hurdle to be addressed is data misuse. Data have inherent limitations in their relevancy to questions they can answer. In a centralized system, data distribution can be limited and therefore inappropriate applications of those data controlled. An openly shared system could potentially lead to greater use of data in contexts not intended by the original providers. On the other hand, a shared system would lead to greater use of data and improved community-wide recognition of a data set's limitations.

Technical Challenges

Despite the many technological advances, significant needs remain before distributed networks will be accepted within scientific and policy communities. Perhaps most significant is security. A data provider must have assurance that making their data available through a distributed network will not adversely impact their operations. Concerns include an increased volume load on their servers which could lead to disruption of their mission-specific operations as well as security breaches due to opening their databases to the outside world.

The underlying emissions data presents its own set of challenges. In many cases, data are not inherently accessible. Emissions inventories are currently not designed for such application and while most of the emissions inventories used during this project were available through the Internet; their web access methods only support single user access. Attempts were made to automate a manual approach through an internally hosted web server but those attempts failed to produce a stable, reliable method for accessing the data. Most access software utilizes some authentication technology that prevents dynamic, server-side access. For example, the products employed by U.S. EPA's databases are designed for single user access through a desktop computer. This limitation prevented automation of the dynamic access of U.S. EPA databases through a web server interface. Security is and will continue to be an important concern in distributed data access and is one of many issues yet to be resolved. Recommendations and approaches to addressing these challenges so that data access can be dynamic and secure are presented in later sections.

Consensus derived standards and protocols are still missing in many aspects of distributed computing, particularly in describing and defining the services for making data available and accessible through distributed tools. It is reasonable to expect that, as distributed computing becomes commonplace, these standards will stabilize and promote the expansion of distributed data networks.

An effective distributed system should be responsive to the user. Accessing large datasets, such as multi-dimensional national emissions inventories with thousands of emission point locations, is currently too cumbersome for efficient user interaction. This performance limitation should not be considered insurmountable. The continually expanding bandwidth of internet networks and more efficient algorithms for handling distributed data promise to make distributed systems fast enough for everyday use by researchers, managers, and the public.

The organization and technical challenges outlined above are certainly surmountable and collaborative efforts in the future are likely to generate an operational distributed North American emissions inventory. The development of this inventory would benefit from a step-wise approach that initially focuses on the most readily available and multi-country comparable data. The preliminary versions of the inventory would help clarify issues related to handling complex queries. Building and using initial versions will assist in creating consensus approaches to issues as straightforward as data naming conventions that could make the exchange of emissions data among the three countries even simpler.

Addressing the Challenges

In progressing toward a distributed emissions inventory, it is important to keep the goals of a distributed emissions inventory at the forefront:

- Minimum burden on data providers;
- Shared and distributed data; and
- Uniform and transparent user interfaces to data.

We contend that the progression toward a distributed emissions inventory can be promoted by continuing to upgrade the state of a distributed emissions inventories using technologies and techniques that do not impose additional burdens on data providers and by encouraging emissions inventory managers to adopt new technologies that foster the sharing of their data with external clients. Among these technologies are web services and related standards, such as the OpenGIS Web Map Server and Web Feature Server.

The implementation of an operational distributed emissions inventory will be realized sooner by focusing on defining a process for dynamically linking emissions inventories rather than imposing data format, software, and hardware standards. Some next steps that could assist in achieving distributed data networks include:

- More complete access to distributed datasets - A process for creating trusted provider-user agreements that would help address issues of security and data misuse.
- More comprehensive content – Current efforts in creating distributed information systems will make a diverse set of data and tools available that could spark additional interest in the technology’s potential;
- Integration – Linking current distributed database efforts together with one another will create a broad base of data and tools and will serve as important examples in testing and demonstrating the effectiveness of distributed databases;
- Metadata .More complete description information about emissions databases would help in relating heterogeneous data. Efforts to use FGDC metadata and the development of standard data catalogs, such as Geo Spatial-One Top are beginning to address this.

The adoption of new approaches to data management and use of information technology will be driven from many sides, such as the desire to make federal data available through Geospatial One-Stop and the goal of integrating state emissions inventories.

Many approaches can be pursued in developing distributed emissions inventories. Two approaches are presented here. One approach uses currently implemented technologies that allow a distributed database to be created through the application of data caching. The second approach focuses on web service technologies that would require additional technology implementation by the data provider in order to become a node on the network. Both of these approaches are being pursued by DataFed.net and embrace the concept of mediated data access; a middle component between the data provider and data user that adapts to the data and user needs in fostering their interaction.

Cached Data Approach

Emissions databases are already available through online interfaces and many can be automatically queried through single user access accounts. These allow single users to download data but do not allow other distributed servers to handle multiple user queries and pass them along to the emissions database. A solution to this problem is to use a mediator structure that dynamically accesses the data and store it as a cached dataset on the mediator's server. This mediator could then supply the data using web services that allow distributed access while still maintaining the original link to the data source.

Instead of accessing the data directly from a database each time a query is executed and then discarding the results after the user is finished with that query, this data system would store the retrieved data in a cache and in a format that allows for efficient access, querying, and analysis of large, multidimensional emissions data.

A cached data solution would avoid the data provider issues outlined in Table 2 as it would construct an intermediate, virtual instance of the data as designed for efficient user access. Instead of a user query going directly to (and burdening) an emissions inventory database, the query would instead only interact with an intermediate form of the data.

The cached data would contain a relevant subset of data and would be dynamically linked to the initial emissions database (thereby benefiting from the advantages of the original data) so that when updates occurred in the original data, the host system of the intermediate information would be notified and updated with appropriate changes. This would ensure that a shared, single version of the data was continually available to end users. Maintenance of the data cache would be automated with minimal associated cost. Human interaction would be required during setup to provide the mapping to the data provider's database but once the link with the data cache was established it would automatically update to reflect changes in the source database.

Data Web Services Approach

In the longer term, it is feasible to think about a peer-to-peer type of distributed access network. Such a network would allow direct access to each emissions database on the network after each data server implemented web services or some alternative web interface method of dynamically accessing the data. Because these web services are self-contained and use Extensible Markup Language (XML)-based standards for describing themselves and communicating with other web resources, they can be reused in a variety of independent applications.

In the web services network approach, mediators serve the role of brokers, providing users with the interfaces for finding available data, dynamically retrieving it, and integrating it with other distributed data sources. These network users can function on an independent level, each addressing local issues of importance. These individual components can then be integrated or modified to handle differing data types dynamically on demand.

Web service technology is still evolving and does not currently provide a convenient off-the shelf software solution. However, many required components are considered standards in peer-to-peer web programming applications and therefore make it possible to create an operational data web service. These components allow computer-to-computer communication in a platform- and programming language independent manner. Additionally, web service technology provides existing software applications with service interfaces without changing the original applications, allowing them to fully operate in the user's existing environment.

Both the near term and long term opportunities provide solutions to the challenges encountered in developing integrated emission inventories in a distributed nature. What continues to remain unclear is which method, combination of the methods, or other methods will best offer the final resolution to an operational distributed emissions inventory over regional, continental, or global scales.

CONCLUSIONS

The development of an integrated North American emission inventory that could be used for strategic planning and management of air quality is feasible and within reach. The technology is at a point where it can be applied in transitioning distributed database concepts to implemented solutions.

Emissions data present unique challenges due to their complex relational dimensionality, heterogeneous formats, and diverse sources. However, collaborative efforts in the near future could generate a distributed North American emissions inventory. An incremental approach that focuses on projects in particular locations, time periods, or pollutants for which there distributed database champions or adopted data standards will generate functional systems. The initial versions would help clarify the issues related to handling complex emissions data types and queries. Using these initial distributed emissions tools will assist in understanding the most effective process for creating more comprehensive emissions inventories.

Web services with their promise, and to some extent proven, ability to provide modular software components for disseminating and accessing data through web interfaces appear to offer a technological solution to address some of the barriers to distributed emissions inventories and are worth serious consideration.

Many challenges, both technical and organizational, remain but available and evolving information technology coupled with the desire by multiple government agencies for collaborative databases have charted a course to a truly integrated, yet distributed, network of emissions inventories a reality in the near future.

REFERENCES

1. General Accounting Office (GAO) "Geospatial Information: Technologies Hold Promise for Wildland Fire Management, but Challenges Remain", Report to Congressional Requestors, GAO-0301047 September 2003**;** http://www.gao.gov/new.items/d031047.pdf

2. Atkins, D.E.; Droegemeier, K.K; Feldman, S.I.; Garcia-Molina, H.; Klein, M.L.; Messerschmitt, D.G.; Messina, P.; Ostriker, J.P.; Wright, M.H. *Revolutionizing Science and engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Technical report, National Science Foundation, January 2003; http://www.cise.nsf.gov/sci/reports/toc.cfm

3. Trovato, R. "e-Gov and Ecoinformatics – Platforms for Collaboration" Presented at the Open Forum 2003 on Metadata Registries; http://www.epa.gov/sor/ecoinformatics_trovato.pps

4. NASA, 2003 Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS), http://lennier.gsfc.nasa.gov/seeds/

5. Commission for Environmental Cooperation (CEC), “Availability and Infrastructure of North American Electric Generating Utility Emission Inventories and Opportunities for Future Coordination”. Prepared for the Commission for Environmental Cooperation by Alpine Geophysics, LLC. October 2003.

6. Falke, S.; Stella, G.; Keating, T. “Demonstration of a Distributed Emissions Inventory using Web Technologies”. Presented to the Commission for Environmental Cooperation and U.S. EPA in Hull, Ontario. February 27, 2004; http://capita.wustl.edu/NAMEN/DemoPresFeb27.pdf

7. Hemming, B.; Falke, S.; Keating, T.. “Networked Environmental Information Systems for Global Emissions Inventories (NEISGEI)”, Presented at the NARSTO Workshop on Innovative Methods for Emission Inventories, Austin, TX. October 2003; http://www.neisgei.org

8. U.S. Environmental Protection Agency (EPA), 2004. National Environmental Information Exchange Network; http://www.exchangenetwork.net

9. Western Regional Air Partnership (WRAP), “Needs Assessment for Evaluation and Design of an Emissions Data Reporting, Management, and Tracking System” Prepared for Western Governor’s Association Western Regional Air Partnership by EA Engineering, Science and Technology July 25, 2003; http://www.pechan.com/edms/Task3_Final_Technical_Report.pdf

10. National Spatial Data Infrastructure (NSDI), 2004. Geospatial One-Stop, http://www.geo-one-stop.gov/

11. Solomon, D.; Pope, A.; Tooly, R. “Managing EPA’s Emission Inventory Databases”, Presented at the International Emission Inventory Conference, "One Atmosphere, One Inventory, Many Challenges." Denver, CO, April 30, 2001.

12. Environment Canada, 2004. Environmental Canada National Pollutant Release Inventory, On-line Data Search; http://www.ec.gc.ca/pdb/npri/npri_online_data_e.cfm.

13. Falke, S.R, 2002. *Environmental Data: Finding It, Sharing It, and Using It,* Journal of Urban Technology, 9(2), 111-124.

ACKNOWLEDGEMENTS

The authors would like to thank Paul Miller at the North American Center for Environmental Cooperation (CEC) for his comments and feedback. The work presented in this paper was supported by the CEC. Rudolf Husar and Kari Hoijarvi at the Center for Air Pollution Impact and Trend Analysis at Washington University in St. Louis designed DataFed.net and provided valuable guidance and suggestions in applying its concepts and services to this project. Stephanus van Schalkwyk assisted in developing data wrappers.

KEYWORD

Web services
Distributed databases
Emissions inventory
Electric utilities