A SCALABLE GRID-ENABLED DATA FRAMEWORK FOR EASTFIRE DECISION SUPPORT

Ruixin Yang, John J. Qu, Zuotao Li, Yuechen Chi, Menas Kafatos
Center for Earth Observing & Space Research (CEOSR)
School of Computational Sciences (SCS)
George Mason University (GMU)
Fairfax, VA 22030
ryang@gmu.edu

Wildfire management in the eastern U.S. is more complex than in the west because of higher population density, increased closeness of housing and people with wildlands, and large spatial and temporal variability of topography, climate, ecosystems, and development patterns. The diversity requires a large group of data sets from heterogeneous sources, such as fuel property and fire characteristics data from *in situ* or remote sensing observations, weather data from observations and model prediction, topography data over long temporal scales but high spatial resolution, socio-economic data, etc. Those data sets are different in spatial and temporal resolutions, data types (data models) and data formats, and are in different distributed sites, accessible in a variety of ways. An integrated data access is needed to support decision making process. Current existing technology in earth science community for data interoperability, and in particular, Grid technology will be valuable for building such an integrated data access system. In this paper, we will identify the data needs and describe a scalable Grid-enabled data framework for wildfire decision support. We will also give overall system architecture and outline the major components.

1. INTRODUCTION

Wildfire is a major natural high risk disaster in the United States. In 2002, for example, there were 88,458 fires that consumed nearly seven million acres with a fire suppression cost for that year estimated at 1.6 billion US dollars (NIFC 2002). In recent years, wildfires have become a significant management and science problem affecting our nation's ecosystems and wildland-urban interfaces (Wangtendonk, Zhu & Lile 2002). They also threaten human life and property and have significant ecological, social, and economic implications (Riebau and Qu 2004; Prestemon *et al.* 2002). The wildfire prevention has been a focal point addressed in the Present's Healthy Forest Initiative (White House 2002).

In order to reduce the aggregate costs and damages arising from wildfires and to support the decision making processes, a great number of studies have been conducted to assess wildfire dangers including a slew of fire indices that rely on some form of fuel conditions and atmospheric inputs (Potter, Goodrick & Brown 2003). Those input data sets are very diverse, in all aspects including data types, formats, sources, and data delivery mechanism. In this paper, we will identify the data needs and describe a scalable Grid-enabled data framework for wildfire decision support. We will also give overall system architecture and outline the major components.

2. GRID-ENABLED DATA FRAMEWORK AND ARCHITECTURE

It is a great challenge to give an integrated risk assessment based on a large variety of wildfire danger risk parameters derived from highly heterogeneous data sources: data from a variety of advanced remote sensing instruments, data from numerical weather predictions, ground measurements as well as human activity related socio-economic data. Some data sources, such as remote sensing instruments and model outputs produce large volumes of highly dynamic data. For example, each file of NCEP (National Centers for Environmental Prediction) ETA model is of about 100MB in size for hourly forecasting, which is updated every 6 hours. Similarly, each granule of MODIS DB (Direct Broadcast) data is of about 380MB, and we have about 10 granules for a specific location each day.

Table 1. High level input data sets to this project.

Data Category	Spatial Resolution	Temporal Resolution	Data Type	Data Format	Data Source	Delivery Mechanism
Remote Sensing	250m-1km	12hours	Swath	HDF	MODIS	automatic ftp
Weather Observation	Various	20minutes	station	coded	Unidata	LDM
Weather Prediction	12km	1 hour	gridded	GRIB	NCEP	GDS/http
Topography	30m	static/years	Raster	GIS	USGS	ftp
Socio-economic	various	static/years	vector	GIS	various	ftp
Fire Products	various	various	various	various	USDA/FS	ftp
Fire Products	various	various	point	ASCII	NOAA/SSD	ftp

Table 1 gives the high level category list of relevant data. The highly heterogeneous data, multi-scale and multi-resolution requirements and current status of data availability pose several challenging IT and related issues. To solve those issues, one needs a strategy for the efficient and economical integration of heterogeneous data and for the development of a scale-adaptive risk assessment framework and a system for supporting collaborative and interactive decision process and knowledge and data dissemination to fire managers and other stakeholders.

Our strategy is to leverage current and future technology such as Unidata Local Data Manager (LDM) (Unidata 2004), NOAA Operational Model Archive and Distribution System (NOMADS) (Rutledge *et al*. 2001), open source software, as well as those developed at George Mason University (GMU).

In the past few years, several earth science information systems were built at the Center for Earth Observing and Space Research (CEOSR) of GMU. The systems include Virtual Domain Application Data Center (VDADC) (Kafatos *et al*. 1997; Yang, Li & Kafatos 1998) which integrate free available software for online data analysis and visualization; the Seasonal to Interannual Earth Science Information Partner (SIESIP) online data search and analysis system, which integrates a Database Management System, Web technology, and GrADS, a data analysis and visualization software package (Kafatos *et al*. 1998); an XML-Based Distributed Metadata Server (DIMES) handling flexible metadata and knowledge about data (Yang *et al*. 2001; Yang, Kafatos & Wang 2002); and an enhanced server which closely integrates an online data server and a metadata server to provide data search, data browsing, data access, and data analysis seamlessly to online data users without downloading any data (Yang *et al*. 2003).

Recently, Grid technology has been receiving more attention in applications for intensive computing and massive data storage (Foster & Kesselman 1999; Foster & Kesselman 2003; Johnson *et al*. 2000). CEOSR has developed an Earth science data grid system based on the Storage Resource Broker (SRB) and metadata catalog service (MCAT) technologies (Rajasekar, Wan & Moore 2002). The CEOSR data grid node is part of the ESIP (Earth Science Information Partner) Federation Data GRID. The system architecture of GMU Earth science data grid is shown in Figure 1, in which components supported at CEOSR are highlighted by color. The SRB provides seamless access to data stored remotely, uniform access to data in heterogeneous storages, and intelligent discovery service to associated metadata. The MCAT based on free PostgreSQL (PostgreSQL 2004) handles the metadata and other resources managed through SRB.

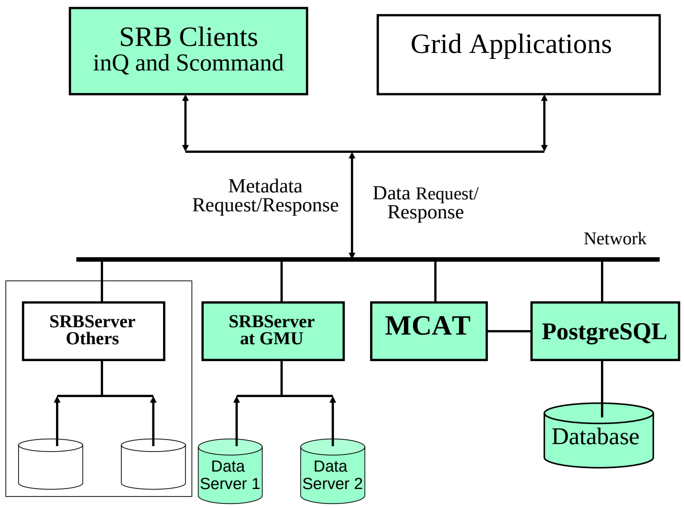

Figure 1. System Architecture of GMU Earth Science Data Grid.

Our expertise in building online information systems, current national distributed information systems, existing Grid technology, and future emerging technology suggest us to propose a scalable Grid-enabled data framework for EastFIRE decision support. Figure 2 demonstrates the functionalities needed for fire danger risk assessment. The highly heterogeneous data including those from ground weather station, environmental monitoring, remote sensing instruments, model outputs and socioeconomic data should be collected and integrated seamlessly. The integrated data should be consumed by the algorithms for risk assessment to translate these raw data into information-rich fire danger risk parameters. These risk parameters along with raw data and accumulated wildfire knowledge base will produce informative risk assessments for fire managers, decision makers and risk responders. Scenario based interactive risk analysis will also be supported by this framework.

Based on above framework, a system for integrated wildfire risk assessments can be developed. Figure 3 sketches the system architecture and high level data flows. The distributed real time or near real time heterogeneous data will be fed into the system via the most efficient and economical mechanism. For example, real time or near real time remote sensing data may come from direct satellite antenna data downlink or through automatic ftp from national remote sensing data centers. The relatively static socio-economic data and historical fire data may be obtained through ftp. The dynamic weather observation data will be fed into the system by using the Unidata LDM mechanism. Weather forecasting data from NWP models will be pulled through data servers with server-side data manipulation capability.

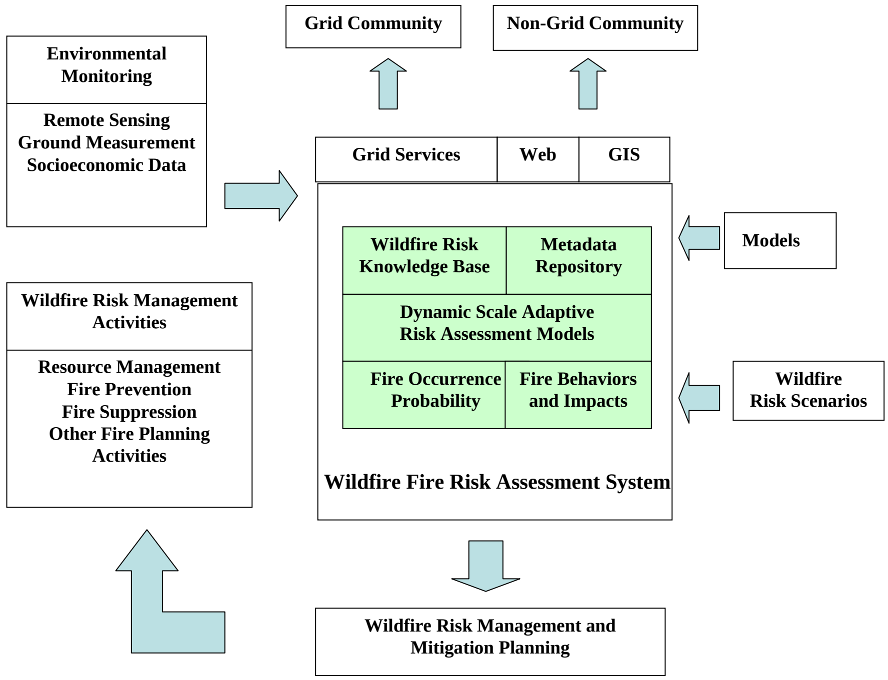

Figure 2. Integrated Wildfire Risk Assessments Framework.

The system will process the integrated data through data grid infrastructure to produce fire danger risk and fire impact assessment. The assessment results and integrated input data will be stored on the data grid and be available to users with grid-enabled client or more traditional tools such as web browsers and GIS. Therefore, the system can be a basis of future deployed information technology infrastructure for fire science and fire management communities.

3. CONCLUSION

The existing data resources, data delivery mechanisms, and IT technology including Grid make it possible to develop a scalable Grid-enabled data framework for EastFIRE and the corresponding decision support system by integrating existing components. With this system, firefighting coordinators and other decision makers will be able to conduct scenario based interactive danger risk analysis and risk maps will be produced through the analysis process. The system will allow fire managers and the general public to access real time online maps of fire danger risk near current wildfire locations using standard web browsers. Fire danger risk data and maps can be updated several times a day based on the input from incident intelligence sources, such as satellite images and meteorological data and field observations, during fire reasons. In addition, this system also allows users to selectively zoom in and out at various scales of spatial regions of interest to display real time fire detail danger risk patterns. This prototype also allows users to access other online important information around these regions, such as schools, houses, highways, hospitals, etc.

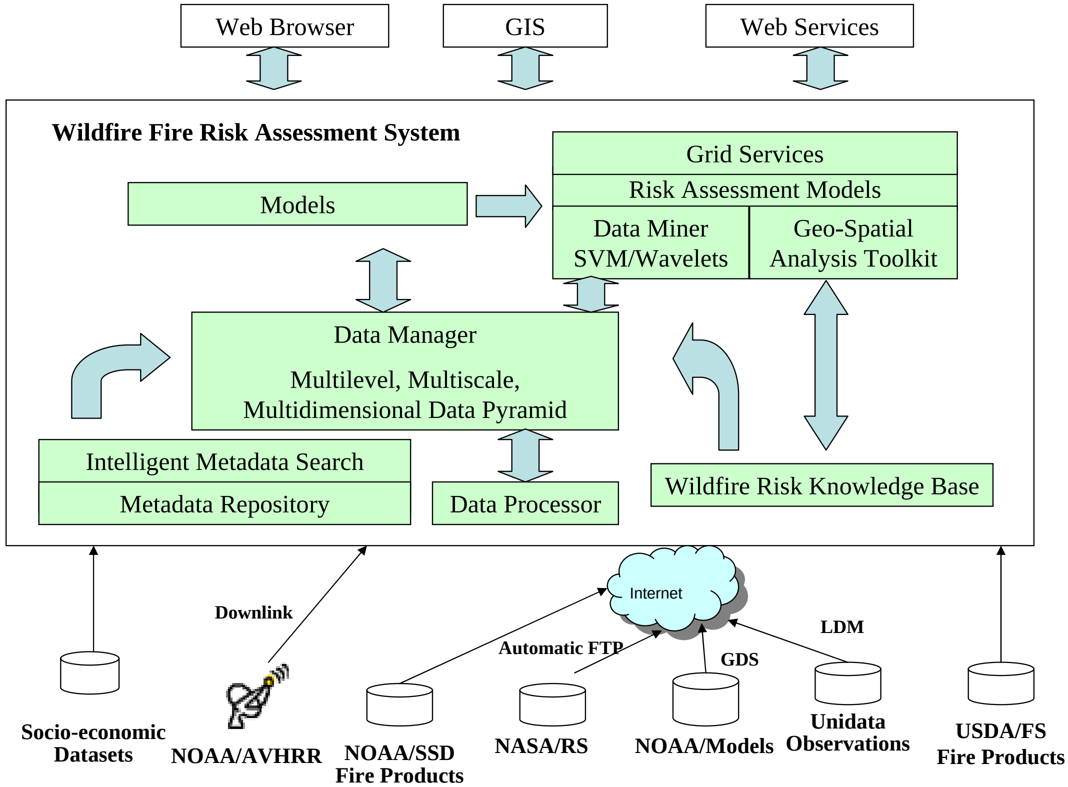

Figure 3. System Architecture Based on the Proposed Framework.

REFERENCE

1. Foster, I. and C. Kesselman, The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann Publishers, Inc. San Francisco: CA. 1999.
2. Foster, I. and C. Kesselman , 2003, “ The Grid 2: Blueprint for a New Computing Infrastructure,” Morgan Kaufmann Publishers, Inc. San Francisco: CA..
3. Johnson, B., D. Gannon, B. Nitzberg , L.A. Tanner , B. Thigpen, and A. Woo, 2000, “Computing and Data Grids for Science and Engineering,” IEEE SC’2000, Dallas, Texas. http://www.sc2000.org/techpapr/papers/pap.pap253.pdf.
4. Kafatos, M., Z. Li, R. Wang, X. Wang, P. Hertz, D. King, H. Wolf, and D. Ziskin, 1997, “The Virtual Domain Application Data Center: Serving Interdisciplinary Earth Scientists,” in *Proceedings of the Ninth International Conference on Scientific and Statistical Database Management*, pp. 264-276. IEEE, Computer Society.
5. Kafatos, M., X.S. Wang,, Z. Li,, R. Yang, and D. Ziskin,, 1998, “Information Technology Implementation for a Distributed Data System Serving Earth Scientists: Seasonal to Interannual ESIP,” in *Proceedings of the 10th International Conference on Scientific and Statistical Database Management*, pp. 210-215. IEEE, Computer Society.
6. NIFC, National Interagency Fire Center, 2002, “Wildland Fire Statistics,” www.nifc.gov/stats/wildlandfirestats.html, last accessed on February 19, 2004.
7. PostgreSQL, 2004, http://www.postgresql.org/, last accessed on February 19, 2004.
8. Potter, Brian E., S. L. Goodrick and T. J. Brown, 2003, “Development of a statistical validation methodology for fire weather indices,” 2nd International Wildland Fire Ecology

and Fire Management Congress, Paper J11.1, American Meteorological Society**,** November 2003, Orlando, FL. http://ams.confex.com/ams/pdfpapers/65540.pdf. last accessed on February 23, 2004.

9. Prestemon, P.J. Pye, J.M., Butry, D.T. Holmes, T.P. and Mercer, D.E, 2002, “Understanding Broadscale Wildfire Risks in a Human-Domainated Landscape,” *Forest Science*, Vo1. 48, No.4.
10. Rajasekar, A., M. Wan, and R. Moore, "*MySRB & SRB -- Components of a Data Grid*", In The 11 th International Symposium on High Performance Distributed Computing (HPDC-11. Edinburgh, Scotland, 2002.
11. Riebau, A. and J. J. Qu, 2004, Application of Remote Sensing and GIS for Analysis of Forest Fire Risk and Assessment of Forest Degradations in the Southwest Pacific Region, Proceeding of Expert Group Meeting on Reducing the Impact of Natural Disasters and Mitigation of Extreme Events in Agriculture, Rangelands, Forestry and Fisheries,Beijing, China,16 - 20 February 2004.
12. Rutledge, G. K., T. Karl, D. Easterling, L. Buja, R. Stouffer, 2001, “The NOAA Operational Model Archive and Distribution System (NOMADS): Bridging the Model/Observations Gap,” (Invited), American Geophysical Union, (AGU) Boston MA, May 2001.
13. Unidata, 2004, “The Unidata LDM,” http://my.unidata.ucar.edu/content/software/ldm/index.html, last accessed on February 19, 2004.
14. Wangtendonk, J. W., Z Zhu and E. Lile, 2002, “Pre-fire Risk Assessment and Fuels Mapping,” White Paper of November 2002, USGS Third Wildland Fire Science Workshop. November 12-15, 2002, Denver, Colorado.
15. White House, 2002, “White House Press Release: Present’s Health Forest Initiative” http://www.whitehouse.gov/news/releases/2002/08/20020822-3.html, last accessed on February 22, 2004.
16. Yang, R., Li, Z. and Kafatos, M., 1998, “Remote Sensing Data Access via WWW-Enabled Search and Analysis Engines,” in *Proceedings of The 4th Pacific Ocean Remote Sensing Conference*, pp. 509-513, Qingdao, China, July 28-31, 1998
17. Yang, R., M. Kafatos and X. Wang, 2002, “Managing Scientific Metadata Using XML,” *IEEE Internet Computing*, v6, no.4, pp. 52-59.
18. Yang, R., M. Kafatos, B. Doty, J. Kinter and L. Pham, 2003, “A Distributed Enhanced Server for Multi-Dimensional Scientific Data Search, Access and Analysis,” *Computing in Science & Engineering*, v5, no.2 pp. 44-52.
19. Yang, R., X. Deng, M. Kafatos, C. Wang and X. S. Wang, 2001, “An XML-Based Distributed Metadata Server (DIMES) Supporting Earth Science Metadata,” in *Proceedings of the 13th International Conference on Scientific and Statistical Database Management*, pp. 251-256, IEEE, Computer Society.