J2.3

LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY (LEAD): ARCHITECTURE, TECHNOLOGY ROADMAP AND DEPLOYMENT STRATEGY

*[1,2]Kelvin K. Droegemeier, [3]V. Chandrasekar, [4]Richard Clark, [5]Dennis Gannon, [6]Sara Graves, [7]Everette Joseph, [8]Mohan Ramamurthy, [9]Robert Wilhelmson, [1]Keith Brewster, [8]Ben Domenico, [1]Theresa Leyton, [7]Vernon Morris, [8]Donald Murray, [5]Beth Plale, [6]Rahul Ramachandran, [10]Daniel Reed, [6]John Rushing, [1]Daniel Weber, [8]Anne Wilson, [1,2]Ming Xue, [4]Sepideh Yalda

[1]Center for Analysis and Prediction of Storms and [2]School of Meteorology
University of Oklahoma
Norman, Oklahoma
[3]Colorado State University
Fort Collins, Colorado
[4]Millersville University
Millersville, Pennsylvania
[5]Indiana University
Bloomington, Indiana
[6]University of Alabama in Huntsville
Huntsville, Alabama
[7]Howard University
Washington, DC
[8]University Corporation for Atmospheric Research
Boulder, Colorado
[9]National Center for Supercomputing Applications
Urbana, Illinois
[10]University of North Carolina
Chapel Hill, North Carolina

1. INTRODUCTION

A 5-year National Science Foundation Large Information Technology Research (ITR) grant – funded on 1 October 2003 and known as Linked Environments for Atmospheric Discovery (LEAD) –is creating the cyberinfrastructure needed to facilitate the identification, access, preparation, assimilation, prediction, management, analysis, mining, and visualization of a broad array of mesoscale meteorological data and model output, independent of format and physical location (Droegemeier et al. 2004).

A transforming element of LEAD is Workflow Orchestration for On-Demand, Real-Time, Dynamically-Adaptive Systems (WOORDS), which allows the use of analysis tools, forecast models, and data repositories not in fixed configurations or as static recipients of data, but rather as dynamically adaptive, on-demand, Grid-enabled systems that can a) change configuration rapidly and automatically in response to weather; b) continually be steered by new data; c) respond to decision-driven inputs from users; d) initiate other processes automatically; and e) steer remote observing technologies to optimize data collection for the problem at hand. Simply put, LEAD is creating the IT needed to allow people (students, faculty, research scientists, operational practitioners) and atmospheric tools (radars, numerical models, data assimilation systems, data mining engines) to interact with weather (Figure 1.1).

LEAD is targeted principally toward the meteorological higher education and research communities, though as described below, LEAD also is developing learning communities, centered around teacher-partners and alliances with educational institutions, to bring the benefits of LEAD technologies to grades 6-12.

* Corresponding Author Address: Prof. Kelvin K. Droegemeier, University of Oklahoma, 100 East Boyd Street, Suite 1110, Norman, OK 73019. Email: kkd@ou.edu. Visit LEAD on the web at http://lead.ou.edu.

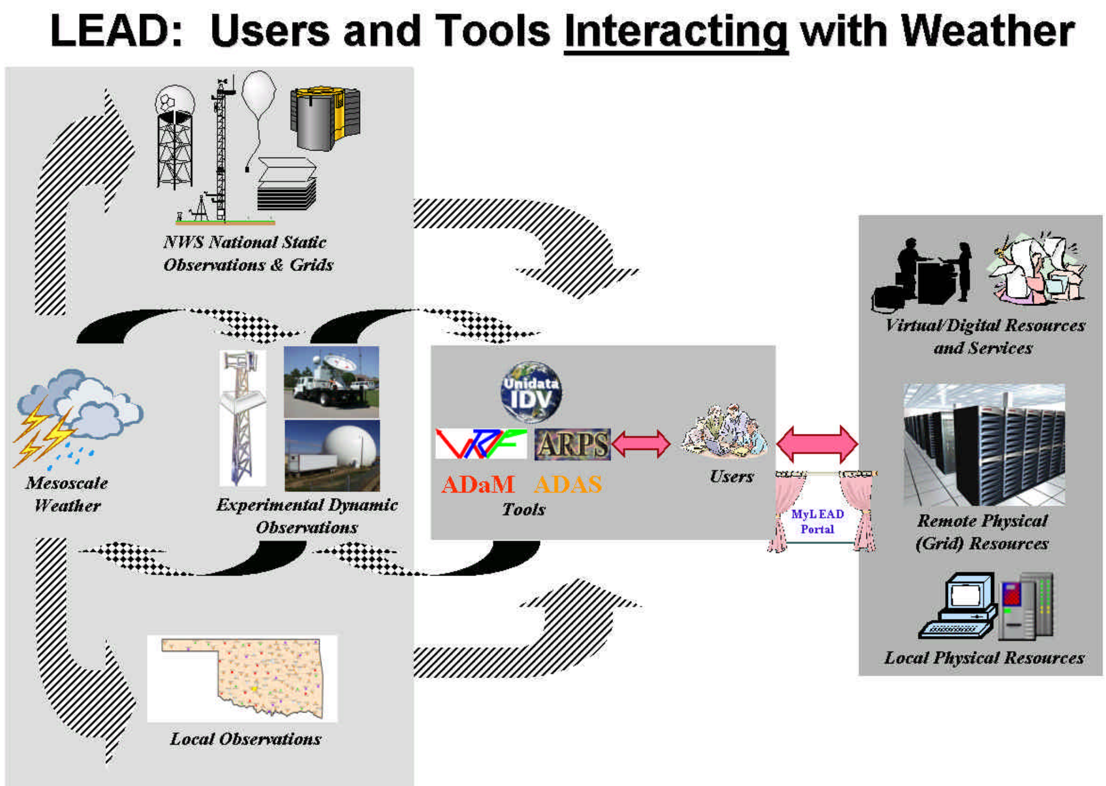

Figure 1.1. The LEAD system consists of two levels of interaction among people, tools and mesoscale weather. First, streaming observations of mesoscale weather (left center of diagram) collected by national and local observing systems feed (cross-hatched arrows) tools that respond to changing weather conditions. Second, these same tools have the ability to drive dynamically adaptive remote sensing systems, such as Doppler radars. All elements of the system make use of local and remote computing, storage, and visualization resources and can operate within the LEAD portal or as stand-alone services.

The field deployment of LEAD is being orchestrated via a phased approach involving a number of test beds and strategic partners. It is taking place via two principal mechanisms, the first of which is the UCAR Unidata program that involves approximately 150 organizations encompassing 21,000 university students, 1800 faculty, and hundreds of operational practitioners. The second is the nascent Developmental Test Bed Center (DTC) at the National Center for Atmospheric Research. The DTC, sponsored by the NSF and NOAA, provides a national collaborative framework in which numerical weather analysis and prediction communities can interact to accelerate testing and development of new technologies as well as techniques for research applications and operational implementation – all in a way that mimics, but does not interfere with, actual forecast operations. It is anticipated that the DTC will become the focal point for mesoscale model experimentation and the transfer of new concepts and technologies into operational practice.

2. THE LEAD SYSTEM AND SUPPORTING RESEARCH

Having been in operation for slightly more than a year, LEAD has created a technology roadmap and architecture for developing its capabilities and placing them within the academic and research environment. The philosophical underpinning of LEAD is *WOORDS – Workflow Orchestration for On-Demand, Real-Time, Dynamically-Adaptive Systems*. WOORDS expressions the notion of users creating, managing, and altering workflows, dynamically and automatically in real time, in the context of systems that adapt themselves to changing circumstances such as weather. As used in LEAD, the components of WOORDS have the following meaning:

- Workflow Orchestration -- The automation of a process, in whole or part, during which tasks or information are passed from one or more components of a system to others -- for specific action -- according to a set of procedural rules.

- On-Demand – The capability to perform action immediately with or without prior planning or notification.

- Real-Time – The transmission or receipt of information about an event nearly simultaneous with its occurrence, or the processing of data or execution of other commands in step with wall clock time.

- Dynamically-Adaptive – The ability of a system, or any of its components, to respond automatically, in a coordinated manner, to both internal and external influences in a manner that optimizes overall system performance.

- System – A group of independent but interrelated elements that operate in a unified holistic manner.

Although mesoscale meteorology and numerical weather prediction represent perhaps prototypical applications of WOORDS, the concept is far more general. As an example, effective suppression of wild land fires will in the future depend upon numerical simulations that use evolving conditions of weather, fuel availability, the location of burn lines, etc. Additionally, embedded sensors that measure roadway conditions and highway traffic flow will in the future be able to reroute traffic so as to optimize throughput based upon models that input continually changing conditions. These and other examples show the importance of LEAD evolving WOORDS as a general notion – a paradigm of broad applicability – from which the non-meteorology community can benefit.

The LEAD system, shown schematically in Figure 2.1, consists of the following principal components:

- User Sub-System – Consists of a Web Portal, which is the principal, though not only mechanism by which users can access LEAD technologies, along with the myLEAD personal work space (including storage and cataloging capabilities) and the Geo-reference graphical user interface.

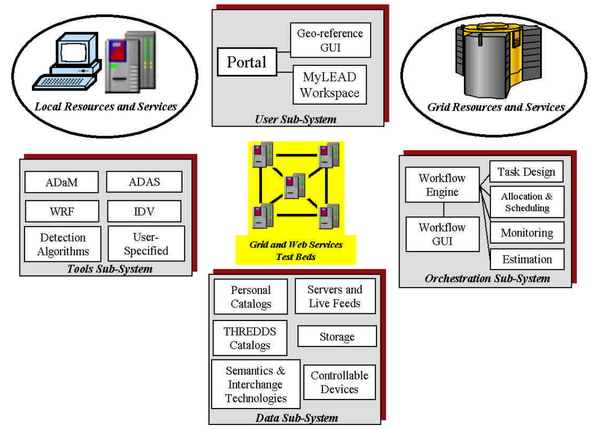

Figure 2.1. The LEAD System.

- Data Sub-System – Handles everything associated with data and meta data, including ingest, cataloging, and storage management of observations (real time streaming, archived, experimental or operational), numerical model output produced by operational or experimental models, and user-generated information.

- Tools Sub-System – Consists of all meteorological and IT tools including but not limited to the WRF model, ADaM data mining engine, ADAS data analysis system, ESML, hazardous weather detection algorithms, data decoders, and "hooks" for user-supplied tools.

- Orchestration Sub-System – Provides the technologies that enable users to manage flows of data and model execution streams, as well as create and mine output. It further provides linkages to other software and processes for continuous or on-demand application, including the steering of remote observing systems.

- Grid and Web Services Test Beds – Located at five of the nine participating institutions (Oklahoma, Unidata, Illinois, Indiana, and Alabama in Huntsville, with a sixth developmental/non-general access site at the University of North Carolina), these distributed computing systems represent a "mini grid," operated by LEAD to ensure strict software compatibility, that serves as a test bed for developing, integrating, and testing all components of LEAD. The LEAD system will be migrated to other grids as the project proceeds, and as grid technologies mature to provide a suitably stable framework.

Much of the research in LEAD is directed toward creating a fabric that links these various sub-systems in a manner that is transparent to the user. It consists of three distinct but highly synergistic elements: (1) basic IT and CS research, driven by the unique needs of mesoscale meteorology, to enable the LEAD system described above; (2) mesoscale meteorology research needed to both enable the LEAD system and use it to address important scientific problems; and (3) the deployment and refinement of tools and technologies by researchers, educators, and operational practitioners.

This research in LEAD is not *ab initio*, but rather builds upon several enabling technologies shown in Figure 2.2. Many of these were pioneered by the LEAD team and are being enhanced for use in LEAD.

Capability/Resource	Principal Technologies
Atmospheric, Oceanographic, Land-Surface Observations	CONDUIT, CRAFT, MADIS, IDD, NOAAPort, GCMD, SSEC, ESDIS, NVODS, NCDC
Operational Model Grids	CONDUIT, NOMADS
Data Assimilation Systems	ADAS, WRF 3DVAR
Atmospheric Prediction Systems	WRF, ARPS
Visualization	IDV
Data Mining	ADaM
NSF NMI Project	Globus Tool Kit
Semantic Interchange and Formatting	ESML, NetCDF, HDF5
Steerable Observing Systems (Radars)	CASA OK Test Bed, V-CHILL
LEAD Portal	NSF NMI Project (OGCE)
Workflow Orchestration	BPEL4WS
Monitoring	Autopilot
Data Cataloging/Management	THREDDS, MCS, SRB

Figure 2.2. LEAD enabling technologies.

The computer science research in LEAD focuses on the following elements:

- Workflow Orchestration – Development of capabilities that will allow users to construct and schedule execution task graphs with data sources drawn from archived as well as real-time sensor streams and output. Particular emphasis is given to workflows that can change dynamically in concert with user needs, data, and output.
- Interaction With and Control Over Dynamically Adaptive Sensors – Research that will produce appropriate protocols, command interfaces, and related linkages between meteorological tools and sensors to effectuate two-way adaptivity.

- Data Streaming – Development of capabilities to support robust, high bandwidth transmission of multi-sensor data in a time-continuous manner with fault tolerance.

- Distributed Monitoring and Performance Estimation – Creation of mechanisms to enable soft real-time performance guarantees by estimating resource behavior to ensure timely completion of tasks – which is especially critical in real time environments.

- Data Management – Creation of the infrastructure needed to support the storage and cataloging of observational data, model output and results from data mining.

- Data Mining – Development of the tools needed to enable users to glean insights from data and model output, particularly with regard to streaming information (e.g., from NEXRAD Doppler radars).

- Semantic and Data Interchange Technologies – Adoption/refinement of technologies to enable the use of heterogeneous data by diverse tools and applications.

The meteorology research in LEAD focuses on the following elements:

- ARPS Data Assimilation System (ADAS) for the WRF Model – Adaptation of the CAPS ADAS to the WRF model to allow users to assimilate a wide variety of observations in real time, especially those collected locally (e.g., from mesonetworks).

- Orchestration System for the WRF Model – Development of a process control system to allow users to manage flows of data, model execution streams, the creation and mining of output, and linkages to other software and processes for continuous or on-demand application, including steering of remote observing systems. This is a highly synergistic effort with the CS workflow component.

- Fault Tolerance in the WRF Model for On-Demand, Interrupt-Driven Utilization – Development of the capabilities needed to accommodate interrupts in streaming data and user execution commands in the WRF model and perhaps other tools.

- Continuous Model Updating – Application of advanced data assimilation techniques, most notably ensemble Kalman filtering, to allow the WRF to be steered continually by observations and thus be dynamically responsive to them (in comparison to the more conventional sequential data assimilation framework that operates intermittently with pre-determined data cut-off times).

- Hazardous Weather Detection and Data Mining – Development of advanced data mining techniques for identifying hazardous weather in gridded forecasts and assimilated data sets as opposed to traditional decision support tools that primarily use information from sensor data alone in their native coordinate systems.

Programmatically, research and development within LEAD is organized around five parallel research thrusts (Figure 2.3), along with two cross-cutting test beds to ensure that the former are tightly and continuously integrated.

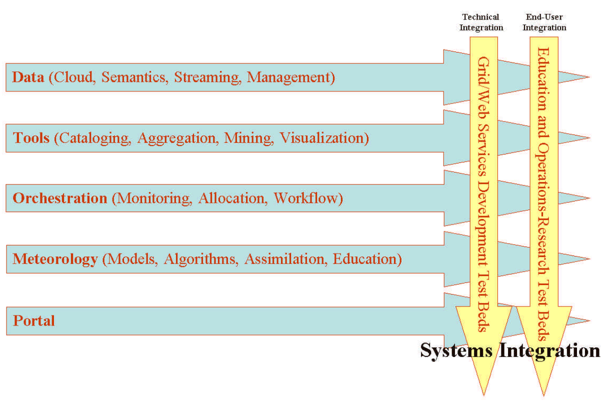

Figure 2.3. Organization of LEAD research and development.

3. DEVELOPMENT PATHWAY

Because LEAD is creating a complex cyberinfrastructure that contains many components, the development and integration of which must be highly coordinated among disciplines and institutions, it is vitally important that an effective technology development process be in place to provide direction while not stifling creativity. Likewise, given that the principal goal of LEAD is to evolve an infrastructure suitable for tackling a wide array of basic research problems in mesoscale meteorology, and for meeting the needs of operations research as well as grade 6 through post-graduate education, it is important that research and education remain the primary motivation behind everything done within LEAD.

To accommodate these requirements, LEAD has created a set of canonical problems in mesoscale meteorology research and education:

- Problem #1: Radar-Based Climatology of Convective Storm Characteristics

- Problem #2: Relating Convective Storm Behavior to the Ambient Environment

- Problem #3: Dynamically Adaptive, High-Resolution Nested Ensemble Forecasts Using Static Observations

- Problem #4: Dynamically Adaptive Remote Sensing of Severe Weather

- Problem #5: Dynamically Adaptive Remote Sensing Coupled with Numerical Prediction

They are not meant to be exhaustive, but rather representative of key problems for which solutions are desired, but which cannot be addressed using present capabilities. These problems serve to organize and focus the LEAD research process (lowest box in Figure 3.1), thus forming its foundation. From them system functional requirements (next box upward in Figure 3.1) are defined and used to identify fundamental scientific and technological barriers – requiring basic research and some development – that are needed to create a specific capability. Proceeding upward in Figure 3.1, the system architecture then is defined, along with appropriate services, leading to the conduct of basic research, the creation of functional prototypes, and the development of test beds. The totality of this work then is integrated, leading to technology generations (Section 8) that can be tested and deployed within the community. The last step, shown by the top box in Figure 3.1, is in fact the starting point – the use of LEAD capabilities to address research and education problems. In this manner the foundation, and the capstone, of LEAD are basic research and a system for application to a broad range of educational activities.

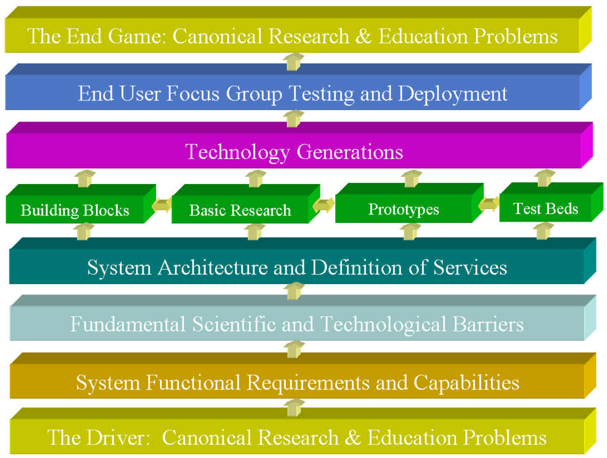

Figure 3.1. The LEAD research process.

4. SERVICE-BASED INFRASTRUCTURE

The LEAD system is being developed as a sequence of prototypes of increasing sophistication that serve to test and refine research concepts, engage end users from the outset, demonstrate tangible capabilities and stimulate new ideas, and provide a mechanism for ensuring continuous integration among the multiple disciplines within LEAD.

The fundamental "building blocks" of these prototypes are a series of web services, shown schematically in Figure 4.1, which themselves can consist of web services and be used as stand-alone applications or as part of the overall LEAD

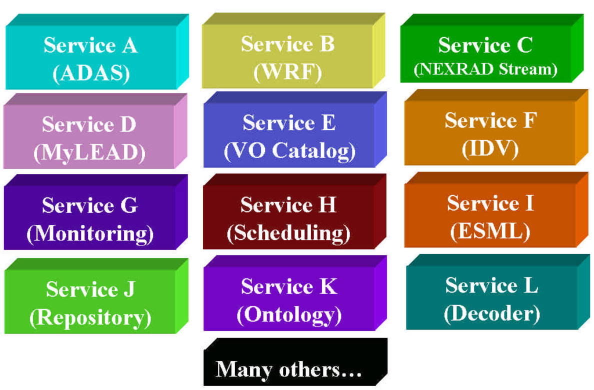

Figure 4.1. Web services, the building blocks of LEAD.

environment. The web service model is especially attractive from the user point of view because it allows for loose coupling of capabilities represented by the services. Within this web service framework, other types of services may be invoked such as machine-specific desktop applications (e.g., high-end visualization tools) that typically require software installation.

Once the full complement of LEAD web services has been developed, users will be able to combine them, via an orchestration GUI in the LEAD Portal (Section 7), in numerous ways (Figure 4.2) to create a wide array of capabilities, adding other services as necessary, or creating and saving workflows of services that themselves can be combined in new workflows to solve increasingly complex tasks.

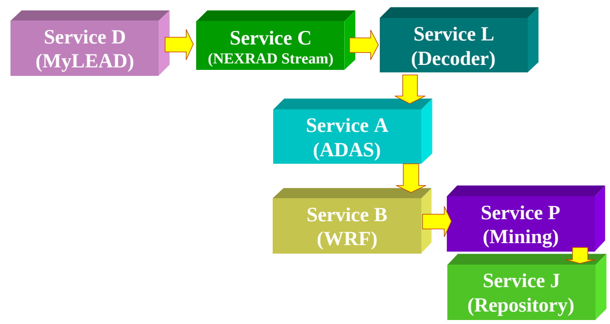

Figure 4.2. By combining LEAD services using the workflow engine, users can create a wide array of capabilities for research and education. Additional services can be added by the user as needed.

5. SYSTEM ARCHITECTURE

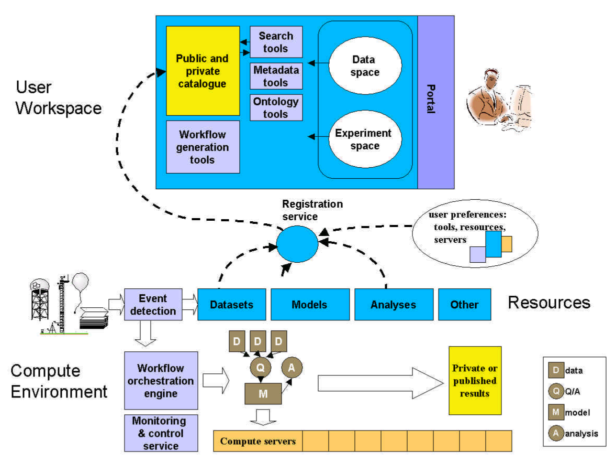

Figure 5.1. LEAD architecture.

The architecture of LEAD (Figure 5.1) can be characterized as service-oriented, i.e., it is organized as a set of services exposed to the network. A *service* is a logical manifestation of some physical or logical resource (such as data bases, programs, devices or people), along with application logic. Services are distributed across spatially dispersed computing resources and interact with one another, and with people, by means of exchanging messages. The capabilities of a service are exposed through a service interface which enables other services to discover the presence of a new service without manual intervention by a user.

As noted in Section 4, a service representation for established tools such as IDV (visualization), WRF (prediction), ADAM (data mining) and ADAS (analysis and assimilation) can be accomplished in two ways: directly, by wrapping the tool with a service interface, or by situating the tool behind a family of services or a gateway service. For instance, a family of services that implements a broker would be used when tools need to run on powerful platforms, e.g., the Teragrid. Additionally, the service framework is being constructed with sufficient generality to allow a variety of tools to be flexibly substituted at key points in the LEAD system, e.g., to allow substitution of the WRF 3DVAR assimilation package in place of ADAS.

Access to LEAD services is facilitated via a web portal (Section 7) or through desktop tools. User-side services such as search tools, metadata tools, ontology tools, and public and private catalogs support the user in her navigation of the data and experiment space. This is depicted in Figure 5.1 as services existing within the user environment to the left of the portal. The public and private catalogs in that same space catalog resources that are important to the organization and the individual user, respectively. Possible resources shown in the middle of Figure 5.1 include models, data sets and analysis tools.

The orchestration of a LEAD task, shown across the bottom of the figure, can be impacted by changing weather conditions. For instance, services can interact with regional data sources that provide information about a significant local weather event. Additionally, data analysis or assimilation of model results could feed back via service calls to experimental radar systems, such as those being developed by CASA (McLaughlin et al. 2005), to place reservations for preferred radar observation patterns near the time and location of the predicted weather.

6. MyLEAD

Mesoscale meteorology forecasting is data-driven, with reasonably stringent timeliness demands on the results. The quantity of data available to an individual researcher is staggering. Data is continuously streaming from hundreds of sources, and the data products of the forecasts themselves represent a data space equally as large. This rapidly proliferating information space can be overwhelming for a scientist without proper tools. We posit that a user must have a sense of control over the vast information space, and that can be achieved through the personalized view. A personalized view is a subset of the information space of interest to an individual; it is the space over which the individual has control. A good analogy to this is the Internet. The Internet is too vast an information space for any one person to completely comprehend. To make sense of the space, a person constructs and maintains a personal view of the Internet. The search engine allows a person to locate objects and add the object to their personal "view" by means of bookmarking pages. Through creating a web page, responding to a message list, or writing to a blog; an individual publishes content to the web. These search, tagging, and publishing tools are key to an individual's perception of control over the vast space. In a similar spirit, we are building a tool called myLEAD that permits personalized searching, storing, and publishing of data products in the LEAD (http://lead.ou.edu) information space.

The personal orientation of the myLEAD service is the primary driver behind many of its interesting technical challenges. A graduate student should be able to conduct experiments in the myLEAD information space for the several years of her graduate study. She must be able to fully trust that the space is protected from malicious intent. Files generated in an experiment cannot be lost, corrupted, or misplaced over the years. An experiment conducted several years ago must be recoverable. The researcher must feel she has complete control over when and how one of her experiment data products is made available to a broader community. The requirements needed to be met in order to bring about such a space are summarized as follows:

- <u>Publishing</u> – mechanisms for publishing a data product to the larger community. User has complete control over when product is published. Issues include publication scope, that is, multiple levels of visibility accorded a product; and mechanisms for discovery of information that has been made public.

- <u>Guarantees</u> – any data products that the user designates as "private" must be kept private. When a request for the product is issued, the product must exist. The guarantee of privacy extends beyond the metadata description to the logical data objects.

- Flexible schema – inherits highly general schema for representing collections, views, and logical files from Globus MCS, on which myLEAD is built. Spatial, temporal, and other geoscience extensions. Supports evolved understanding of data product over time by means of extended attributes,

- Immutable investigations – user can reuse collections, views, and logical files from earlier investigations without destroying the integrity of earlier investigation,

- Proactive agent – infers new metadata attributes from context of an active experiment using case-based reasoning. Agent represents user in negotiations with other web services such as workflow service.

7. WORKFLOW ORCHESTRATION AND THE LEAD PORTAL

Workflow orchestration, as defined in the LEAD project, involves the organization, scheduling and sequencing of the tasks that are involved in an experiment. The collection of tasks and their dependencies in a particular experiment is called a workflow. The process of orchestration involves defining, or composing, the workflow as well as its execution, the latter of which is referred to as the workflow enactment. Individual tasks in a workflow include extracting data from live streams or from archives, mining these streams to identify interesting events, transforming data objects from one form to another, preprocessing data, performing simulations, mining simulation output, rendering simulation output into movies or still images, and using simulation output to retarget instruments such as radars (Figure 7.1).

This particular set of tasks in LEAD will initially be carried out on our grid and web services test beds (Figure 7.2), and we will migrate elements of the global workflow to larger resources such as those provided by the NSF centers and TeraGrid. In a typical experimental weather prediction scenario, many simulations may need to be generated based upon the analysis of current atmospheric conditions. As the atmosphere evolves, some of

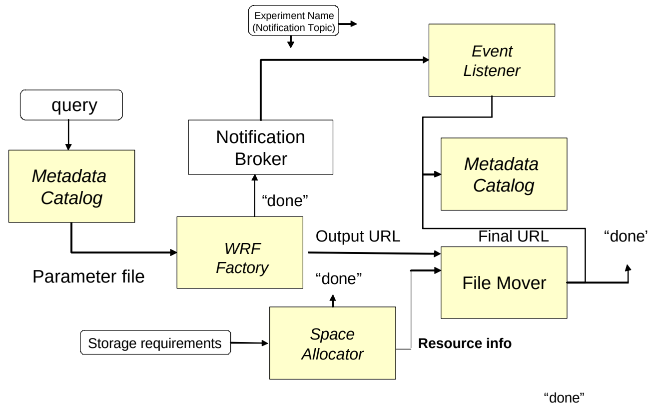

Figure 7.1. Sample LEAD tasks in the context of workflow orchestration.

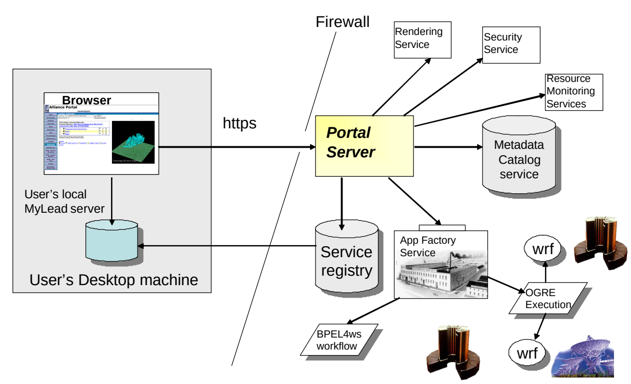

Figure 7.2. The global context of LEAD's testbed environment (which is the location for enactment of the workflow).

these simulations will diverge while others will accurately track the atmosphere's evolution (i.e., ensemble forecasting). It is possible that some simulations might be terminated and more resources directed toward those that are showing promise. Hence, the workflow must be very dynamic in its structure and enactment as it responds to real and simulated events, resource availability and human interaction. LEAD workflows also must manage computations and instruments that are distributed over large geographic regions. In addition, these workflow enactments must be very robust and quickly recover from failure or loss of resources.

LEAD presently is focusing on two approaches to workflow specification and enactment. In the first, the computational prototypes are primarily static in structure: a fixed set of tasks are chained together in a well-defined order and then executed without an ability to be modified during execution (Figure 7.3). For these simple workflows, we are using an extension of a Java-based build tool called the Open Grid computing environments Runtime Engine (OGRE). OGRE, which was developed at the National Center for Supercomputing Applications (NCSA), allows a user to "program" a workflow which may be launched by Globus on a

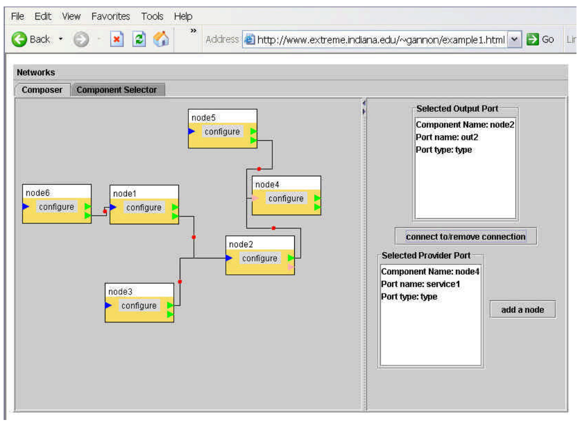

Figure 7.3. Web Applet graphical workflow composer used in LEAD. It allows a user to create a "network" of services (e.g., data feeds, decoders, visualization tools, models) that can then be executed (enacted).

remote host to manage a set of task instantiations and data movements using Grid FTP, HTTP, and other protocols. OGRE scripts also can publish execution events to a remote event channel so that the user can keep track of the state of the workflow enactment. OGRE utilizes a set of user interface tools (within which the capabilities of OGRE are integrated) to launch OGRE workflows and monitor their progress.

In our second application to workflow, we are using OGRE as a basis for integrating web service clients. With this capability, we gain a number of advantages including easy publication of automatically generated metadata to describe attributes to a data collection, as well as integration into user interfaces for easily generating user interface elements to provide access to a data collection.

Linked closely with workflow orchestration is the LEAD portal, which is a user-configurable web-based access point to web or other services, including data bases, that often lie behind a firewall (Figure 7.4).

The purpose of the LEAD portal is to provide a customizable user interface to enable mesoscale meteorology research and education. The portal capability must combine complex functionality, including streaming and archived data, distributed resources, advanced tools, remotely-controllable observing systems, and continually changing data formats and policies into in a simple user experience.

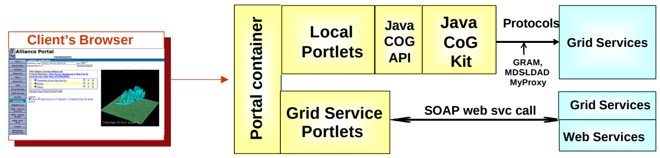

Figure 7.4. Basic LEAD portal architecture.

The portal maintains a persistent "context" for each "logged in" user; and this context is retained from session to session and contains credentials, such as identity certificates and service functionality, that are managed in a manner hidden from the user. Depending upon user capabilities, different levels of services may be accessed. In the context of LEAD, we have identified three principal (initial) users to be supported by the portal:

- Casual Browsers who wish to see content and exercise limited capabilities without establishing a portal login identity, e.g., access to pre-computed weather forecasts, current conditions, or general educational material on weather.

- Students (K-12 or university) taking a class that has been authorized to run certain simulations or data mining experiments. These applications will be wrapped together as web services and managed by a resource allocation system that automatically limits the resources available to students. The portal provides them with a user interface to these services and students interact with this interface in order to run the application and analyze data. In many cases, the application will consist of a pre-packaged workflow that allows the user to supply parameters and, when execution completes, provides them access to output. It is not assumed that these users have accounts on the back-end resources.

- Researchers who are developing simulations and workflow scenarios. The portal must provide them with easy mechanisms for deploying new applications in the form of services as well as easy-to-use tools for orchestrating workflows. It is assumed that these researchers have accounts on the back-end resources.

The portal work for LEAD focuses on two key areas. First, the construction of an "application factory" (Figure 7.2) that allows a LEAD user to convert any deployed application into a web service -- which itself may be accessed by a portal user or incorporated into a larger workflow. The LEAD researcher describes the application deployment and parameters (e.g., the URL of user supplied input files) in an XML document via an application generator tool. When this application descriptor is created, a web service is automatically generated that will allow another authorized user to execute it.

The second involves experiments in which users wish to integrate important meteorological graphical user interfaces into the portal framework and coupling them with web services. A geo-locater applet, now incorporated into the test portal, is an example of such an interface. It allows users to select a geographic region of interest and store it in the portal database. This region descriptor then can be use as an input to other services.

Finally, a variety of monitoring capabilities are being built into LEAD. Monitoring involves the creation of "virtual instruments" for tracking and estimating the performance of both applications and workflows as they are executing.

8. TECHNOLOGY ROAD MAP

As shown schematically in Figure 8.1, LEAD is evolving three distinct yet related generations of technology for community deployment. In *Generation 1*, workflows will be static, i.e., all tasks to be performed, including their order of execution, data dependencies, and computational resources, will be determined prior to job launch and cannot be changed until the job concludes. As Generation 1 is evolving, research in dynamic workflow for *Generation 2* will be underway (indicated as "Look-Ahead Research" in Figure 8.1). Note that Generation 1 will continue to exist throughout the project, though will become "frozen" in its capabilities by the end of 2006. In Generation 2, the early instantiation of which will become available late in 2006, workflows can be modified by the user during execution, or by the workflow itself, in response to any number of conditions (e.g., loss of data, identification of new features in output or observations, availability of computing resources). Further, on-demand capabilities will become available in Generation 2, requiring sophisticated monitoring and performance estimation resources (See Section 7).

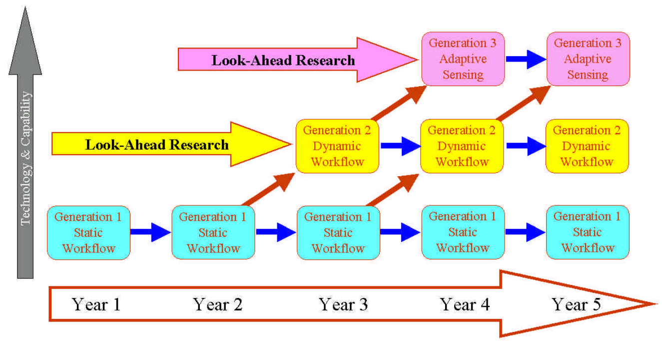

Figure 8.1. LEAD technology generations.

Generation 3 will provide the capability for meteorological tools to mutually interact with adaptive remote sensors, most notably the CASA Doppler weather radars – the first test bed of which will be located in Oklahoma and become available in late 2005 or early 2006. Preliminary work by LEAD in adaptive sensing will be performed using the CHILL radar system at Colorado State University, with look-ahead research beginning in year-2.

9. DEPLOYMENT

The field deployment of LEAD is being orchestrated via a phased approach involving a number of academic and research test beds and strategic partners in the context of three principal mechanisms. The first is the UCAR Unidata program that involves approximately 150 organizations encompassing 21,000 university students, 1800 faculty, and hundreds of operational practitioners. By integrating LEAD into its broad portfolio of resources, Unidata will be the focal point for bringing LEAD capabilities to the higher education community.

The second mechanism is the nascent Developmental Test Bed Center (DTC) at the National Center for Atmospheric Research. The DTC, sponsored by the NSF and NOAA and involving the Forecast Systems Laboratory, provides a national collaborative framework in which numerical weather analysis and prediction communities can interact to accelerate testing and development of new technologies as well as techniques for research applications and operational implementation – all in a way that mimics, but does not interfere with, actual forecast operations. The DTC is expected to become the focal point for bringing LEAD capabilities to the operations research community as well as to academic and other researchers on an international scale.

Finally, through its education test beds and learning communities (see Section 10), LEAD will provide to all levels of education, especially grades 6-12, classroom content that is consistent with national science education standards.

10. LEAD LEARNING COMMUNITIES

LEAD has established six Education Testbeds to coordinate and manage a three-phase plan designed to assess the effectiveness of LEAD technologies for research and education, to provide critical input and feedback to LEAD developers, and to facilitate knowledge transfer to a community of end users (educators, researchers and students).

Phase I establishes the education objectives that are helping shape the evolution and the user environment of LEAD and fuse the goals and enabling technologies into applications that are scalable and congruent with educational requirements, specifications, and standards. *Phase II* is linked to the flow of prototypes from the IT developers for evaluation and refinement, and is underway with the release of prototype 1a. *Phase III* will focus on deployment and integration of LEAD applications into the learning environment to incite curricular change and to drive innovation in meteorology and computer science, and to other disciplines through its inherent extensibility (e.g., oceanography, ecology).

A significant component of the LEAD education and outreach program are the so-called LEAD Learning Communities (LLC). The LEAD Learning Communities comprise a collaborative network of teachers, researchers, and students who interact to address a variety of issues related to LEAD that are common to their community. The principal driver in the development of LEAD Learning Communities is to enable and foster a two-way interaction between the LEAD developers and LEAD users to maximize the overall effectiveness of LEAD. Each learning community will 1) Integrate LEAD applications into their domain of activity, 2) develop new applications, 3) create performance outcomes for both education and research, 4) provide feedback to LEAD developers so that user needs drive the design and development process, and 5) determine potential relevance of LEAD technologies to other related fields and initiatives.

In light of the many ways in which students, teachers, and researchers might be clustered together, we have chosen to develop two LEAD Learning Communities: 1. Teaching and Learning Community and 2. Research and Operations Community. The Teaching and Learning Community will focus on activities related to teaching and learning that are important to all levels of the education enterprise ranging from middle school to graduate school. The idea for the LEAD Teaching and Learning Community leverages on the new paradigm of teaching and learning where students are engaged in the learning process by constructing their own knowledge. The activities of this community include but are not limited to the deployment and integration of LEAD applications as they are developed into the various learning environments, and to the evaluation and assessment of these applications. The assessment and evaluation process will be used to generate recommendations that will be communicated to the developers to drive further modification and refinement. The purpose of the LEAD Research and Operations Community is to focus on issues related to meteorology and computer science research and applications, particularly those requiring advanced cyberinfrastructure capabilities that otherwise are not available. This learning community will foster an interaction between the basic and applied research communities utilizing LEAD capabilities and LEAD developers. Furthermore, LEAD Learning Community members will serve as liaisons to the broader communities of education and research, and the two communities form a strand that will share common themes to better capitalize on new opportunities created by data accessibility and Web Services technology.

11. ACKNOWLEDGMENTS

LEAD is funded by the National Science Foundation under the following Cooperative Agreements: ATM-0331594 (Oklahoma), ATM-0331591 (Colorado State), ATM-0331574 (Millersville), ATM-0331480 (Indiana), ATM-

0331579 (Alabama in Huntsville), ATM03-31586 (Howard), ATM-0331587 (UCAR), and ATM-0331578 (Illinois at Urbana-Champaign). The first and second authors also were supported by NSF Cooperative Agreement EEC-0313747 to the University of Massachusetts at Amherst, which funds the Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere.

12. REFERENCES

Droegemeier, K.K. and Co-Authors, 2004: Linked environments for atmospheric discovery (LEAD): A cyberinfrastructure for mesoscale meteorology research and education. Preprints, *20th. Conf. on Interactive Info. Processing Systems for Meteor, Oceanography, and Hydrology*, Seattle, WA, Amer. Meteor. Soc. (http://lead.ou.edu/pdfs/IIPS LEAD AMS Preprint.pdf)

McLaughlin, D.J. and Co-Authors, 2005: Distributed Collaborative Adaptive Sensing (DCAS) for improved detection, understanding and prediction of atmospheric hazards. Preprints, Ninth Symp. On Integrated Obs. Systems for the Atmos., Ocean and Land Surface, San Diego, CA, Amer. Meteor. Soc.