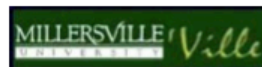




L I N K E D  
E N V I R O N M E N T S  
F O R A T M O S P H E R I C  
D I S C O V E R Y



# Final Report

## October 2009

NSF Cooperative Agreements ATM03-31574, 31578, 31579, 31480, 31586, 31587, 31591, 31594

## 1. Introduction and High Level Summary

High-impact local weather occurs daily across a breadth of application areas -- aviation, surface transportation, harvesting and application of pesticides and herbicides in agriculture, building construction, massive public recreation venues such as Walt Disney World – and are characterized by one common factor: they share an inability to apply weather technologies when, where and how needed in order to deal with high-impact weather.

Today's weather technology in general, and its use in research, operations and education, even as 2009 comes to a close remains *far from optimal when applied to any particular situation involving high-impact local weather*. In recognition of the need for fundamental research in this area, in October 2003 the National Science Foundation funded a 5-year Large Information Technology Research (ITR) grant known as Linked Environments for Atmospheric Discovery (LEAD). A multidisciplinary effort, involving nine institutions and more than 100 scientists, students and technical staff, LEAD subsequently created an integrated, scalable, web service architecture (SOA) in which meteorological analysis tools, forecast models and data repositories can operate as dynamically adaptive, on-demand, grid-enabled systems that (i) change configuration rapidly and automatically in response to weather, (ii) respond dynamically to inputs from users, (iii) initiate other processes automatically, and (iv) steer remote observing technologies to optimize the data collection for the problem at hand. Specifically, LEAD focused efforts on:

- Developing capabilities to allow models and other atmospheric tools to *respond dynamically* to their own output, to observations, and to user inputs so as to operate as effectively as possible in any given situation;
- Developing appropriate *adaptive capabilities* within supporting IT infrastructure.

Achieving these goals required more than translating existing capabilities (e.g., batch processing of numerical models configured using name list input files) into new IT frameworks built upon symbolic workflow task graphs, web services and grid computing. Rather, it required fundamental changes – underpinned by basic research in meteorology and computer science – in how experiments are conceived and performed, in the structure of user application tools and middleware, and in methodologies used to observe the atmosphere. *The stretch goal for LEAD was to begin ushering in this paradigm change*. After 6 years, LEAD produced 421 papers and presentations; 9 PhD degrees, 14 Masters degrees, and 1 Bachelors degree were awarded directly related to the project. LEAD engaged the meteorology education and research communities through workshops, and through making the LEAD framework available in existing venues such as the NOAA NCAS Weather Camp at Howard University or the NOAA Hazardous Weather Testbed (HWT) Spring Experiment. LEAD had and still maintains a strong multi-year presence at the AMS yearly meeting and the TeraGrid conference. Bill Gates mentioned LEAD in his 2005 keynote speech at Supercomputing 2005. LEAD tools and services have been and continue to be used in other projects, and continue to be enhanced with funding from follow-on sources. 5 TB of user generated forecast and analysis has resulted from research and education use of the LEAD Science Gateway. It is safe to say based on the kinds of activities we highlight here and in the body of the report that LEAD has succeeded in making substantial inroads towards advancing a paradigm change in how the meteorology community envisions responses to local high impact weather.

The impact of a large, interdisciplinary project such as LEAD is impossible to assess purely by quantitative metrics. In this final project report we attempt to capture LEAD's impact through quantitative measures, but also through qualitative examples and highlights of research and tool

contributions to the meteorology, computer science, and Teragrid communities, and to communities beyond.

Assessed through scholarly publications and presentations, the impact of LEAD is substantial as is shown in Table 1, where the number of publications over the grant life was 254 and the presentations reached 167. It is interesting to note the publication output in Year 1 of the project. Meteorology research in Year 1 yielded 9 publications where computer science research yielded no publications. While there are many possible explanations, it does challenge a commonly held belief in collaborative interdisciplinary research involving computer scientists that the domain science research can only begin once software and hardware infrastructure is in place. The fact that CS had no publications in Year 1 could be explained by the nature of computer science work in interdisciplinary settings where computer scientists spend the first year building software. Patterson et al. backs this up (Patterson et al. 1999) by pointing out that systems research often has a slower publication rate because the time to develop prototype software can be long. It should also be noted that 12.5% of the presentations in Year 2 are in the educational track.

Publications and Presentations/	2004	2005	2006	2007	2008	2009 (NCE)	Total by category
General	1	3	3	2	2	0	<b>11</b>
Meteorology	9	14	28	28	12	3	<b>94</b>
Computer science	0	29	38	20	33	7	<b>127</b>
EOT	0	4	4	2	3	0	<b>13</b>
Social Science	0	3	3	3	0	0	<b>9</b>
Presentations	16	40	40	40	11	20	<b>167</b>
<b>Total by Year</b>	<b>26</b>	<b>93</b>	<b>116</b>	<b>95</b>	<b>61</b>	<b>30</b>	<b>421</b>

Table 1. Summary of scholarly production measured by publications broken out by discipline, and presentations.

The qualitative impact of the project can be seen in part anecdotally. As an example, Howard University master’s student Hector Lima used the LEAD Portal to combine data and results from WRF simulations, which enabled him to study the development of gravity waves events within the vicinity of the Appalachian Mountains, and was able to investigate limitations of the model to forecast these events. This success was aided by a cyberinfrastructure system that reduced the time for Mr. Lima to obtain the data and model results, so his efforts could be focused on the science and not on learning cryptic software scripts and code. We saw this repeatedly, most often for researchers interested in using WRF in non-meteorology research. For these researchers the level of model configurability provided through the LEAD workflow configuration tools was sufficient. Hard core meteorologists, on the other hand, were better served by the WRF Domain Wizard adaptation that Craig Mattocks (UNC) did in collaboration with the WRF Portal team under supplemental funding in the 5<sup>th</sup> year of the project.

Additionally, LEAD pioneered research in workflow systems, and was one of the first systems to run real, complex long running workflows. As such it gained recognition and continues to inspire research in e-Science as the following two recent inquiries attest to:

*"I am a PhD student at the University of Basle, Switzerland, working in the field of Scientific Workflows. For the preparation of experiments, I was looking for real-life workflows. It seems to me that the LEAD applications are pretty "spot-on" to serve as example, motivation, and also use-cases for my work. In fact, I have already used one workflow in an earlier publication as motivation [URL included]." Dec 2009.*

*"I have been investigating the LEAD project for the past week or so and I'm interested in obtaining large sample workflows (i.e., 20-30 services) used in the project. I'd like to use them as motivation for my research in Real-time SOA." Oct 2009.*

LEAD struck a delicate balance during the 6 years of the project (5 plus a no-cost extension year) between carrying out leading-edge research across meteorology and e-Science, and providing a computational platform upon which users could get real work done. This choice was heavily influenced by the research nature of the NSF ITR program that funded LEAD as a research project and not, for instance, as an engineering development effort. In many ways this choice was a success on both fronts because the project advanced understanding of Science Gateways, ensemble forecasting, service oriented architectures (SOA), data integration, mining and metadata, adaptive and responsive algorithms for data assimilation. But there are also drawbacks to this balanced approach, the biggest of which had an impact on the realization of the LEAD sustainability plan. While a sustainability objective is now being achieved, LEAD like many other large NSF funded cyberinfrastructure efforts, faced a highly uncertain future as the project ended.

The sustainability of LEAD has taken shape in part in the form of LEAD-in-a-Box, the next generation LEAD, led by Indiana University in a partnership with Microsoft Research that incorporates Microsoft technology in key cyberinfrastructure pieces. In particular it has adopted the Trident Scientific Workflow Workbench, Windows Workflow Foundation, and Windows HPC Server, reducing substantially the maintenance cost and increasing reliability of the system. An early concept of LEAD-in-a-Box was successfully demonstrated at Supercomputing in Portland OR November 2009. LEAD also attracted the attention of the FAA and is engaged in the NextGen air traffic control effort.

In the following sections we highlight the particularly high impact research, outreach, and education resulting from the project.

## **2. Research and Community Impact in Physical Sciences**

A cornerstone of the LEAD mission has been to develop data, modeling, and cyber environments that can operate as dynamically adaptive, grid-enabled systems, and LEAD demonstrated this capability for the first time during the 2007 Spring Experiment of the NOAA Hazardous Weather Test Bed (HWT) in Norman, Oklahoma. This capability was repeated and enhanced in 2008 and 2009. Using substantial resources on the NSF TeraGrid (including PSC, NCSA, the University of Indiana, and NCIS), LEAD-supported weather analysis and prediction components, along with other scripting software at CAPS, we produced, on a daily basis for a 7-week period, a multi-member, 4 km grid spacing, two-thirds Continental-US (CONUS) WRF ensemble, and a single 2 km resolution forecast (1-km in 2009) over the same domain. Beginning in 2008 NEXRAD Level II data from about 100 radars across that domain were analyzed for initializing all the WRF model forecasts. In 2009 4-km members using the ARPS model were added to the ensemble, enhancements were made to the radar wind analysis, and the domain was expanded. Also, in 2007 and 2008 1-km forecasts over a specific targeted area were generated on the LEAD portal based on the HWT forecasters' interests for same-day high resolution forecasts,

thus incorporating a human in the loop.

This unprecedented capability provided operational forecasters and researchers with the ability to compare, in a real time operational setting, fine-scale deterministic and ensemble forecasts generated with and without NEXRAD data in the initial conditions. These forecasts were examined in real time not only by a dozen or more operational forecasters, but also evaluated by numerous senior faculty and researchers who visited Oklahoma during the course of the experiment each spring. Analysis of the impact of the radar data showed a marked improvement in the first few hours of the forecasts, an improvement that lasted for several hours. In individual cases the impact of the radar data can be seen as long as 24-30 hours into the forecast. Although the research outcomes are only beginning to emerge, they suggest that the capabilities offered by LEAD are poised to transform our understanding of and ability to predict mesoscale weather.

Also during the spring periods the LEAD-supported broker system at NCSA spawned more than 1000 limited area forecasts at 2 km grid spacing. These forecasts were triggered automatically in response to tornado watches and mesoscale discussions issued by the National Weather Service, demonstrating automated adaptive forecasts. During the course of the LEAD project, refinements were made to the automated forecast launch system, for example to avoid redundant forecasts when multiple tornado watches are issued for the same region. The NCSA Seige software for spawning multiple forecasts can be employed as a means of simultaneously testing a wide range of tuning parameter settings in a forecast for a particular case and has the potential to dramatically improve productivity by automating such studies.

The LEAD portal was used in 2009 to submit to the TeraGrid once-daily targeted 1-km resolution forecasts covering the daily domain of interest for the VORTEX2 project. The VORTEX2 project has a need to accurately deploy mobile radars and other mobile observing systems in order to observe potentially tornadic supercell thunderstorms. Their operations domain covers a wide expanse of the Great Plains, yet they can only cover a small part of that on a given day. The use of LEAD will be expanded and enhanced for the 2010 VORTEX2 field campaign using the LEAD system ported to a Windows-based HPC environment, LEAD-in-a-Box. We expect that this use will pave the way for future use of the LEAD system to aide scientists in deploying observing systems for field data collection.

The LEAD project developed wind and cloud analysis tools that were also used by the Collaborative Adaptive Sensing of the Atmosphere (CASA) NSF Engineering Research Center. In the spring of 2007 through 2009 this software was used to produce high resolution real-time analyses that were assimilated into 1-km forecasts to test the utility of the low-level CASA radar data. This has proven quite effective; for example the assimilated data and short term forecast produced a remarkably accurate depiction and short-term forecast of the formation and evolution of the 14-May-2009 Anadarko tornado.

Related to adaptive sensing is the issue of adaptive forecasts, and LEAD supported work that examined how forecast configurations can adapt to produce the best forecasts given a limited set of computing resources. The WRF model was used to examine a scenario of dynamically adaptive numerical weather prediction in which, under the constraint of fixed computing time, a single nested grid of variable spacing and size, but fixed location, was placed within a baseline or pseudo-operational forecast run at 5 km grid spacing. The case examined was an idealized, isolated supercell storm with only warm rain microphysics and no terrain, radiation or surface physics. It was found that phase error (dispersion) was the dominant error in all experiments, and the mean square or total error tended to increase for a fixed grid spacing (nesting ratio) as domain size decreased. Not all variables exhibited this behavior, and for some the “best” forecast was not

produced by the largest domain and finest grid spacing owing to the influence of the nesting ratio, which increased as grid spacing decreased.

Also related to the CASA-LEAD interaction, LEAD also supported research in Ensemble Kalman Filter (EnKF) data assimilation for streaming observations at CAPS. In an OSSE study under the perfect model and imperfect storm environment assumptions, a series of experiments with surface observing site separation ranging from 32 km to 2 km were conducted to examine the impact of surface observations on storm analysis. The rms error of the EnKF analysis relative to that of a corresponding radar-data-only case decreased linearly with decreases in the mean surface network spacing until the spacing was close to the truth-simulation grid interval of 2 km. This relative error is approximately proportional to the inverse square root of the number of stations when the network is relatively coarse. Under the LEAD support, an adaptive observation system based on the Ensemble Transform Kalman Filter (ETKF) was developed that aims to determine the flow-dependent optimal scanning strategies of CASA type radars that minimizes the forecast error, as measured by predicted variance of ensemble forecasts. This work continues within CASA.

Dr. K. Baker, a professor of Geography at Western Michigan University, utilized workflows in the LEAD portal to create 5-km resolution forecasts of surface weather conditions in the Great Lakes Region on TeraGrid. These forecasts were used for her research on forecasting potato late blight in order to provide guidance for crop management, including optimal application of fungicides. This use of the LEAD is part a larger effort to produce a working multi-scale, multi-crop, multi-regional crop disease forecasting system based on publically available data sources, NSF funded cyber-infrastructure, and web based information delivery systems. This research use represents a major impact for LEAD in facilitating the use of complex meteorological tools and advanced high performance computing resources to advance science and applications outside of the field of meteorology, and in this case, with significant potential for economic and ecological benefits.

LEAD received supplemental funding from NSF to create linkages between LEAD and the NOAA/ESRL/GSD WRF Portal, principally in the area of namelist and other model configuration parameters. This work, led by Dr. C. Mattocks of UNC, greatly expanded the utility of both systems by setting up a customized WRF Domain Wizard to ease set-up of the WRF Model domain. The WRF Preprocessing System (WPS) was also implemented as a workflow in the LEAD portal to flatten the learning curve for new WRF users. Additionally, namelist upload and edit capabilities were added to the LEAD portal to allow researchers to edit or generate namelist files offsite or control specific model processes interactively. Namelist variables can be checked for errors and validated against known working combinations of parameters, thus reducing user frustration and conserving computing resources by ensuring that simulations submitted to remote job queues don't fail due to easily preventable errors. Changes were made to allow more complex specification of the WRF vertical coordinate, allowing for higher vertical resolution where needed while conserving the total number of model grid points. Also, a new batch capability for generating graphics products from WRF model output is now available in LEAD to complement the interactive visualization provided by the Integrated Data Viewer (IDV). These automated NCAR Command Language (NCL) scripts produce plots of several commonly-examined fields. This quick-look plotting capability gives users an overview of their WRF simulation results. The quality is high enough to use the images in scientific publications, presentations or web pages. All-told, these enhancements add a rich set of tools to the LEAD portal that are easy to use for the beginning WRF user but incorporate features demanded by advanced weather researchers.

Results from these studies and news about the LEAD tools development have been shared in the meteorological journals, as noted in the publication list, and in several conferences, including special LEAD sessions in the 2007 and 2008 annual meetings of the American Meteorological Society.

### **3. Research and Community Impact in Computer Science**

There are five common components to the kind of Grid portal that LEAD developed: 1.) data search and discovery over public and local data collections, 2.) security, 3.) user private data storage that maintains user's analysis results "in the cloud", 4.) tools for designing and conducting computational experiments, and 5.) tracking data provenance. The real-time nature of high-impact local weather motivated computer science research in dynamic data mining and workflow triggering as well. At the time LEAD began, pieces of a solution already existed, such as with the Unidata THREDDS server, the NCSA led myProxy credential management service, IDV visualization tool, ADaM data mining toolkit but in the areas of data discovery, personal data storage, and experiment composition and execution, much was still hypotheses and research ideas. The LEAD portal, which provided a web browser front end to the entire system, was constructed from core technology provided by the NSF NMI Middleware project (specifically, the Open Grid Computing Environments, or OGCE) and the Gridsphere portal container effort. Below we give highlights of each of functional category, but first mention the impact that LEAD had on the NSF funded Teragrid [www.teragrid.org/](http://www.teragrid.org/).

**LEAD and Teragrid.** The execution of analysis tools and forecast models requires powerful compute resources on which to run. While some researchers have access to ample compute resources, resources in academic departments are often limited. Further university computational resources are often batch scheduled so lack responsiveness to the dynamic nature of weather. LEAD approached these problems by utilizing the NSF TeraGrid (specifically Mercury at NCSA and Big Red at Indiana University) to run its computationally intense workflows. LEAD pioneered the Gateway as a simple means by which researchers can use the TeraGrid without having to solve the complex setup required, particularly with respect to security mechanisms and resource allocation grants. Eliminating the need for the individual scientist to write a proposal for compute cycles (an allocation proposal) to justify their work made rapid start-up and smaller scale, broader, more experimental use possible.

LEAD is one of the first successful Teragrid Science Gateways (Wilkins-Diehr et al 2008). Because of the timeliness demands of mesoscale weather, LEAD also pushed the TeraGrid in new directions – towards responsiveness to the immediacy of severe, local weather, and towards the bursty nature of a weather gateway. LEAD collaborated with research efforts in the process including the Virtual Grid Application Development Software Project (VGrADS) and Special PRiority and Urgent Computing Environment (SPRUCE) to improve response time and decrease queue wait times. SPRUCE/LEAD was demoed at Supercomputing 2007, and VGrADS/LEAD demos have been held yearly at Supercomputing.

**Workflow and Portal.** The way in which an experiment is initiated in the LEAD portal is through an Experiment Builder. Based on the idea of a software "wizard", a program which leads the user through the steps for installing or using software, the experiment builder guides the user through the setup and execution of an analysis workflow. Workflows are represented as a combination of dataflow and control flow in the form of a directed graph. Each node of the graph represents an application service with multiple inputs and one or more outputs. The inputs and outputs may be as simple as a numerical or string value, or as complex as a descriptor of a set of files, which must be fetched and staged on the compute host that is running the application. (This

staging is done automatically by the application service instance.) An example is a data interpolator used in data assimilation to produce lateral boundary conditions. The LEAD workflow system is a comprehensive system that utilizes the XBay workflow composition tool, the Apache ODE workflow engine, the GFAC application wrapping service, and additional services that work with Globus GRAM to schedule jobs on Teragrid resources. The workflow suite, originally funded by LEAD and first deployed there, is now a key piece of the Open Grid Computing Environments (OGCE) <http://www.collab-ogce.org> source repository, and has been used as a basis in whole or in part for other Teragrid Science Gateways.

The LEAD Ensemble Broker, also known as the Parameterized Workflow Engine (PWE), was developed with the express goal of managing the full range and complexity of parameter and ensemble research and forecasting. In order to carry out many-membered ensembles and to efficiently manage their execution on a high performance computer, several steps had to be taken. First, simple expressions are employed to define parameter sets for simulations to be carried out. Then, following the glide-in mechanisms pioneered by the Condor group, a scalable glide-in version of the ELF job container was developed that is launched in the context of a single batch job. This avoids asking the resource manager to launch and monitor hundreds to thousands of jobs in a short period of time (something that is not feasible) by separating container and job management. Because the single job is submitted to the resource manager via ssh (rather than Globus GRAM), it is possible to obtain the JobID and the nodes allocated to the job. Further, if an individual simulation failed it could be resubmitted automatically. Finally, a user interface was crafted with Eclipse Rich Client Platform technology that allows for easy workflow editing, launching and monitoring, *including a relatively concise and general way for expressing the parameterization itself*. The workflow system has been tested and used to launch thousands of WRF ensemble members in research studies led by Dr. Brian Jewett at the University of Illinois. It is being adapted to run on Blue Waters to not only support atmospheric science research but to also enable other researchers with different models to carry out large parametric studies in a similar manner.

**Dynamic Mining and Workflow Triggering.** Dynamic forecasting, where forecasts are triggered automatically in response to emerging severe weather, was explored in LEAD using a couple approaches. The first is through data mining model or observational data, as explained here. The second is by watching the stream of weather watches from the National Weather Service, and triggering forecasts based on severe events. This was discussed in Section 2.

The adaptive capability of LEAD was advanced through the development of several new data mining algorithms for feature detection of model generated and observational data (NEXRAD Doppler radar data). The *Phenomena Extraction Algorithm (PEA)* is a novel, effective and efficient method for phenomena detection based on the statistical properties of the dataset and therefore is data, domain and application independent. Unlike thresholding methods, the PEA examines both the intensity and the variation in a region, and is currently being used in other projects outside LEAD. A flexible *Storm Detection Algorithm (SDA)* that works on multiple types of data sets such as the WSR88-D and numerical models outputs was also developed. The algorithm retains the data points with intensities higher than the specified threshold and uses a region-growing technique to build 3-dimensional (3D) volumes. The *Mesocyclone Detection Algorithm* identifies 1-dimensional (1D) shear segments based on Rankine Vortices, detects two-dimensional (2D) shear regions in a scan sweep, and aggregates collocated shear regions into 3D mesocyclone signatures. The *Storm Clustering algorithm* spatially groups the severe weather detections and determines which of these groups are most interesting. This information is used to spawn limited finer resolution models providing the dynamic and adaptive forecasting capability within LEAD. *Coverage Based Ensemble Algorithm (CBEA)* is an incremental anytime learning

algorithm well suited for dealing with streaming data, as in a sensor network, and when there are changing conditions or goals. This is accomplished through the dynamic selection of an ensemble of classifiers – each classifier tuned to excel in a specific area. Taken together, the development of these algorithms both advances the ability of researchers to quickly mine their data for features and to track real-time events. This provides a means to link model execution with sensor data generation in a dynamic feedback environment.

**Data Management and Search.** A significant challenge in LEAD was in bringing together observational and model generated data from public repositories, local research lab collections, and user generated model and analysis derived data products and making that data available on demand and with minimal latency to the workflow system. This data integration problem is known to be an important problem for which good solutions are still being investigated. We addressed this by evaluating existing community solutions to metadata representation based on 1.) conformance to a widely accepted standard and 2.) ability to support data discovery through querying. We studied solutions closest to home, including the THREDDS metadata schema and the Earth System Modeling Framework (ESMF) and settled on a LEAD-specific profile to the Federal Geographic Data Committee (FGDC) standard as it best satisfied both requirements. The LEAD profile is an extension and subset to the full FGDC schema and was structured to handle the LEAD specific notion of collections and extensions to the schema brought about by separate schemas (requiring multiple namespaces.) We established interoperability with existing metadata schemas such as the THREDDS schema by means of cross-walking from one schema type to the other. The utilization of FGDC elements allows LEAD data catalog to easily interoperate with other US government geospatial efforts such as the Geospatial One-Stop and the NSDI Clearinghouse. The schema has been requested by a committee member of the American Meteorological Society (AMS) Ad Hoc Committee on the Nationwide Network of Networks [www.ametsoc.org/boardpages/cwce/docs/NoN/2009-10-Update.pdf](http://www.ametsoc.org/boardpages/cwce/docs/NoN/2009-10-Update.pdf) to serve as an example in their discussions.

Semantic search over public and LEAD local weather sources is explored in the *Noesis semantic search tool* that uses ontologies to associate semantic information with the search process. It provides a guided refinement of search queries producing successful searches and reducing the user's burden to experiment with different search strings. Furthermore, Noesis queries many different public catalogs and indexes of various types and aggregates the resources returned. Noesis utilizes a domain ontology that encodes all the terms in the AMS Glossary as concepts and their constraints and the relationships, and an ontology developed in LEAD for mapping the Climate and Forecasting (CF Convention) controlled vocabulary to the concepts in Atmospheric Science ontology. The CF ontology allowed the Noesis search engine to mediate search queries across different catalogs. Noesis is now being used in several other projects outside of LEAD.

LEAD has made advancements in reproducibility of atmospheric science research outcomes through *provenance collection* with the *Karma provenance collection tool* (Cao et al, 2009, Simmhan et al. 2008). The provenance of a data product is a record of its lineage, or trace of the execution history that resulted in the product. Provenance enables data to be properly attributed and captures critical parameters about the model run so the quality of the result can be ascertained. Proper provenance is essential to providing *reproducible* scientific computing results. The Karma project has gone on to receive funding from NSF OCI SDCI, Eli Lilly Corporation, NASA and the NSF-funded GENI program.

The public data collections used in LEAD were made available through THREDDS servers hosted at Unidata and Indiana University. The private user workspace is supported by XMC Cat (Jensen and Plale 2006). THREDDS, Thematic Realtime Environmental Distributed Data Services, is middleware that allows students, educators and researchers to publish, contribute,

find, and interact with data relating to the Earth system in a convenient, effective, and integrated fashion. Just as the World Wide Web and digital-library technologies have simplified the process of publishing and accessing multimedia documents, THREDDS is building infrastructure needed for publishing and accessing scientific data in a similarly convenient fashion. The THREDDS Data Server (TDS) is a web server that provides metadata and data access for scientific datasets, using OPeNDAP, OGC WMS and WCS, HTTP, and other remote data access protocols. The THREDDS Data Server is implemented in Java, and is contained in a single war file, which allows easy installation into a servlet container such as the open-source Tomcat web server. Configuration is made as simple and as automatic as possible, and we have made the server as secure as possible. The library is freely available and the source code is released under the (MIT-style) netCDF library [license](#).

Motivated by LEAD and other use cases and to advance the goals and capabilities of the THREDDS Data Repository technology, RAMADDA, Repository for Archiving, Managing and Accessing Diverse Data, is an open data management framework that provides a publishing platform, content management system, and collaboration services for Earth Science data. RAMADDA enables a data provider or a community of users to upload, manage, and share large data holdings, has been designed to run in multiple contexts, ranging from local desktop use to real time and case-study data archives. RAMADDA incorporates a number of rich information management facilities, including an extensible metadata framework, grouping, tagging, annotations, and associations. Based on an extensible Model-View-Control (MVC) framework, RAMADDA is a Java-based server that runs under Tomcat or can be run stand-alone and it offers the ability to plug-in a variety of ways to access and deliver information.

#### **4. Contributions to the Social Sciences**

LEAD made early and substantive contributions to the understanding of virtual organizations by engaging a team of social scientists from the University of Michigan's School of Information (Tom Finholt and Katherine Lawrence) under financial support from the NCSA with the goal of identifying issues and challenges associated with promoting broad adoption of the LEAD system by a diverse community of researchers, educators, students and operational practitioners. The group participated in more than 130 hours of project-related meetings, held semi-structured, in-depth interviews with members of the LEAD research and development team. The group also visited six of the nine participating institutions in person, spending more than 15 days on site, viewing each site's resources and work environment and interviewing members of the project team.

A second survey was administered to members of the target audience for LEAD. The survey sample was derived from the members of the American Meteorological Society (AMS), which has the mission of promoting the development and dissemination of information and education on atmospheric sciences as well as the related oceanic and hydrologic sciences. Three hundred members were selected, representing a random, stratified sample proportional to the geographic distribution of membership across the United States. The survey response rate was 68%. Of those respondents, the largest number represented individuals active in weather forecasting and analysis (44%). Most respondents were unfamiliar with the LEAD project (80%). The respondents were 83% male and 9% female (the remainder did not answer this question). Approximately one-third of the respondents worked for a government agency (35%), while the remainder reported affiliations with an academic institution (23%), business or service provider (20%), or other type of organization (22%). Their education was almost evenly divided between bachelor's (36%), master's (31%), and PhD (33%) degrees.

From these surveys the community engagement group produced several reports and has identified aspects of the project requiring additional support for successful community engagement. Specifically, for the project team, additional support for decision making and tracking components has been necessary, along with alternative ways of modulating the large amounts of communication and document sharing necessary for a large project the size of LEAD. For the target audience, the surveys identified key technological requirements as well as barriers to adoption. The community engagement team has published their results for broader dissemination in Lawrence 2006, Lawrence and Kim 2006, Lawrence, Finholt and Kim 2006.

Large, multidisciplinary, geographically distributed collaborative research projects are known to have more difficulty bringing about objectives than projects of the same size that reside in a single institution (Lawrence et al 2006). LEAD, however, is a successful cross-institution project, and thus contributes to the deeper understanding of virtual organizations. As reported in Lawrence et al 2006,

*“LEAD relied heavily on technologies to support interaction, but these overshadowed the true core of their success: the team members’ collaboration history and interpersonal skills.”*

## **5. Advancements in Education**

From its inception LEAD has advanced initiatives that aspired to democratize the availability of weather technologies for research and education, lower the barrier to entry, empower application in a grid context, increase the realism of how technologies are applied, and facilitate rapid understanding, experiment design and execution. LEAD was organized according to thrusts with the Education and Outreach Thrust (EOT) responsible for ensuring that the design, development, and implementation of the features of this web-based service-oriented architecture provided user interfaces that were functional, appealing, and congruent with instructional approaches and classroom capabilities. The LEAD EOT was a diverse group of educators, atmospheric and computer scientists and students, external partners that specialized in assessment and instructional design. We also had strong ties with other sponsored education initiatives such as NCSA Cybereducation Group and WeatherRATS, and many of these collaborations will continue as an enduring testimony to the community engagement efforts put in place as part of the LEAD organizational strategy.

Advancements in education that grew out of this collaborative effort include a suite of LEAD-to-Learn (L2L) modules linked to the portal that served the dual purpose of testing the functionality and capabilities developed by the cyberinfrastructure thrusts, while providing the greater education community with interactive web-based classroom supplements that incorporate expert-designed multilevel learning strategies, goals, and outcomes.

Over the lifetime of the LEAD project, L2L modules evolved from self-contained, web-based educational materials linked to the LEAD portal that utilized innovative software and middleware to explore the atmosphere phenomena (e.g. lake-effect snows, fronts, the polar jet stream, land/sea breezes, the skew-T diagram, and quasi-geostrophic theory), to advanced interactive, inquiry-based learning environments that incorporated features such as an ontology vocabulary specific to mesoscale meteorology, glossary, and mouse-over definitions of terms tailored for novice, intermediate, and advanced users. Collaboration with the NCSA Cybereducation Group and instructional design specialist led to a host of new features, including learning outcomes, which helped transform the L2L into the interactive, inquiry-based, and self-guided experience that they are today. The modules are primarily designed for undergraduate education and have

been successfully incorporated into specific upper-level undergraduate courses at Millersville University and the University of Oklahoma.

Complementing these activities, the LEAD EOT team engaged teacher partners from an area secondary school, to assist in extending the use of the modules to pre-college students. L2L modules included “exploring air temperature and pressure through visualizations,” “exploring the structure of the jet stream,” and “investigating surface observations” and are being used in high school advanced placement courses and have been shared with the educator community through their participation in the LEAD-NCSA educator workshop, CASA WeatherRATS workshop, the National Science Teachers Association LEAD workshop, and an IEEE e-science workshop. Notably, in an effort toward democratizing LEAD and expanding the use of L2L modules to include the pre-college classroom, a modified version of Unidata’s Integrated Data Viewer (IDV) called Basic IDV was developed primarily by team of students at NCSA that reduced the number of parameters resulting in greater accessibility using minimally configured computers often found in secondary schools. Unidata has since adapted this feature in IDV.

Team collaboration has been a central modus operandi of the LEAD experience. Nowhere has this collaboration been more effective than between the NCSA Cybereducation Group led by Edee Wiziecki and the Millersville EOT team, which includes the teacher-partners and the instructional design expert, Dr. Lynne Davis (UCAR). These efforts have resulted in co-institutional presentations in a number of venues including two TeraGrid Conferences (2007 and 2009). TeraGrid highlighted LEAD’s impact on education in their 2009 publication *TeraGrid: Education, Outreach and Training* (see TeraGrid Campus Champions; <http://teragrid.org/eot/campuschamps.html>).

Another meaningful collaboration was between the LEAD EOT and the NOAA Center for Atmospheric Science (NCAS) Weather Camp (WC) at Howard University. The NCAS Weather Camp is funded through the NOAA Educational Partnership Program for Minority Serving Institutions (EPP/MSI). Students working with LEAD PIs at both institutions prepared demonstration and visualization materials that were used by high school students from across the country to learn about and interact with the weather. With considerable assistance from LEAD, WC incorporated the LEAD Portal’s forecasting and visualization capabilities in WC 2008 (LEAD year 5) forecasting contest. Model runs were submitted each day between 2 and 4pm, and the results were evaluated and discussed during the weather forecast labs that were conducted each evening. The students worked in groups of two and each group ran two WRF runs per day. The first run (default) was a NAM initialized, 36 hour, CONUS run. For the second run, the groups changed one of the namelist options -- for example some changed the land parameterization, the cloud microphysics, or cloud parameterization options. WC instructors contextualized the physical meaning of these changes at the experience level of the students. Overall the students were intrigued by use of the portal and by producing actual forecasts from their own model runs.

LEAD EOT, with strong and invaluable participation from the full LEAD team, played a key role in organizing workshops that disseminated LEAD capabilities to the education community. These include:

- LEAD-NCSA Workshop for pre-college teachers held at NCSA, Campaign, IL
- 2006 Unidata Users Workshop: Expanding the Use of Models as Educational Tools in the Atmospheric and Related Sciences, Boulder, CO.
- AMS Workshop on *Linked Environments for Atmospheric Discovery* (LEAD): An Emergent Information Technology Environment for On-Demand, Dynamically Adaptive

Interaction with Weather for Research and Education, 20 January 2008, New Orleans, LA

- Fourth IEEE International Conference on e-Science, Real Time Geoscience Data Use in Education Workshop, 8 December 2008 at IUPUI, Indianapolis, IN

LEAD EOT also organized one of the bolder community outreach activities of the project by engaging *WxChallenge*, the national collegiate forecasting competition sponsored by the University of Oklahoma, to offer a pilot program where a select subset of contest participants were given access to a production version the LEAD portal and its tools and services. Over a period of seven weeks, 62 undergraduate and 13 graduate students from 10 institutions were given authorization to build experiments, compose their own workflows, use computing resources to run the WRF numerical prediction system, save the model output in a personal myLEAD workspace, and visualize the WRF output using Unidata's IDV. Students integrated the high-resolution WRF output products into their personal schema for preparing a forecast for stations previously selected by WxChallenge. A team effort involving all LEAD partners allowed users to: 1) establish a 5-km resolution nested WRF domain inside a prescribed 20 km CONUS; 2) select between the ARPS Data Assimilation System (OU) or NAM gridded fields for model initialization and determining lateral boundary conditions; 3) launch the WRF model over a user-specified domain that ran over distributed TeraGrid computing resources; 4) transfer the model output into the student's own myLEAD workspace; and finally, 5) visualize and analyze the output products for enhanced forecast decision making. The pilot project exposed strengths and weaknesses of the distributed service-oriented architecture that were used to improve the functionality and reliability in subsequent production build-outs. *Moreover, in a survey of participants following the pilot, favorable ratings were given for the vision of LEAD as an advanced capability with the potential for changing the paradigm for access to data and high-resolution mesoscale models.*

One of the most satisfying aspects of the LEAD project has been the active, dedicated, and enthusiastic involvement of undergraduate and graduate students across all thrusts and institutions. Over the five years, the EOT team alone involved 30 undergraduates at Millersville University working with six computer science undergraduates at NCSA, and graduate students at Howard and Indiana Universities. These students interacted directly with PIs, research scientists, computer technicians and software engineers, teacher-partners, instruction specialists, and peer-to-peer. LEAD EOT students presented results of their efforts at AMS conferences, LEAD sponsored workshops, the NSF site visit in year 2, and two TeraGrid Conferences (see publications list). Students were the primary contingent for testing the LEAD portal and its underlying functionality, reporting problems to the Indiana University and Unidata technical support team, and evaluating improvements and upgrades to the production versions. They also were principals in the development of the L2L modules and software for novice users<sup>1</sup>, creating materials for the NCAS Weather Camps, working closely with teacher-partners, and collaborating with their peers in the CyberEducation Group at NCSA. Students advancing LEAD education and outreach contributed significantly to the broader impacts of this collaborative research.

---

<sup>1</sup> Basic IDV and NCAnnihilator

## 6. Further Advancements in Outreach and Diversity

LEAD organized and hosted a workshop September 17-18, 2009 that attracted nearly 60 researchers from the geosciences, and cloud computing and collaborative technologies research areas. The workshop, titled "Cloud Computing and Collaborative Technologies in the Geosciences" addressed the suitability of cloud computing and collaborative technologies, two emerging technological advancements in facilitating geoscience research and education. A report is forthcoming.

In December 2007, Dr. Daniel Atkins, then Director of the NSF Office of Cyberinfrastructure, asked LEAD to be part of an NSF effort to create a set of short-format video "cyberinfrastructure exemplars" of exciting narratives of leading edge applications for discovery and learning. The principal audience was Congress, but with outreach to more general audiences. The video is available in HD format and a lower-resolution version is on the web at [http://kkd.ou.edu/LEAD\\_Video.mpg](http://kkd.ou.edu/LEAD_Video.mpg)

The LEAD diversity strategy emphasized engagement of women and minorities in atmospheric and computer sciences research and development, capacity building at minority serving institutions (MSIs), and targeted outreach. In terms of engagement of women and minorities in research the representation in LEAD exceeded national diversity norms in key categories. For example among women faculty in computer science LEAD was above the norms by 33%. In atmospheric science where the production of minority PhD students has historically been among the lowest of the natural sciences, LEAD's impact was notable. LEAD contributed to the production of five minority PhDs: three through partial direct support and two indirectly through leveraged involvement in LEAD activities. Among this group two were women. LEAD contributed to the production of two minority atmospheric science MS students and six minority MS computer engineering students. Considering that between 1973 and 2003 the national production of minority PhDs in atmospheric science was 51 (3 women) or 1.7 per year these results are impactful on the discipline.

MSI capacity-building and empowerment in graduate meteorology research is another means by which LEAD has contributed to diversity enhancement. As addressed in previous reports LEAD aimed to enable MSI graduate students to conduct research on advance topics in mesoscale meteorology by providing easy access to sophisticated models and data assimilation tools without the need for time-consuming software engineering or significant departmental resources to configure these tools for their study. The following are examples of how this was demonstrated. Howard University graduate student Segalye Walford's association with LEAD helped her to complete her dissertation research project on convective initiation. As part of her study Dr. Walford used [the WRF Variational Data Assimilation System \(WRF-Var\)](#) to assimilate water vapor profiles derived from a mesoscale network of LIDARs across the mid-Atlantic region. She investigated the extent to which assimilation of this data improved quantitative precipitation forecasts with WRF. Part of Ms Walford's research with WRF-VAR was motivated to contribute to efforts at Howard to investigate the application of WRF-VAR in the LEAD Portal. Involvement in LEAD enabled students, such as Howard University masters student Hector Lima, whose story was given in the introduction of this report, to engage in much richer and deeper scientific discovery in meteorological course work and research.

With regard to outreach activities designed to impact diversity, LEAD supported the Howard University Weather Camp as discussed in more detail in the E&O section above. One of the main goals of Weather Camp is to expose traditionally disadvantaged pre-college population to atmospheric science to help improve the numbers of these students that enter the field at the baccalaureate level. Although statistics on the success of the program towards these goals are yet to

be analyzed, LEAD support enriched Weather Camp and impacted close to 100 high school students within this population.

		CS/IT Faculty	CS/IT Non-Faculty	CS/IT Grad Students	CS/IT U-Grad Students	Metr Faculty	Metr Non-Faculty	Metr Grad Students	Metr U-Grad Students	Total
<b>National Diversity Norms</b>	<b>Women</b>	9%	29%	26%	N/A	15%	33%	16%	N/A	
	<b>Ethnic Minorities</b>	14%	26%	42%	N/A	14%	23%	8%	N/A	
<b>LEAD Year 1 (2003-2004)</b>	<b>Women</b>	25%	9%	13%		16%	8%	7%		
	<b>Ethnic Minorities</b>	25%	6%	0%		16%	0%	50%		
<b>LEAD Year 2 (2004-2005)</b>	<b>Women</b>	40%	21%	8%	0%	20%	25%	14%	75%	
	<b>Ethnic Minorities</b>	20%	14%	33%	0%	20%	13%	57%	25%	
<b>LEAD Year 3 (2005-2006)</b>	<b>Women</b>	40%	17%	32%	0%	20%	33%	17%	23%	24%
	<b>Ethnic Minorities</b>	0%	13%	37%	0%	20%	17%	83%	15%	23%
<b>LEAD Year 4 (2006-2007)</b>	<b>Women</b>	40%	24%	21%	33%	13%	0%	14%	38%	23%
	<b>Ethnic Minorities</b>	0%	15%	57%	0%	50%	22%	57%	8%	27%
<b>LEAD Year 5 (2007-2008)</b>	<b>Women</b>	67%	19%	7%	33%	13%	0%	29%	25%	18%
	<b>Ethnic Minorities</b>	0%	4%	13%	33%	13%	0%	57%	0%	11%

Table 2. Diversity representation in LEAD over the life of the project. Note that national norms are shown in the first rows.

## 7. Year 6 Publications and Talks

1. Droegemeier, K.K., B. Plale, M. Ramamurthy and C. Mattocks, 2009: A new approach for using web services, grids, and virtual organizations in mesoscale meteorological research. Preprints, *25th Conf. on Integrated Information and Processing*, Phoenix, AZ, Amer. Meteor. Soc., CD-ROM Paper 6.B2.
2. Droegemeier, K.K. and Y. Wang, 2009: Dynamically adaptive numerical weather prediction, models, observations and cyberinfrastructure responding to the atmosphere. Preprints, *23rd Conf on Wea. Analys. And Forecasting and 19th Conf. on Num. Wea. Pred.* 1-5 June, Omaha, NE, Amer. Meteor. Soc., Paper 14A.1.
3. Scott Jensen and Beth Plale, Extended Abstract: Schema-Independent and Schema-Friendly Scientific Metadata Management, *4th International IEEE Conference on e-Science*, Indianapolis, IN Dec 2008.
4. Li, X., B. Plale, N. Vijayakumar, R. Ramachandran, S. Graves and H. Conover, 2008: Real-time Storm Detection and Weather Forecast Activation through Data Mining and Events Processing. *Earth Science Informatics, Springer Verlag, Vol 1, no 2, pp. 49-57.*
5. Suresh Marru, Marlon Pierce, Chathura Herath, Srinath Perera, Poster: Open Grid Computing Environment's Workflow Suite for E-Science Projects, 4th International IEEE Conference on eScience, pp.332-333, 2008
6. Mattocks, C., K.K. Droegemeier and R.B. Wilhelmson, 2009: Integration of LEAD and WRF Portal technologies to enable advanced research, operations and education in mesoscale meteorology. Preprints, 23rd Conf on Wea. Analys. And Forecasting and 19th Conf. on Num. Wea. Pred. 1-5 June, Omaha, NE, Amer. Meteor. Soc., Paper 12B1.
7. Srinath Perera, Dennis Gannon, "Enforcing User-Defined Management Logic in Large Scale Systems," IEEE Congress on Services, pp.243-250, 2009.
8. Srinath Perera, Suresh Marru, Thilina Gunarathne, Dennis Gannon, and Beth Plale, Application of Management Frameworks: A Case Study on Managing Workflow-based Systems, International Conference on Web Services (ICWS), Los Angeles, July 2009.
9. Srinath Perera, Suresh Marru, Dennis Gannon, Poster: Monitoring and Managing E-Science Cyber-Infrastructures: A Case Study, 4th International IEEE Conference on eScience, pp.402-403, 2008
10. Yiming Sun, Scott Jensen, Chathura Herath, Jeff Cox and Beth Plale, "Tutorial: Event Processing in Weather Responsive Scientific Workflow Cyberinfrastructure". *ACM Distributed Event Based Systems*, Nashville, July 2009.

### Year 6 Talks

1. A novel approach to detect regions of phenomena from NAM model outputs. Xiang Li, with R. Ramachandran, S. Graves, K. Brewster, S. M. Lazarus, and B. T. Zavodsky, Conference on Integrated Information and Processing (IIPS) 2009
2. Applying LEAD in NCAS Weather Camp. Michael Mogil, with D. K. Bond, S. St. Gerard, E. Joseph, D. V. R. Morris, S. Yalda, T. Baltzer, and S. Marru, 18<sup>th</sup> Symposium on Education, co-located with IIPS 2009.
3. Beyond LEAD: Impact, Education, and Future Plans, Beth Plale, Supercomputing, Austin, TX, November 2008
4. How the NCSA/LEAD Workflow Broker Manages Complex Workflows. Poster (11.1) presentation at the 9th Annual WRF User's Workshop, Alameda, J.C., A. L. Rossi, S. D. Hampton, B. F. Jewett, and R. B. Wilhelmson, National Center for Atmospheric Research, Boulder, CO. June 23-27, 2008.

5. Dynamically adaptive numerical weather prediction, models, observations and cyberinfrastructure responding to the atmosphere. Droegemeier, K.K. and Y. Wang, 23rd Conf on Wea. Anal. And Forecasting and 19th Conf. on Num. Wea. Pred. 1-5 June, Omaha, NE, 2009
6. Hands-on Tutorial on using LEAD/OGCE Workflow Suite, Suresh Marru, SciDAC Users Workshop, June 2009, San Diego, CA.
7. Integrating Weather/Climate Research Models and Data Into End-Use Knowledge, Beth Plale, Invited Talk Notre Dame, February 2009.
8. An Interactive, Integrated, Instructional Pathway to the LEAD Science Gateway, Wilhelmson, R., with e. Wiziechi, A. Knecht, T. Daley, W. Davis, S. Yalda, R. D. Clark, S. States, R. Junod, S. Cecelski, L. David, TeraGrid Science Presentation, June 22 - 25, 2009, Arlington, VA.
9. An interactive, integrated, instructional pathway to the LEAD Science Gateway, Sepideh Yalda, et al. Joint session between 18<sup>th</sup> Symposium on Education and 25<sup>th</sup> Conference on Integrated Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology. 2009
10. It's All In the Data: Workflow Systems and Weather, Beth Plale, American Geophysical Union 2009 Joint Assembly Meeting, Toronto, Canada, May 2009
11. The LEAD cyberinfrastructure: A technical view of lessons learned and current plans, Suresh Marru, Supercomputing, Austin, TX, November 2008.
12. LEAD-to-LEARN: A demonstration of an instructional pathway for undergraduate meteorology education using a science gateway. Sepideh Yalda, with R. D. Clark, S. States, S. Cecelski, B. Kerschner, R. Junod, L. K. Davis, E. N. Wiziecki, T. Daley, and H. Gadde. 25th Conference on Integrated Information and Processing (IIPS), 2009.
13. Numerical Modeling of Atmospheric Phenomena with Brokered Workflows on the TeraGrid, Al Rossi and Robert Wilhelmson, TeraGrid '09, Arlington, VA, June 2009
14. Numerical Modeling of Atmospheric Phenomena with Brokered Workflows on the TeraGrid, B. Jewett, TeraGrid Education Presentation, June 22 - 25, 2009, Arlington, VA.
15. Overview and Tutorial on LEAD/OGCE Workflow Suite, Suresh Marru, TeraGrid'09, June 2009, Washington D.C,
16. Provenance and Metadata in Data-intensive Computing, Beth Plale, Invited talk at Howard University, Washington, DC, October 2008
17. Science Gateways as a Way to Support e-Science and Virtual Organizations via the TeraGrid, Marlon Pierce, Suresh Marru, Stephen Simms, and Craig Stewart, Tutorial at e-Science 2008, Indianapolis, Indiana December 7 2008.
18. Storm Clustering for Data-driven Weather Forecasting, X. Li, 88th AMS Annual Meeting, January 20-28, 2008 New Orleans, LA.
19. TeraGrid Computing at the Center for Analysis and Prediction of Storms, Keith Brewster, TeraGrid '09, Arlington, VA, June 2009
20. Using NCSA/LEAD's Workflow Broker to Study Storm Interaction with WRF. Poster presentation. B. Jewett, 9th Annual WRF User's Workshop, National Center for Atmospheric Research, Boulder, CO. June 23-27 2008.

## References

- Cao, B., B. Plale, G. Subramanian, E. Robertson, and Y. Simmhan (2009), Provenance Information Model of Karma, *IEEE 2009 Third International Workshop on Scientific Workflows (SWF'09)*, Los Angeles, CA July 2009.

- Gannon, D., B. Plale, M. Christie, Y. Huang, S. Jensen, N. Liu, S. Marru, S.L. Pallickara, S. Perera, S. Shirasuna, Y. Simmhan, A. Slominski, Y. Sun, N. Vijayakumar (2007), Building Grid Portals for e-Science: A Service Oriented Architecture. *High Performance Computing and Grids in Action*, IOS Press - Amsterdam, Lucio Grandinetti editor.
- Jensen, S., and B. Plale (2008), Extended Abstract: Schema-Independent and Schema-Friendly Scientific Metadata Management, *4th International IEEE Conference on e-Science*, Indianapolis, IN Dec 2008.
- Lawrence, K. A. (2006), Walking the Tightrope: The Balancing Acts of a Large e-Research Project. *Computer Supported Cooperative Work (CSCW): The Journal of Collaborative Computing*, Springer, Netherlands.
- Lawrence, K. A. and Kim, I. (2006). A Tight Squeeze: Finding Room for HCI in the Space between Research and Development. *CHI 2006 conference workshop "Usability Research Challenges for Cyberinfrastructure and Tools,"* Montreal, Canada.
- Lawrence, K. A., Finholt, T. A., and Kim, I. (2006). Warm Fronts and High Pressure Systems: Overcoming Geographic Dispersion in a Meteorological Cyberinfrastructure Project. *HICSS-40 (Hawaii International Conference on Systems Sciences)*, IEEE.
- Patterson, D., L. Snyder, and J. Ullman (1999) Best Practices Memo, Evaluating Computer Scientists and Engineers For Promotion and Tenure, August 1999, <http://www.cra.org/main/cra.pubs.html#bes>
- Simmhan, Y., B. Plale, and D. Gannon (2008), Karma2: Provenance Management for Data Driven Workflows, Extended and invited from ICWS 2006. *Int'l Journal of Web Services Research*, IGI Publishing, Vol 5, No 2, 2008.
- Wilkins-Diehr, N., D. Gannon, G. Klimeck, S. Oster, and S. Pamidighantam (2008), TeraGrid Science Gateways and Their Impact on Science, *Computer*, IEEE Computer Science Press, pp. 32-41, Nov 2008.