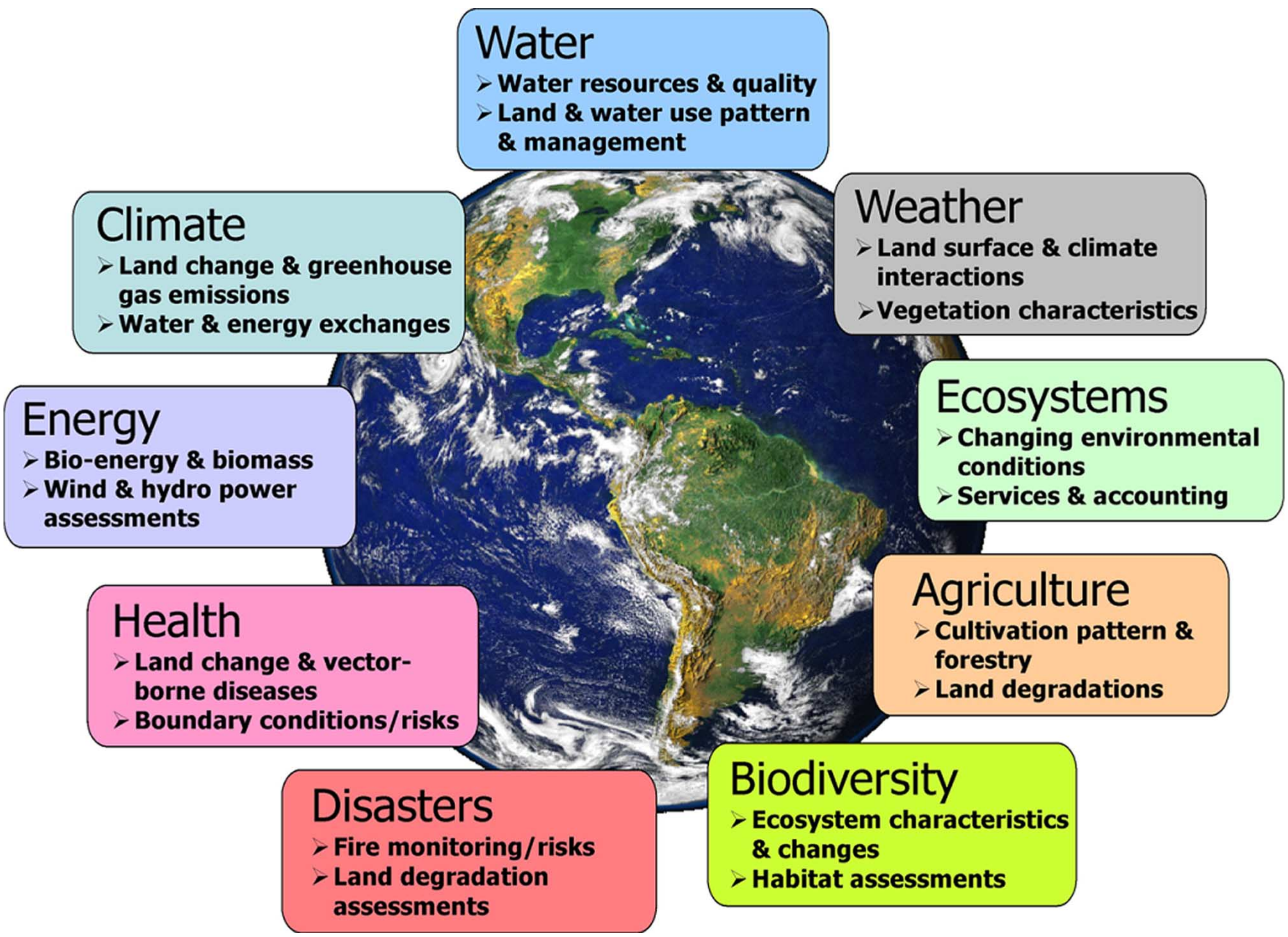
IEEE

# SYSTEMS JOURNAL

A PUBLICATION OF THE IEEE SYSTEMS COUNCIL

SEPTEMBER 2008 VOLUME 2 NUMBER 3 ISJEB2 (ISSN 1932–8184)




GEO areas of societal benefits and key land-cover observation needs emphasize the multitude of services from continuous and consistent global terrestrial observations.

# DataFed: An Architecture for Federating Atmospheric Data for GEOSS

Rudolf B. Husar, Kari Hoijarvi, Stefan R. Falke, Erin M. Robinson, and George S. Percivall

***Abstract*—DataFed is a distributed web-services-based computing environment for accessing, processing, and rendering environmental data in support of air quality management and science. The flexible, adaptive environment facilitates the access and flow of atmospheric data from provider to users by enabling the creation of user-driven data processing value chains. The approach of DataFed is mediation between users and data providers. DataFed non-intrusively wraps datasets for access by standards-based web services. The mediator software, composed of web services, provides homogeneous data views (e.g., geospatial, time views) using a global multi-dimensional data model. Application software written using web services are data browsers, including Google Earth, and analysis tools for distributed AQ data. Its federated data pool consists of over 100 datasets and the tools have been applied in several air pollution projects. From the point of view of GEOSS, DataFed contributes air quality data (as services) to the shared data pool through the GEOSS Common Infrastructure. It also hosts a Decision Support System (DSS) in the societal benefit area of air quality. The developers of DataFed actively participate in the GEOSS process included work with Architecture and Data Committee (ADC) and the User Interface Committee (UIC) as well as in interoperability experiments.**

***Index Terms*—Earth environment, information system, terrestrial atmosphere.**

## I. Introduction

RECENT developments offer outstanding opportunities to fulfill the information needs for atmospheric sciences and air quality management. High-resolution satellite sensors with global coverage now provide near-real-time snapshots which depict the spatial and temporal pattern of haze, smoke, dust, and other atmospheric constituents in stunning detail. The data from surface-based monitoring networks now routinely provide detailed chemical composition of the atmosphere. The "terabytes" of data from these surface and remote sensors can now be stored, processed, and delivered in near-real time and 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 in real and virtual workgroups.

Nevertheless, atmospheric scientists and air quality managers face significant hurdles. The production of Earth observations and models are rapidly outpacing the rate at which these observations are assimilated and metabolized into actionable knowledge that can produce societal benefits. The "data deluge" problem is especially acute for analysts interested in aerosol pollution, since the aerosol processes are inherently complex, the numerous relevant data range form detailed surface-based chemical measurements to extensive satellite remote sensing and the integration of these requires the use of sophisticated models. As a consequence, Earth Observations (EO) are under-utilized in making societal decisions. A remedy is anticipated from the Global Earth Observation System of Systems (GEOSS).

This paper is an early report on the application of the GEOSS concepts in the federated data system, DataFed. The paper focuses on the architectural aspects of the DataFed design, as a user-driven contribution to the emerging architecture of GEOSS. It is recognized that it represents just one of the many configurations that is consistent with the spirit of GEOSS. The implementation details and the various applications of DataFed are reported elsewhere [3]–[5].

## II. GEOSS and DataFed

GEOSS is an emerging public information infrastructure for finding, accessing and applying diverse data useful for decision makers [1]. GEOSS is proactively pursuing the linking together of Earth observing systems so that data can be combined and turned into actionable knowledge for many societal benefit areas (SBAs). The Group on Earth Observations (GEO)[1] is the body that coordinates the construction of GEOSS by the year 2015, it ensures universal access to EO data as public good and helps building trust and collaboration among the diverse stakeholders in GEOSS.

Earth observations are conducted by all nations, by diverse organizations, and performed through many independent

Manuscript received July 2, 2007; revised July 10, 2008 and July 28, 2008. Current version published September 17, 2008. The core design and implementation of DataFed was supported in part (2001–2004) by the National Science Foundation under Grant 0113868, by US EPA, by NASA under Grant 1322-59743, by EPA, and by the Regional Haze, Regional Planning Organizations.

R. B. Husar, K. Hoijarvi, and E. M. Robinson are with Washington University, St. Louis, MO 63130 USA (e-mail: rhusar@wustl.edu; hoijarvi@wustl.edu; emr1@wustl.edu).

S. R. Falke is with Washington University, St. Louis, MO 63130 USA and Northrop Grumman, St. Louis, MO 63101 USA (e-mail: stefan@wustl.edu).

G. S. Percivall is with the Open Geospatial Consortium, Crofton, MD, 21114 USA (e-mail: percivall@opengeospatial.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSYST.2008.2003292

[1]The group on Earth Observations was established by a series of three ministerial-level summits. It currently includes 68 member countries, the European Commission, and 46 participating organizations.

projects. Hence, centralized design and management of a Global Earth Observation System is not an option. A unique contribution of GEOSS is the adoption and promotion of the advanced "system of systems" (SoS) approach toward the integration of the multiplicity of autonomous EOs and models. SoS is a fledgling field of science and there are a number of different definitions and interpretations. According to an authoritative assessment [2], SoS consists of autonomous constituents that are managed independently, the constituents evolve independently and an SoS composed of such constituents acquires an emergent behavior. Furthermore, in SoS no stakeholder has a complete insight and understanding; central control is limited; distributed control is essential and the users/stakeholders must be involved throughout the life of SoS.

Given these unusual attributes, GEOSS is an ambitious, untried, and somewhat risky undertaking. It requires rethinking much of the traditional systems approach that was applied to the design of individual EO systems. Clearly, the SoS approach places considerable burden on the architecture, implementation, maintenance, governance and the overall functionality of the information systems of GEOSS. These substantial challenges can only be matched and exceeded by the major societal benefits that GEOSS may produce.

Since about 2001, our research group, along with others and with the support of multiple US agencies, has been developing the federated data system, DataFed, with the vision of supporting air quality/atmospheric composition management and science by more effective use of relevant data. DataFed is not a centrally planned and maintained data system but a facility to harvest and harness the growing EO resources on the Internet by powerful dynamic data integration technologies and through a collaborative federation philosophy.

The specific goals of DataFed are: 1) facilitate the access and flow of atmospheric data from provider to users; 2) support the development of user-driven data processing value chains; and 3) to participate in specific air quality application projects. The federation currently mediates access to over 100 datasets which includes both near real-time and historical observations and models. Since 2004 DataFed has provided IT support to a number of air quality management applications. Virtually all the content and a record of its evolution is accessible through the community workspace wiki at datafedwiki.wustl.edu. DataFed is now an applied system used in everyday research by several air quality analysis groups. Hence, DataFed is an autonomous information system (IS) with a purpose of its own, managed and evolving independently.

From the point of view of GEOSS, DataFed has two roles.

1) It contributes air quality data (as services) to the shared data pool through the GEOSS Common Infrastructure (GCI).
2) DataFed is a Decision Support System (DSS) that transforms data into useful knowledge for decision makers in the societal benefit area of air quality.

As a data contributor, DataFed is a registered component of GEOSS. It publishes an array of services for standards-based access to a broad range of atmospheric composition and emission data though its catalog. The data catalog is accessible for harvesting by the GEOSS Clearinghouse and the registered services will be broadly accessible through the evolving GCI. In this way, the autonomous DataFed system is a member of the loosely coupled GEOSS SoS.

The data processing, i.e., decision support part of DataFed is a consumer of standards-based data access services. The current networking to the distributed data services accomplished through *ad-hoc* established links. As the GEOSS-mediated data pool grows, we intend to harvest the accumulating communal resources to benefit the air quality community.

From our research group's perspective, the rationale for linking DataFed to the GEOSS common infrastructure is largely self-interest. As new air quality-related GEOSS components and standards-based services will be registered by other systems, DataFed will be able to access and utilize those "for free" and benefit from the "network effect" made possible by the GEOSS service-sharing infrastructure.

Since several goals and needs of DataFed broadly coincide with those of GEOSS, a significant fraction of the DataFed effort was invested toward connecting and coevolving with the GEOSS program. Since 2005, the connections to the GEOSS process included work with Architecture and Data Committee (ADC) and the User Interface Committee (UIC) of GEOSS. The work with ADC involves linking DataFed as a provider/user to the GCI. Work with the ADC consisted of participation in interoperability experiments, architectural studies on SoS, and demonstrations of loosely coupled applications using Service Oriented Architecture. More recent work with the UIC involved refining the user requirements for air quality. DataFed being a DSS for several air quality management programs provides a suitable testbed for developing use pattern, defining user classes and studying data user-producer communication mechanisms. An additional link to UIC is participation in the organization of an air quality Community of Practice (CoP) of GEOSS, which is in progress.

## III. DATAFED ARCHITECTURE

Architecture is a broad concept with many meanings. We are aware of at least two complementary frameworks for describing the architecture DataFed. The first framework is offered by the SoS community which focuses on the human stakeholders of SoS. These include *users*, i.e., the people who benefit from system, *developers* who construct the system, *acquirers* who contract and pay for the system, *testers* who evaluate system for suitability, *sustainers* who keep the system up to date, *trainers* who insure that the users know how to use it and *researchers* who provide the next generation of ideas [2]. The architecture of the enterprise is then described in terms of the activities of these stakeholders and the system components and connectivity that they require.

Another, more formal, approach describes a system's (not SoS) architecture using the reference model for open distributed processing. RM-ODP,[2] which is used the ISO 19100 series of geographic information standards. The GOESS architecture is also described using the RM-ODP framework. Using this reference model, the information system (IS) is described from five different points of view: *enterprise*; *information*; *computational*; *engineering*; and *technology* using a specific language

[2]The reference model of open distributed processing (RM-ODP: ISO/IEC 10746) provides a coordinating framework for the standardization of ODP.

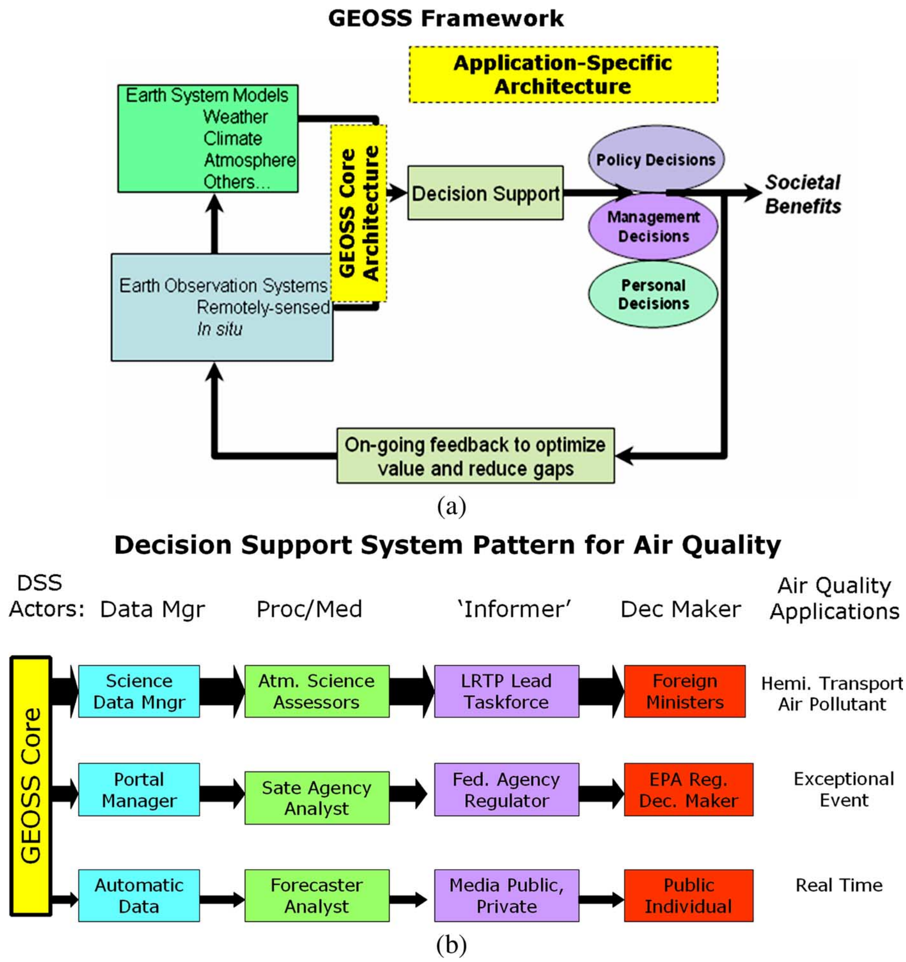



Fig. 1. (a) GEOSS Integrated System Solution framework. (b) Decision support system pattern for air quality.

applied to each viewpoint. We are not aware of a reference model that is applicable to both systems and system of systems. For sake of simplicity and fluidity, in this report we use RM-ODP viewpoint categorization as a loose guide.

### A. *Enterprise Viewpoint—Value Creation From Earth Observations*

The enterprise viewpoint focuses on the purpose, scope, and policies of the SoS IS. From the perspective of the entire GEO Enterprise, the purpose of the system (of systems) is to provide societal benefits by better informed decision-making through the use of EOs and models. The overall model of the GEO Enterprise is shown in Fig. 1(a), depicting a sensory-motor feedback loop. EOs and models provide the sensory and scientific input into the Decision Support System (DSS) which, in turn, produces actionable knowledge for decision-making at policy, management and personal levels. The adaptive functionality of GEOSS is maintained by the feedback loop which drives the adjustments to the observations, models as well as to the GEOSS infrastructure. For EOs and models, the role of GEOSS is to facilitate universal access to EO data as public good. This is to be accomplished by the GEOSS Core Architecture which serves as a broker between service providers and service users. Through this mediation, the GEOSS core infrastructure acquires the characteristics of a "value network" [6]. The core architecture that mediates the service exchange is agnostic of the end user and the SBAs it may serves.

The right side of Fig. 1(a) includes the decision section, depends strongly on the application area where the societal benefits are to be derived. For air quality, the components of the decision making system can be expressed in more detail in Fig. 1(b). The schematic indicates that the actors participating in air quality decision support system include data mangers, data processing technical analysts, and "informers" who prepare the technical information for the decision makers. These classes of actors are necessary for most air quality decision support systems, including international policy making regarding hemispheric transport of pollutants, regulatory decisions as part of routine air quality management and in DSS for informing the public through real-time data delivery and forecasting.

It is worth highlighting that key users of air quality decision systems are technical analysts and the IS needs to be tailored primarily to their needs. Also, much of the communication along the value chain in the DSS is between the human participants through reports and verbal communication rather than computer-computer interactions.

In summary, the role of DataFed in the GEO enterprise is to provide air quality-related data as services and to participate in the testing of the GEOSS core architecture. DataFed is also a decision support system for air quality management and con-

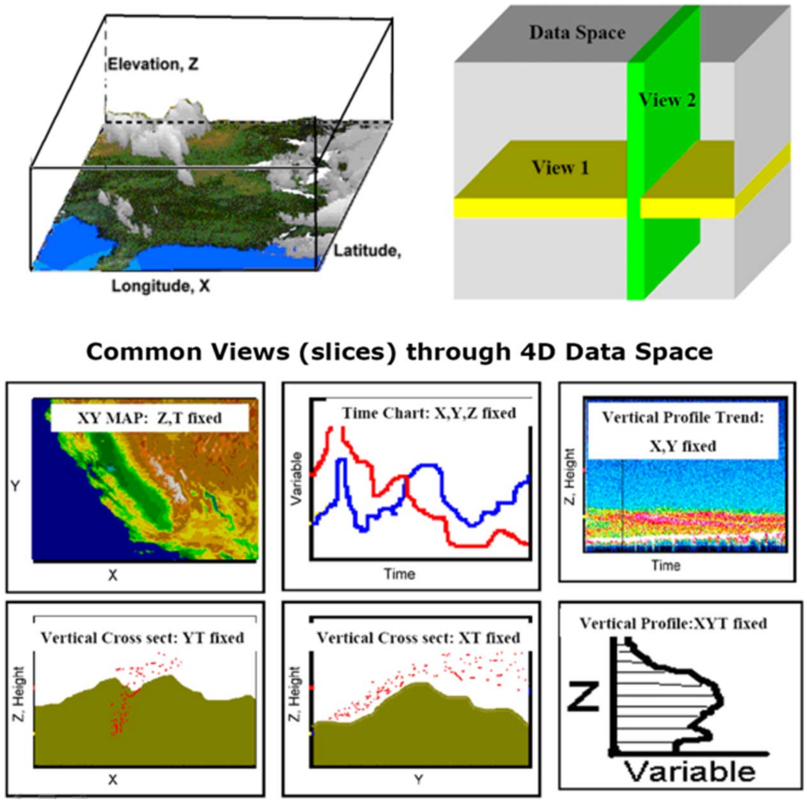



Fig. 2. 4-D data model and typical slices through the 4-D data cube.

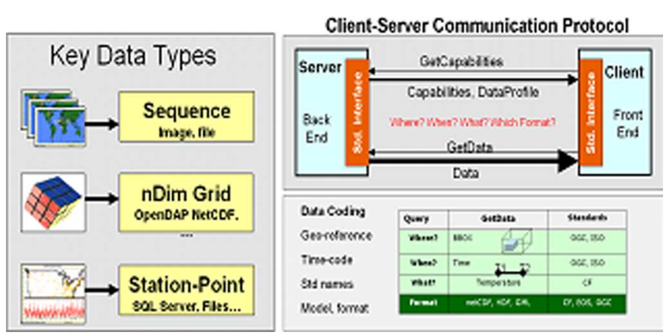



Fig. 3. (a) Key data types: sequential images, multidimensional grids and station-point data. (b) Schematics of OGC standard protocols, WMS, and WCS.

tributes to the refinement of the user requirements in this field. As the GEOSS public infrastructure evolves, DataFed as a DSS, will seek to access and use the GEOSS-mediated resources.

### *B. Information Viewpoint—Earth Observations*

The information viewpoint focuses on the semantics of the information and information processing. The abstract data model used in DataFed is that of a four dimensional data cube (x, y, z, t) (see Fig. 2). The cube dimensions are expressed as the physical dimensions of the Earth system as latitude-longitude, elevation, and time units. The 4-D data model is particularly applicable for representing space-time varying phenomena in Fluid Earth Sciences which include the atmosphere and the oceans.

All the data queries are formulated in terms of the 4-D data cube. Typical queries are slices across different dimensions of the data cube as shown schematically in Fig. 2. This simple data model is consistent with the Open Geospatial Consortium (OGC) standard data access protocols, Web Map Service (WMS), and Web Coverage Service (WCS). For example, the WMS query returns an image representing a slice in the latitude-longitude plane. The WCS query is similar, but it returns numeric data rather than images. WCS is particularly applicable for representing space-time-varying phenomena in Fluid Earth Sciences, atmosphere, and oceans.

The main data types used in DataFed are map images, $n$-dimensional grids and tables representing station point monitoring data [see Fig. 3(a)]. These have direct correspondence with the formally defined data types in the OGC standards [see Fig. 3(a) and (b)] for Web Map Service (WMS), Web Coverage Service (WCS), and Web Feature Service (WFS), respectively. In some cases, it is more convenient to serve point monitoring

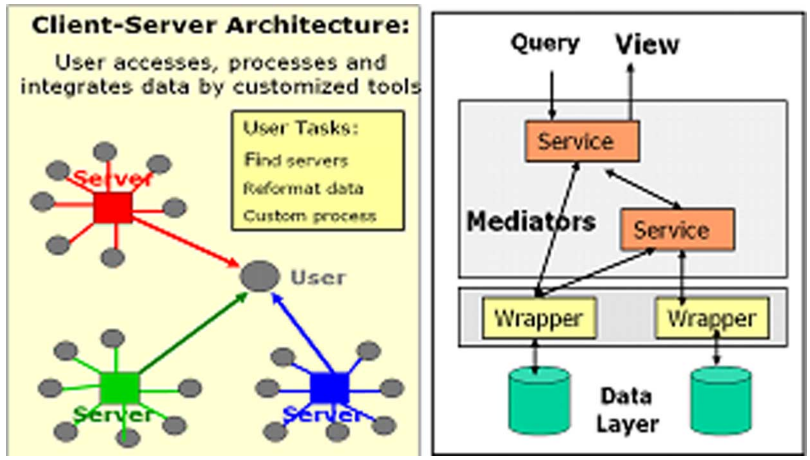



Fig. 4. Schematics of traditional client-server and mediated architectures.

data through the WCS protocol since it allows data access through the powerful spate-time query.

### C. Computational Viewpoint—Service Oriented Interoperability

This viewpoint of the IS architecture describes the functional relationship between the distributed components and their interaction at the interfaces. This viewpoint highlights the key difference between the traditional client-server architecture and the loosely coupled, networked architecture (see Fig. 4).

In the client-server architecture the individual servers are designed and maintained as autonomous systems, each delivering information in its own way. Users who need to access multiple servers carry the burden of finding the servers, formulating the data access procedures, repacking the query results, and performing the necessary integration and homogenization of results. The chores of homogenizing the distributed, heterogeneous datasets are accomplished by wrappers and mediators.

*Wrappers* provide a uniform interface to heterogeneous data by compensating for physical access and syntactic differences. Each wrapper has two sides, one facing the heterogeneous data source that requires custom programming. Data wrappers incorporate the physical server location, perform the space-time subsetting services, execute format translations etc. The other side of the wrapper faces outward toward the internet cloud and presents the uniform interface to the heterogeneous data, i.e., turning data into machine-consumable services.

The wrapper can be physically located on the same server as the data source. However, in a networked environment the wrapping process can be performed as a service by a third party. This is non-intrusive, third party wrapping approach was taken by DataFed for most of the data lacking a standard interface. Initially, our intention was to use the wrappers only for legacy data systems that cannot be equipped with WMS/WCS interfaces. Experience over the past four years has shown that the placement of lightweight wrapper and adopter components between network nodes is desirable for all network links, not only for legacy connections. They allow non-intrusive modification of service connections in response to environmental changes, e.g., an update of an interface standard. 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.

*Mediators* are aggregate software components that accept user queries and return data views. Mediator queries are analogous to the view queries in relational data systems. In fact, the database theoretician Ullman [8], argues that the main role of mediators is to allow such view-based queries against a heterogeneous data system. Data integration through mediators has been an active research area since its inception by Wiederhold [7] in 1992. It provides a flexible way of connecting diverse systems which may include legacy databases. The emergence of SOA has provided an additional impetus and technologies for data federation using mediators. Standardization of service interfaces now allows easy creation of mediators using workflow software.

In DataFed, the orchestration of processing services is performed by a custom-designed workflow engine using SOAP/WSDL web service interfaces. The workflow is designed for chaining both DataFed services as well as other, external web services. Likewise, DataFed's services are available to, and have been integrated with, other organization's workflow software.

The SOA of DataFed is used to build data views 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 browsers for spatial-temporal exploration, multi-view consoles, animators, multi-layer overlays, etc. Fig. 4 illustrates a map view consisting of four independent data layers. The view shows the intrusion of forest fire smoke from Quebec to the N.E. United States. The color NASA satellite image is accessed through an OGC WMS data access service. The point monitoring data are accessed from an SQL server through a wrapper, which formulates the SQL queries, based on the geographic bounding box, time range, and parameter selection in the OGC WCS query.

A data view is a user-specified representation of data accessible through DataFed. Data views consist of a stack of data layers, similar to the stack of spatial GIS data except that DataFed views can represent temporal and other dimensional

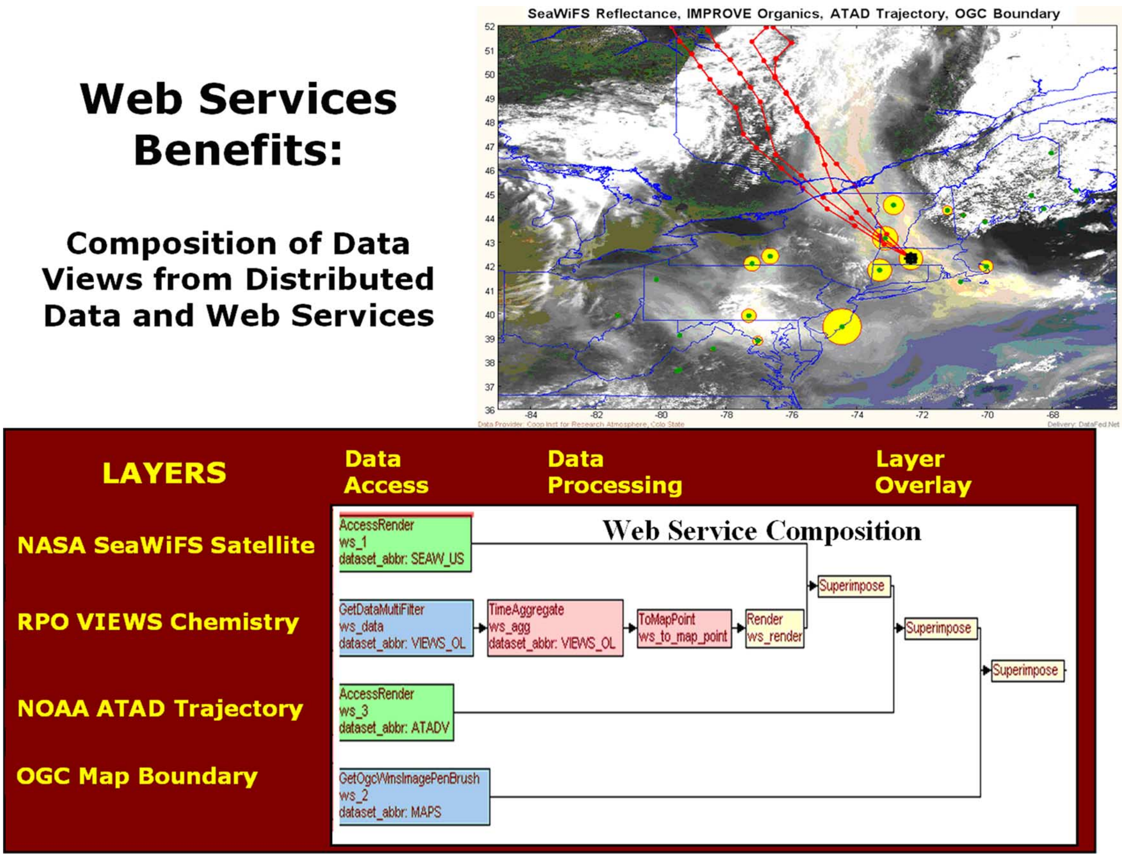



Fig. 5. Typical web service program for the creation of multi-layer data views.

pattern. Each data layer is created by chaining a set of web services, typically consisting of a DataAccessService which is followed by the services for processing, portrayal etc. Data views are defined by an XML file which contains the instructions to create a data view. The view file is also used to store the state, i.e., the input settings of the view. Thus, given a valid view file, the DataFed workflow engine can execute the set of web services either the SOAP or the HTTP Get (REST) protocol. The workflow engine for the orchestration of web services is unique in the sense that the service flows generate data views that can be controlled and embedded directly into application software.

DataFed has its own catalog where data can be registered for standards-based access for processing, visualization, and exploration. DataFed has been registered as a catalog-serving component in the GEOSS registry. For this reason, the datasets are made available for harvesting by the GEOSS Clearinghouse. The interface to the GEOSS Core Architecture is yet to be implemented.

The SoS approach also places new demands on governance, in particular on determining the responsibilities of the participating component systems. In the case of DataFed, the responsibility for providing data lies with the data providers/custodians. The responsibility for the wrappers and mediators lies with the DataFed community. Data discovery is through the data/service registries while the application programs are in the purview of the end user. Formal mechanisms for governance (e.g., service contracts) for such loosely coupled end-to-end applications are not yet developed.

### D. Engineering Viewpoint—Interaction Between Components

This viewpoint focuses on the mechanisms required to support interaction between distributed components. It involves the identification of key components and their interactions through standard communication and data transfer protocols. In DataFed, we have adopted the OGC WMS and WCS protocols as the "convergence" protocols [2] for the standards-based access for all datasets.

OGC WCS is particularly applicable for representing space-time-varying phenomena in Fluid Earth Sciences, atmosphere and oceans. OGC WCS version 1.1 is limited to grids, or "simple" coverages, with homogeneous range sets but future revisions of the standard are anticipated to include support a broader set of coverages, including point coverages. An attractive feature of these services is that: 1) they can be executed using the simple, universal HTTP GET/POST Internet protocol and 2) the services are described by formal XML documents ("GetCapabilities," "DescribeCoverage") and the output formats can be advertised in those service documents.

### E. Technology Viewpoint—Component Instances

This viewpoint identifies the specific key technologies implemented in the system. The key DataFed technologies are data wrappers, web services and service orchestration as discussed above. Additional technologies are applied for building web-based applications fro the users.

A further data federation service is data caching, i.e., local storage or pre-calculation of frequently requested queries. In DataFed the caching consists of densely packed numeric "data cubes" suitable for fast and efficient queries for spatial and temporal views. The data cubes are updated hourly, daily or intermittently as dictated by the data availability and user-need. DataFed also implements image caching. For frequently used views, the data are pre-rendered and stored as images for fast retrieval and interactive browsing.

Since data views are themselves dynamically generated images, these can be embedded into web pages and their content controlled through standard controllers and through Javascript. By design, such light-weight web application pages can be designed and hosted on the user's server. The only relationship to DataFed is that the view images are produced dynamically by the DataFed workflow.

The entire DataFed system was developed and operating in the Microsoft. NET environment and the production server consists of a pair of Dell redundant servers.

### F. User Viewpoint—Integrated User Tools, Workspaces

Beyond the RM-ODP viewpoints, this section highlights DataFed functionality from the user perspective. DataFed is an integrated exploration and analysis system for advanced data analysts. It was designed and implemented by data analysts, themselves users of the system. The key user tools include the following.

*Data Registry* facilitates registering data in the federation and consists of a user-completed form that includes fixed entries required by the Catalog as well as parameters needed data access. For several key data sources (e.g., SQL Servers, sequential images, netCDF files), reusable wrapper templates were developed. There are multiple wrapper classes that produce different data views, i.e., map view and time view. Depending on the data source, data wrappers also require custom programming through Python scripts. *Catalog* is used for finding federated data. Each catalog record describes the dataset and the services to access to data.

*Data Browser and Workflow Editor* is the primary tool for the exploration of spatial-temporal pattern of pollutants. The multi-dimensional data are sliced and displayed in spatial views (maps) and in temporal views (time series). Each data view also accepts user input for point and click navigation in the data space. The DataFed browser is also an editor for data processing workflows using a dedicated SOAP-based workflow engine. A typical workflow for map view is shown in Fig. 4. *Consoles or Dashboards* are for displaying the state of the atmosphere through a collection of data views data from a variety of disparate providers where the sampling time and spatial subset (zoom rectangle) are synchronized. *Google Earth Data Browser*, is a software mashup between DataFed and Google Earth. The two applications are dynamically linked and the user can select and browse the spatial views of any federated dataset. The Google Earth user interface is particularly suitable for the overlay and display of overlapping, multi-sensory data. The temporal animation of sequential data in Google Earth is also instructive for the visualization of air pollutant dynamics and transport.

*DataSpaces* are hybrid structured and unstructured wiki pages are wiki pages dedicated to registered datasets. These DataSpaces are designed to harvest user-contributed metadata and connect data providers, mediators and users. These workspaces includes structured dataset descriptions, data lineage, data quality, and user feedback as well as other user-submitted content relevant to a dataset.

## IV. Interoperability Demonstrations and Pilot Studies

Interoperability among the components services of GEOSS requires continuous development and testing. Interoperability demonstrations and pilot studies are an integral part of the GOESS development process and these require the active participation of its autonomous members. DataFed has participated in these developments of the SoS as a "dual citizen"; it has its own internal and custom architectural elements to meet its particular objectives but at the same time implements standard architectural elements to allow connecting and sharing with other systems. Beyond coexistence, it strives to cooperate, co-evolve and ultimately merge with other data federations. To that end, the developers of DataFed have actively participated in a wide range of interoperability studies.

GALEON (Geo-interface to Atmosphere, Land, Earth, Ocean netCDF) is an OGC Interoperability Experiment[3] to support open access to atmospheric and oceanographic modeling and simulation outputs. This is an active and productive group working on the nuts and bolts of ES data modeling and interoperability. The GEOSS Services Network (GSN) is a persistent network of a publicly accessible OGC services for demonstration and research regarding interoperability arrangements in GEOSS. GSN is the basis for demonstrations in the GEOSS Workshop series.[4] The DataFed group, has actively participated in the Beijing and Denver workshops and organized the interoperability experiment for the Barcelona workshop [10].

The ESIP Air_Quality_Cluster[5] is an activity within the Federation of Earth Science Information Partners, ESIP.[6] It connects air quality data consumers with the providers of those data. The AQ Cluster aims to: 1) bring people and ideas together; 2) facilitates the flow of earth science data to air quality management; and 3) provide a forum for individual AQ projects. The DataFed group is active in the evolution of the OGC WCS specification to air quality data. A specific goal is to include into OGC WCS point coverages arising from surface-based monitoring networks.

[3]OGC GEOSS services network (GSN); [Online]. Available: http://www.ogcnetwork.net/node/56.

[4]GEOSS Services Network (GSN) now the GEOSS Web Services (GWS) is a persistent network of publicly accessible OpenGIS-accessible services for demonstration and research regarding interoperability arrangements in GEOSS.

[5]The ESIP Air Quality Cluster formed to connect AQ data providers and users on current AQ projects.

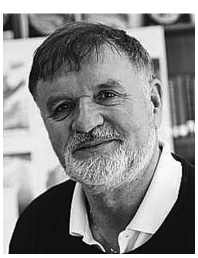
[6]The Federation of Earth Science Information Partners (ESIP) is a unique consortium of more than 90 organizations that collect, interpret and develop applications for remotely sensed Earth observation information.

## ACKNOWLEDGMENT

The authors would like to thank Dr. J. D. Husar for her contributions.

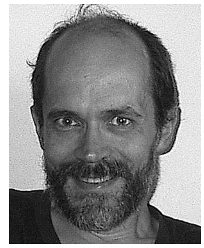
## REFERENCES

[1] ESA Publications Division, The Netherlands, "Global Earth observation system of systems 10-year implementation plan," 2005.
[2] United States Air Force Scientific Advisory Board, Washington, DC, "Report on system-of-systems engineering for air force capability development," SAB-TR-05-04, 2005.
[3] R. Husar and R. Poirot, "DataFed and FASTNET: Tools for agile air quality analysis," *EM, Air Waste Management Assoc.*, pp. 39–41, Sep. 2005.
[4] R. B. Husar, S. R. Falke, and K. Hoijarvi, "Interoperability of web service-based data access and processing: Experience using the DataFed system," presented at the ESTO Meeting, 2006, Paper A6P2.
[5] R. B. Husar and K. Hoijarvi, "DataFed: Mediated web services for distributed air quality data access and processing," in *Proc. IEEE Int. Geosci. Remote Sensing Symp.*, Jul. 2007, pp. 4016–4020.
[6] C. B. Stabell and O. D. Fjeldstad, "Configuring value for competitive advantage: On chains, shops, and networks," *Strategic Management J.*, vol. 19, no. 5, pp. 413–437, 1998.
[7] G. Wiederhold, "Mediators in the architecture of future information systems," *Computer*, vol. 25, no. 3, pp. 38–49, 1992.
[8] J. D. Ullman*, A First Course in Database Systems*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[9] OGC, "The user and the GEOSS architecture V," presented at the IGARSS, Denver, CO, 2006.
[10] OGC, "The User and the GEOSS Architecture XIV," presented at the IGARSS, Barcelona, Spain, 2007 [Online]. Available: http://www.ogcnetwork.net/node/265



**Rudolf B. Husar** studied at University of Zagreb, Zagreb, Croatia, and received the degree of Dipl. Ing. from the Technische Universiat, West Berlin, Germany, and the Ph.D. degree in mechanical engineering from University of Minnesota, Minneapolis, in 1971, followed by a post doctoral research at California Institute of Technology, Pasadena.

He is currently a tenured Professor with the Energy, Environmental, and Chemical Engineering Department and Director of the Center for Air Pollution Impact and Trend Analysis (CAPITA) with Washington University, St. Louis, MO. His research interests span atmospheric chemistry and transport as well as environmental informatics.



**Kari Hoijarvi** received the degree of Dipl. Eng., (equivalent to M.Sc.) from University of Lappeenranta, Lappeenranta, Finland.

He is currently a Research Associate with the CAPITA Research Group, Washington University, St. Louis, MO. His main research interests include software engineering and design.

**Stefan R. Falke** received the D.Sc. in environmental engineering from Washington University, St. Louis, MO, in 1999, and the AAAS Science & Technology Policy Fellowship from the US EPA, Washington, DC.

He is currently an Assistant Research Professor of energy, environmental and chemical engineering with Washington University and Manager of Geospatial Services for Energy and Environment with Northrop Grumman Corporation, St. Louis, MO. His research interests include the development of standards-based web information systems for multi-organizational environmental monitoring, analysis, and forecasting.

**Erin M. Robinson** is currently pursuing the Ph.D. degree in energy, environmental, and chemical engineering from Washington University, St. Louis, MO, under the direction of Prof. Husar in the CAPITA research group.

Her research interests include atmospheric science and environmental informatics.

**George S. Percivall** received the B.S. degree in physics and the M.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana-Champaign, in 1983 and 1984, respectively.

He is currently Chief Architect for the Open Geospatial Consortium and Executive Director of OGC's Interoperability Program, Crofton, MD.