Linked Environments for Atmospheric Discovery (LEAD): An Overview

17 November, 2003
Boulder, CO

Mohan Ramamurthy
mohan@ucar.edu

Unidata Program Center
UCAR Office of Programs
Boulder, CO

LEAD is Funded by the National Science Foundation
Cooperative Agreement:ATM-0331587

The 2002-2003 Large ITR Competition: Facts & Figures

- 67 pre-proposals submitted; 35 invited for full submissions
- 8 projects were funded;
- LEAD is the first Atmospheric Sciences project to be funded in the large-ITR category
  - LEAD Total Funding: $11.25M over 5 years

LEAD Institutions

K. Droegemeier, PI

University of Oklahoma
(K. Droegemeier, PI)
Meteorological Research and Project Coordination

University of Alabama in Huntsville
(S. Graves, PI)
Data Mining, Interchange Technologies, Semantics

UCAR/Unidata
(M. Ramamurthy, PI)
Data Streaming and Distributed Storage

Indiana University
(D. Gannon, PI)
Data Workflow, Orchestration, Web Services

University of Illinois/NCSA
(R. Wilhelmson, PI)
Monitoring and Data Management

Millersville University
(R. Clark, PI)
Education and Outreach

Howard University
(E. Joseph, PI)
Meteorological Research Education and Outreach

Colorado State University
(Chandra, PI)
Instrument Steering, Dynamic Updating

unidata

Motivation for LEAD

Each year, mesoscale weather – floods, tornadoes, hail, strong winds, lightning, hurricanes and winter storms – causes hundreds of deaths, routinely disrupts transportation and commerce, and results in annual economic losses in excess of $13B.

The Roadblock

- The study of events responsible for these losses is stifled by rigid information technology frameworks that cannot accommodate the
  - real time, on-demand, and dynamically-adaptive needs of mesoscale weather research;
  - its disparate, high volume data sets and streams;
  - its tremendous computational demands, which are among the greatest in all areas of science and engineering
- Some illustrative examples…

Cyclic Tornadogenesis Study

Adlerman and Droegemeier (2003)

- A parameter sensitivity study
- Generated 70 simulations, all analyzed by hand

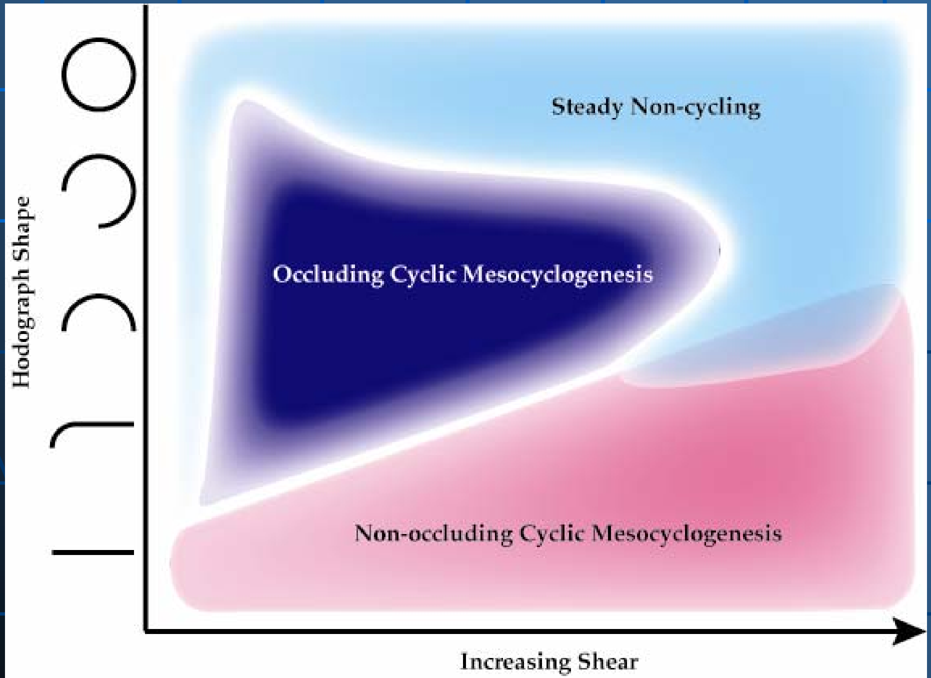
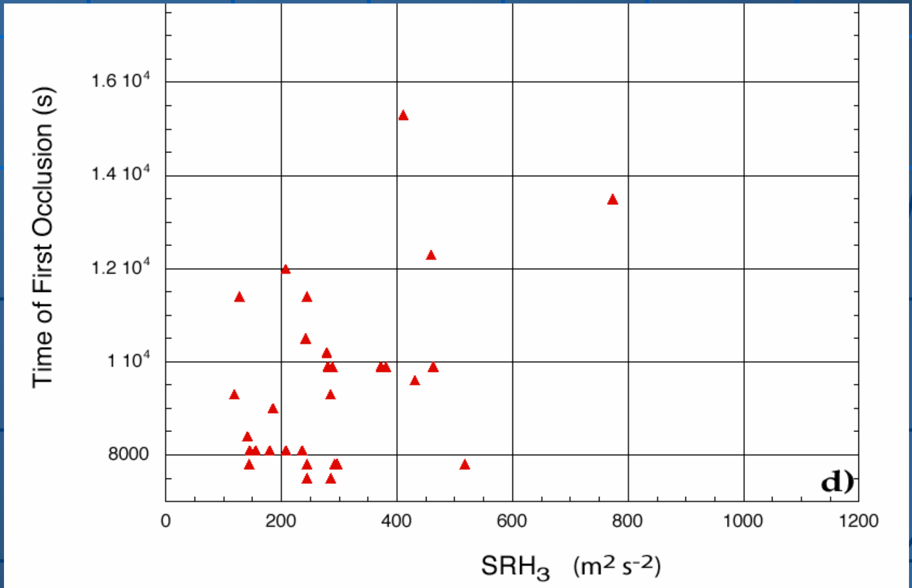
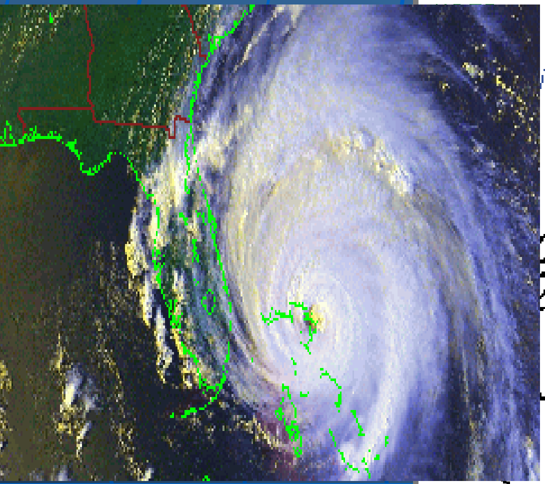


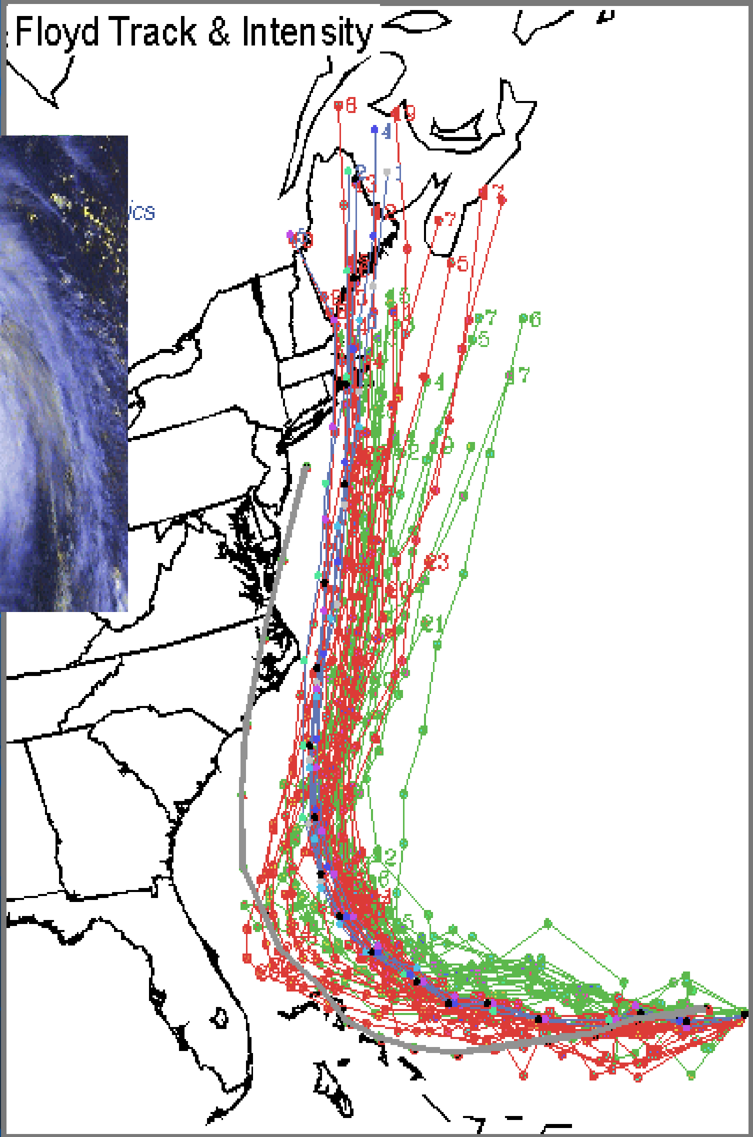
Figure 3.10.1: Summary of cycling behavior for average shear versus hodograph shape .

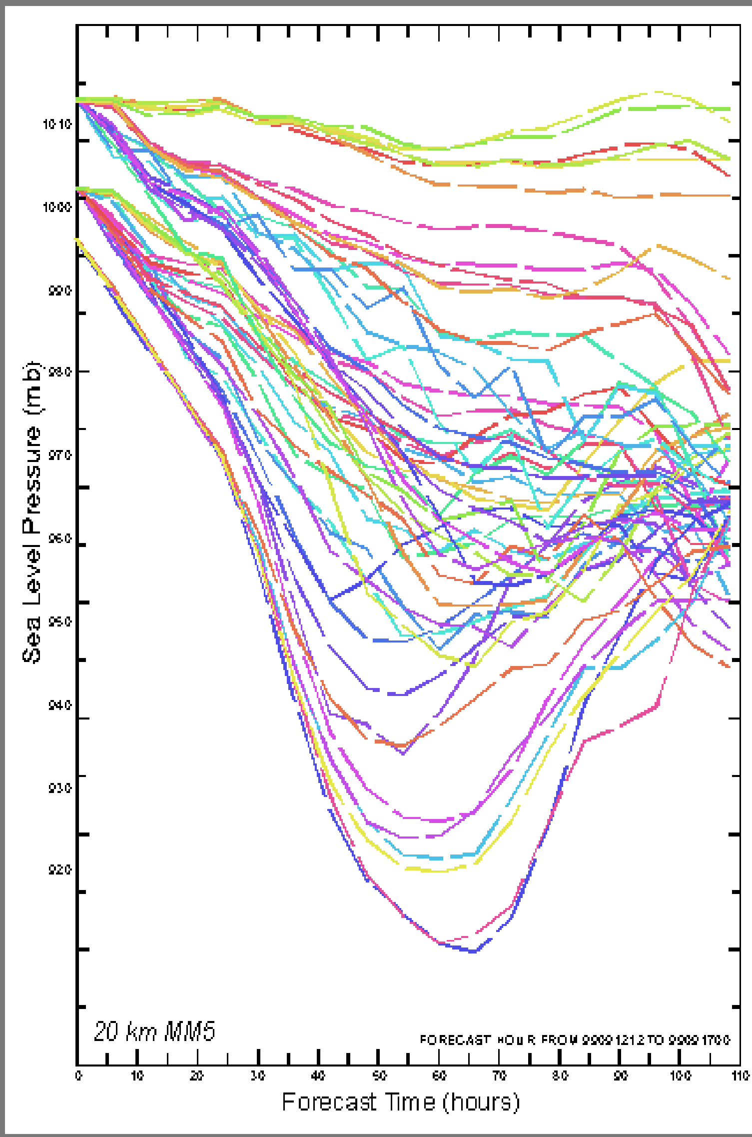


Hurricane Ensembles

Jewett and Ramamurthy (2003)





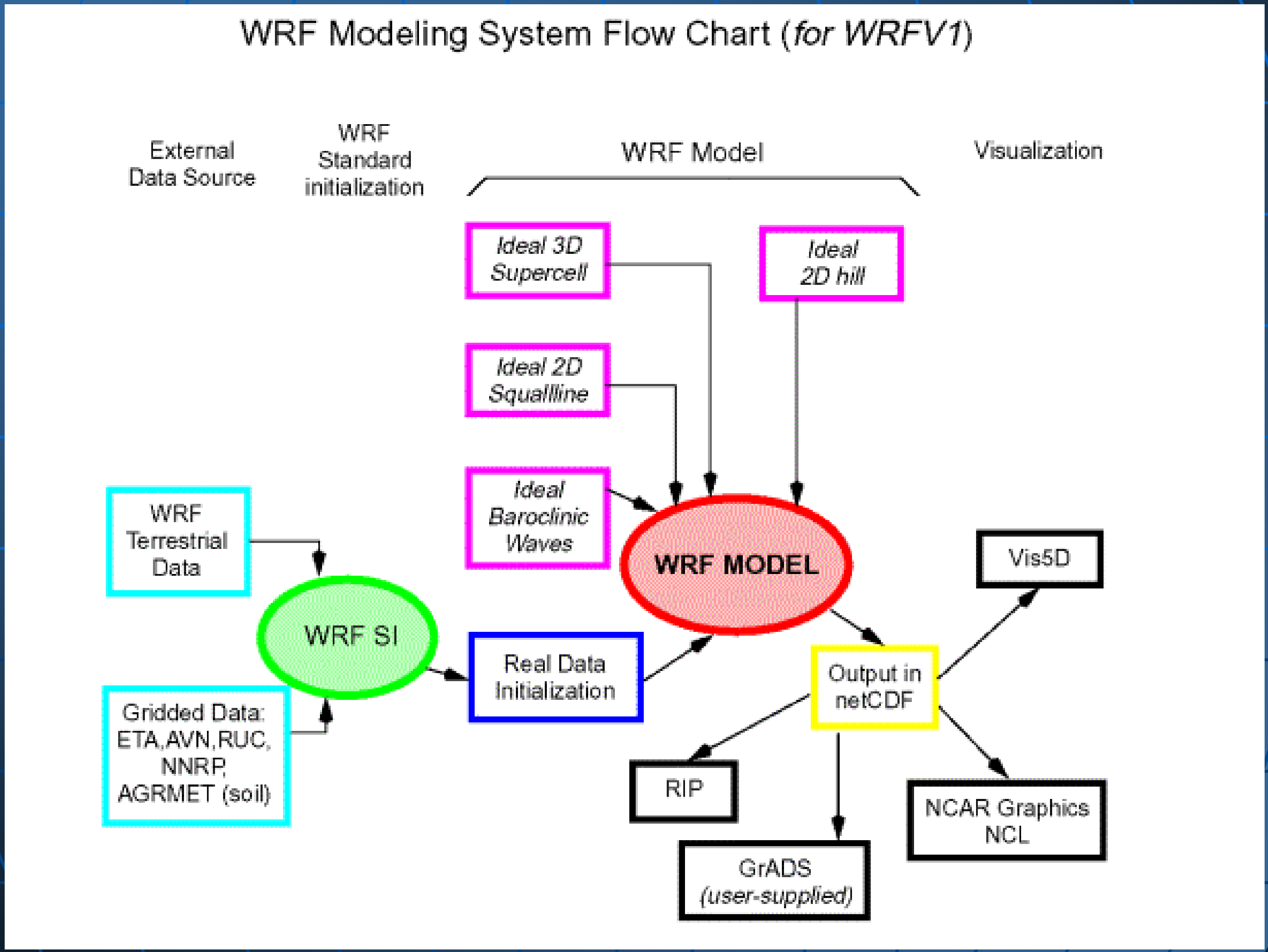


Local Modeling in the Community

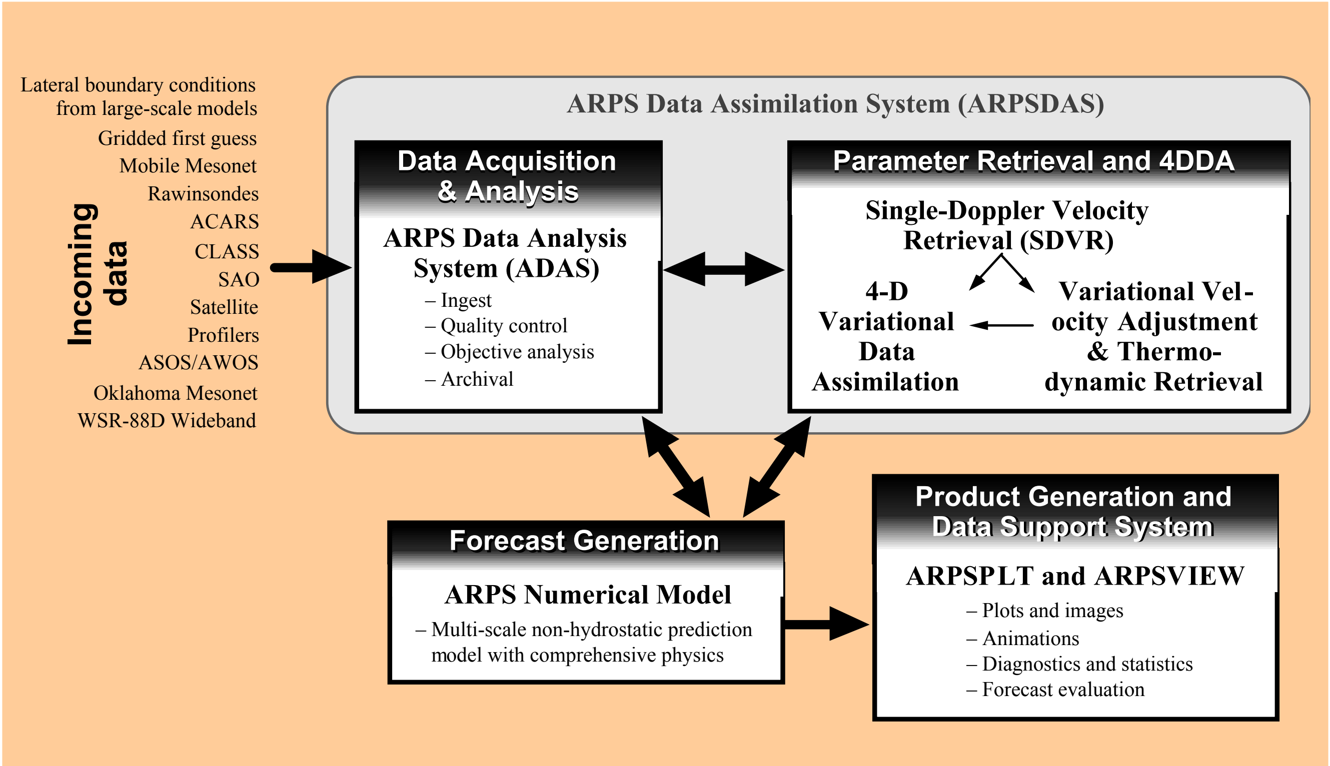
- Mesoscale forecast models are being run by universities, in real time, at dozens of sites around the country, often in collaboration with local NWS offices
  - Tremendous value
  - Leading to the notion of "distributed" NWP
- Yet only a few (OU, U of Utah) are actually assimilating local observations – which is one of the fundamental reasons for such models!

•Applied Modeling Inc. (Vietnam) MM5
•Atmospheric and Environmental Research MM5
•Colorado State University RAMS
•Florida Division of Forestry MM5
•Geophysical Institute of Peru MM5
•Hong Kong University of Science and Technology MM5
•IMTA/SMN, Mexico MM5
•India's NCMRWF MM5
•Iowa State University MM5
•Jackson State University MM5
•Korea Meteorological Administration MM5
•Maui High Performance Computing Center MM5
•MESO, Inc. MM5
•Mexico / CCA-UNAM MM5
•NASA/MSFC Global Hydrology and Climate Center, Huntsville, AL MM5
•National Observatory of AthensMM5
•Naval Postgraduate School MM5
•Naval Research Laboratory COAMPS
•National Taiwan Normal University MM5
•NOAA Air Resources Laboratory RAMS
•NOAA Forecast Systems Laboratory LAPS, MM5, RAMS
•NCAR/MMM MM5
•North Carolina State University MASS
•Environmental Modeling Center of MCNC MM5 MM5
•NSSL MM5
•NWS-BGM MM5
•NWS-BUF (COMET) MM5
•NWS-CTP (Penn State) MM5
•NWS-LBB RAMS
•Ohio State University MM5
•Penn State University MM5
•Penn State University MM5 Tropical Prediction System
•RED IBERICA MM5 (Consortium of Iberic modelers) MM5 (click on Aplicaciones)
•Saint Louis University MASS
•State University of New York - Stony Brook MM5
•Taiwan Civil Aeronautics AdministrationMM5
•Texas A\&M UniversityMM5
•Technical University of MadridMM5
•United States Air Force, Air Force Weather Agency MM5
•University of L'Aquila MM5
•University of Alaska MM5
•University of Arizona / NWS-TUS MM5
•University of British Columbia UW-NMS/MC2
•University of California, Santa Barbara MM5
•Universidad de Chile, Department of Geophysics MM5
•University of Hawaii MM5
•University of Hawaii RSM
•University of Hawaii MM5
•University of Illinois MM5, workstation Eta, RSM, and WRF
•University of Maryland MM5
•University of Northern Iowa Eta
•**University of Oklahoma/CAPS** ARPS
•University of Utah MM5
•**University of Washington** MM5 36km, 12km, 4km
•University of Wisconsin-Madison UW-NMS
•University of Wisconsin-Madison MM5
•University of Wisconsin-Milwaukee MM5

Current WRF Capability



The Prediction Process: Current Situation



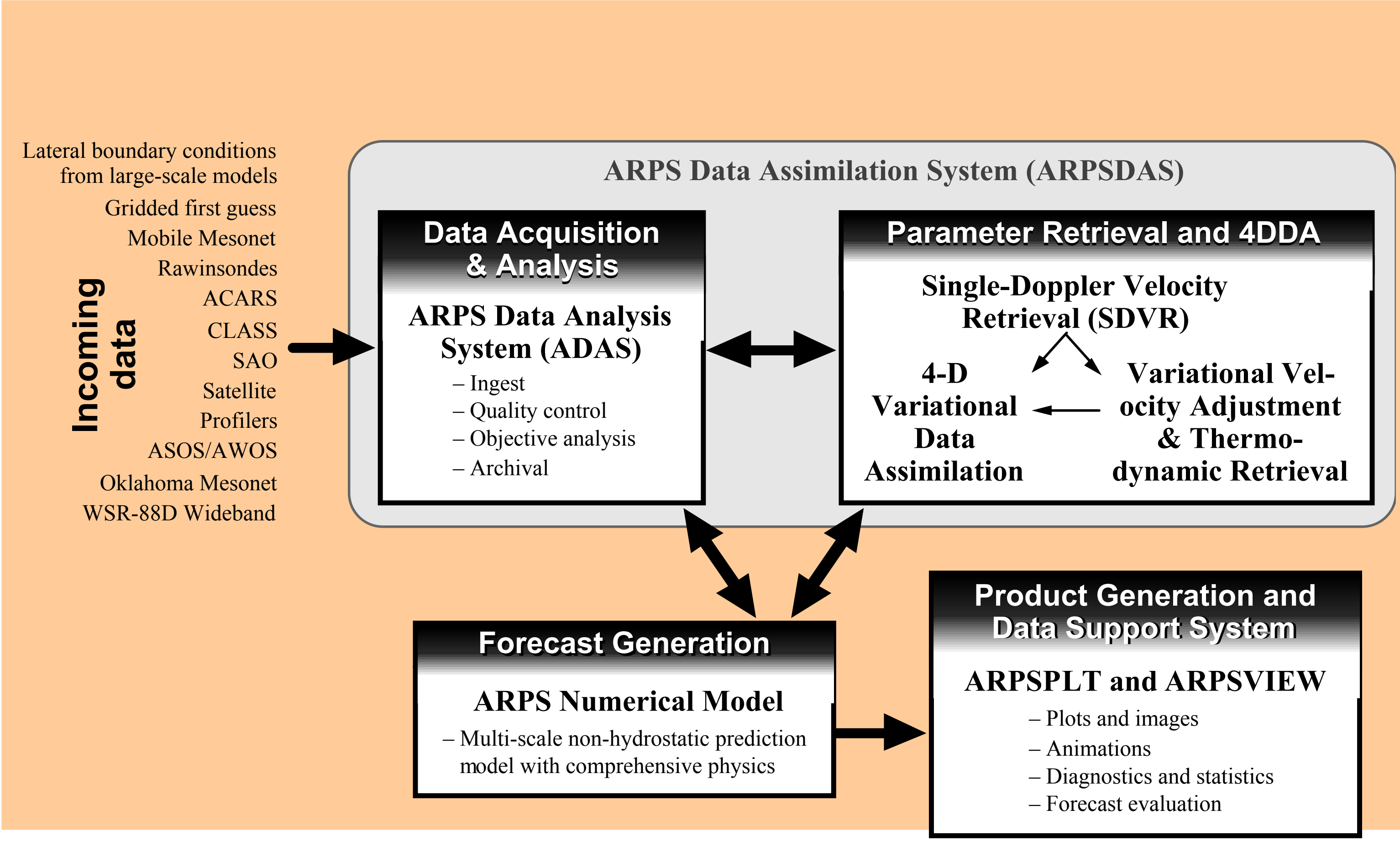
This process is very time-consuming, inefficient, tedious, does not port well, does not scale well, etc.

As a result, a scientist typically spends over 70% of his/her time with data processing and less than 30% of time doing research.

The LEAD Goal

- To create an end-to-end, integrated, flexible, scalable framework for…
  - Identifying
  - Accessing
  - Preparing
  - Assimilating
  - Predicting
  - Managing
  - Mining
  - Visualizing
- …a broad array of meteorological data and model output, independent of format and physical location

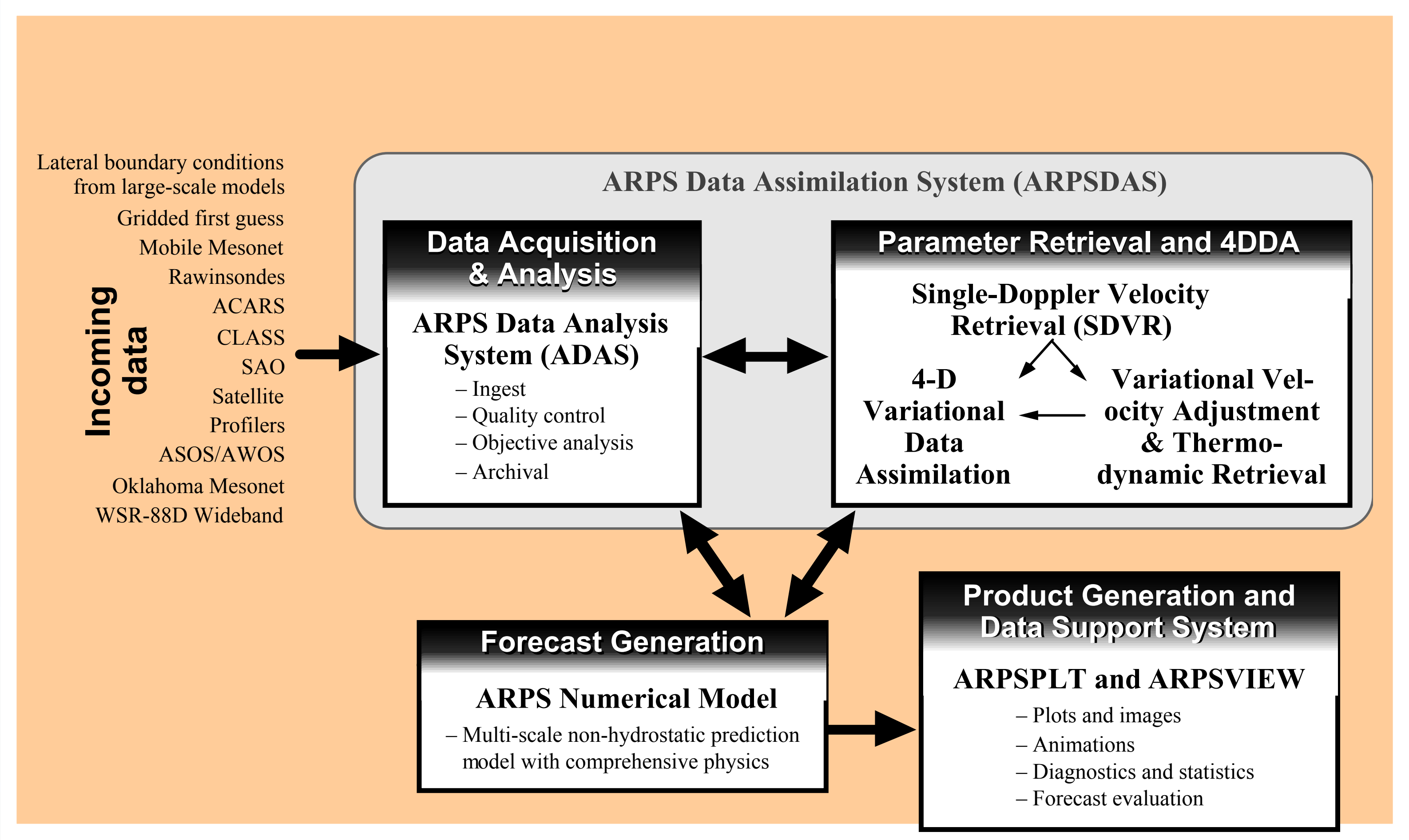
The Prediction Process



How do we turn the above prediction process into a sequence of chained Grid and Web services?

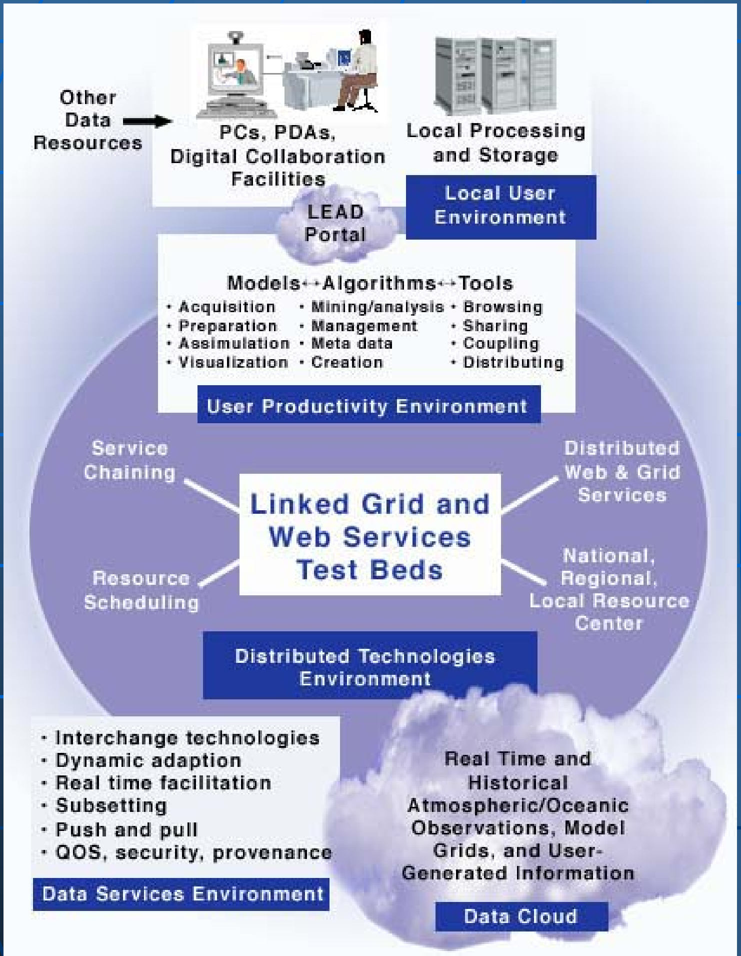
The modeling community HAS TO DATE **NOT** looked at this process from a Web/Grid Services perspective

The Prediction Process - continued



Key Issues: Real-time vs. on-demand vs. retrospective predictions – what differences will there be in the implementation of the above sequence?

LEAD Testbeds and Elements



- Portal
- Data Cloud
- Data distribution/streaming
- Interchange Technologies (ESML)
- Semantics
- Data Mining
- Cataloging
- Algorithms
- Workflow orchestration
- MyLEAD
- Visualization
- Assimilation
- Models
- Monitoring
- Steering
- Allocation
- Education

LEAD Testbeds at UCAR, UIUC, OU, UAH & IU

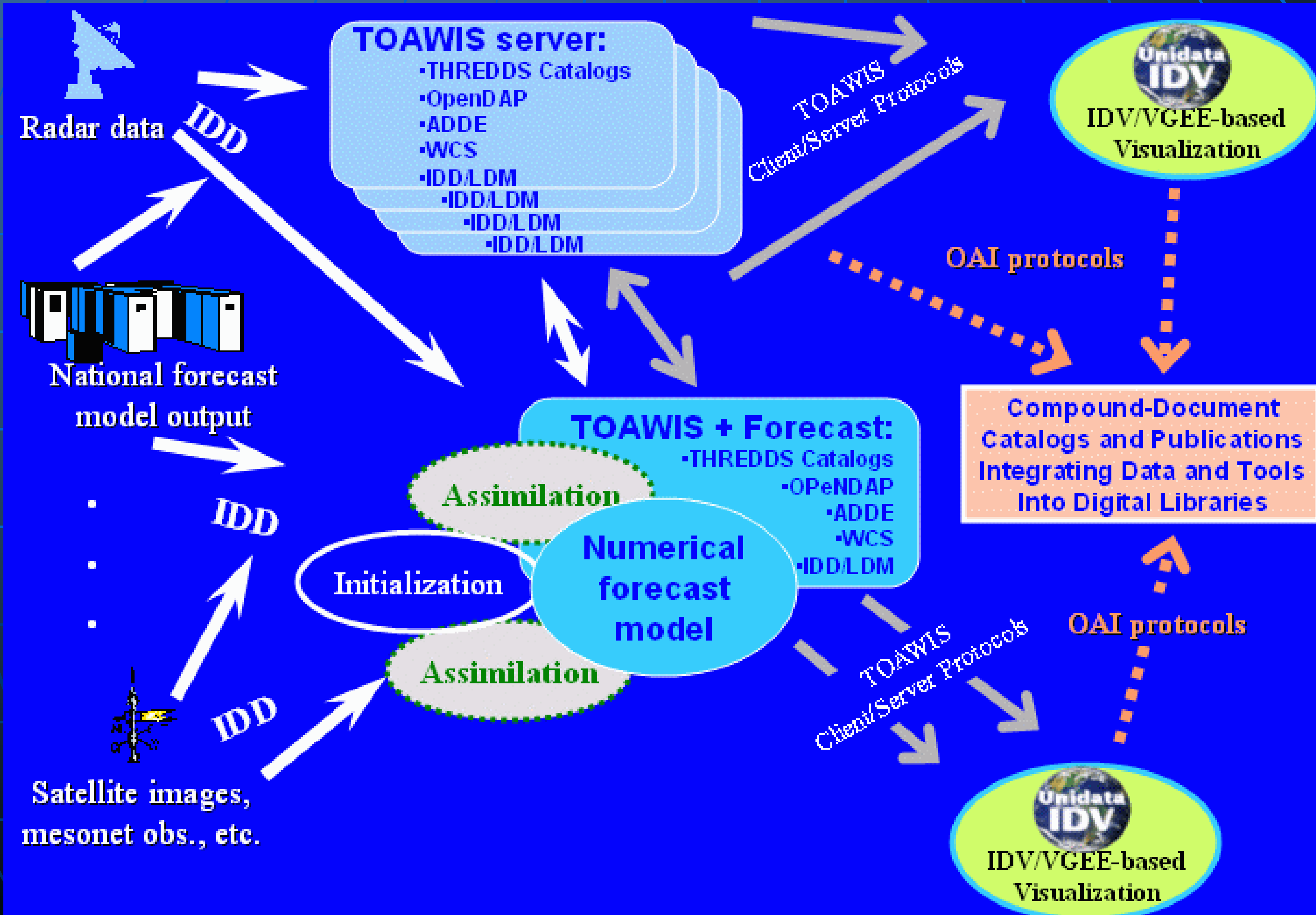
So What's Unique About LEAD?

- Allows the use of analysis and assimilation tools, forecast models, and data repositories as *dynamically adaptive, on-demand* services that can
  - change configuration rapidly and automatically in response to weather;
  - continually be steered by unfolding weather;
  - respond to decision-driven inputs from users;
  - initiate other processes automatically; and
  - steer remote observing technologies to optimize data collection for the problem at hand.

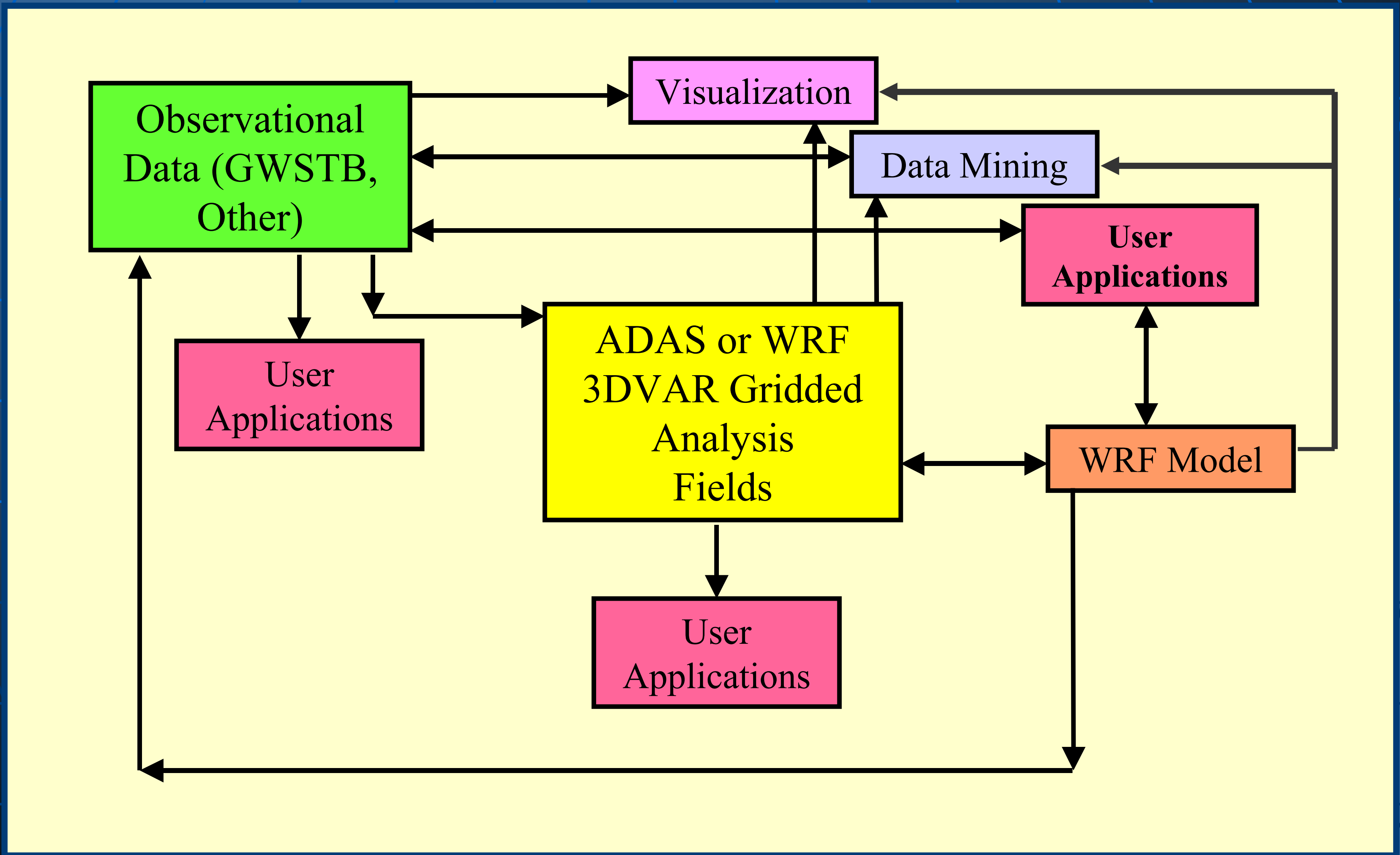
When You Boil it all Down…

- The underpinnings of LEAD are
  - On-demand
  - Real time
  - Automated/intelligent sequential tasking
  - Resource prediction/scheduling
  - Fault tolerance
  - Dynamic interaction
  - Interoperability
  - Linked Grid and Web services
  - Personal virtual spaces (myLEAD)

Testbed Services: An Example



Lead User Scenario: An Example



Web Services

- *They are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web.*
- The XML based Web Services are emerging as tools for creating next generation distributed systems that are expected to facilitate program-to-program interaction without the user-to-program interaction.
- Besides recognizing the heterogeneity as a fundamental ingredient, these web services, independent of platform and environment, can be packaged and published on the internet as they can communicate with other systems using the common protocols.

Web Services Four-wheel Drive

- **WSDL** (Creates and Publishes)
  - Web Services Description Language
  - WSDL describes what a web service can do, where it resides, and how to invoke it.
- **UDDI** (Finds)
  - Universal Description, Discovery and Integration
  - UDDI is a registry (like yellow pages) for connecting producers and consumers of web services.
- **SOAP** (Executes remote objects)
  - Simple Object Access Protocol
  - Allows the access of Simple Object over the Web.
- **BPEL4WS** (Orchestrates – Choreographer)
  - Business Process Execution Language for Web Services.
  - It allows you to create complex processes by wiring together different activities that can perform Web services invocations, manipulate data, throw faults, or terminate a process.

The Grid

- *Refers to an infrastructure that enables the integrated, collaborative use of computers, networks, databases, and scientific instruments owned and managed by distributed organizations.*
- The terminology originates from analogy to the electrical power grid; most users do not care about the details of electrical power generation, distribution, etc.
- *Grid applications often involve large amounts of data and/or computing and often require secure resource sharing across organizational boundaries.*
- Grid services are essentially web services running in a Grid framework.

TeraGrid: A $90M NSF Facility

Capacity:

20 Teraflops

1 Petabyte of disk-storage

Connected by 40GB network

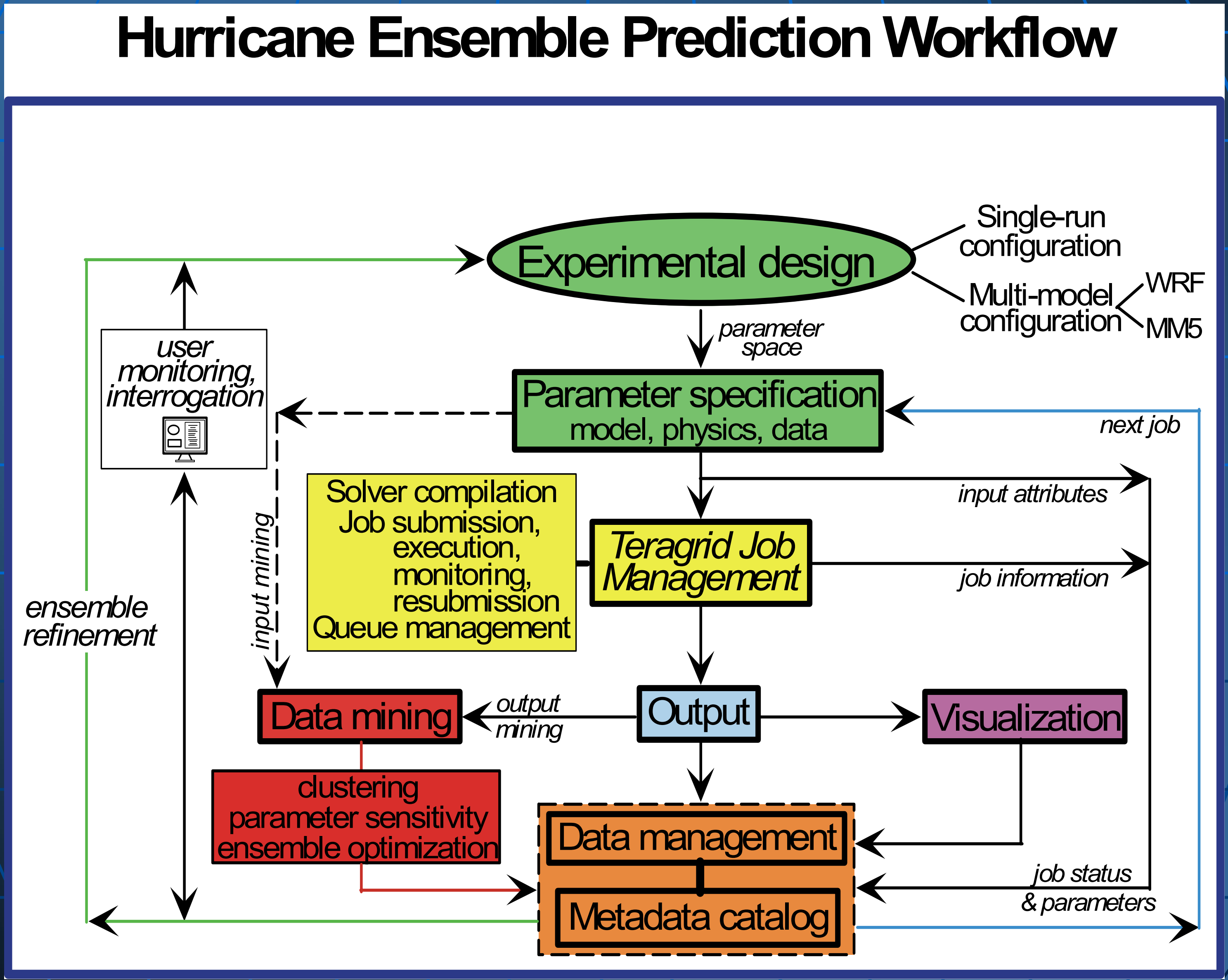
The LEAD Grid Testbed facilities will be on a bit more modest scale!

NSF Recently funded three more institutions to connect to the above Grid

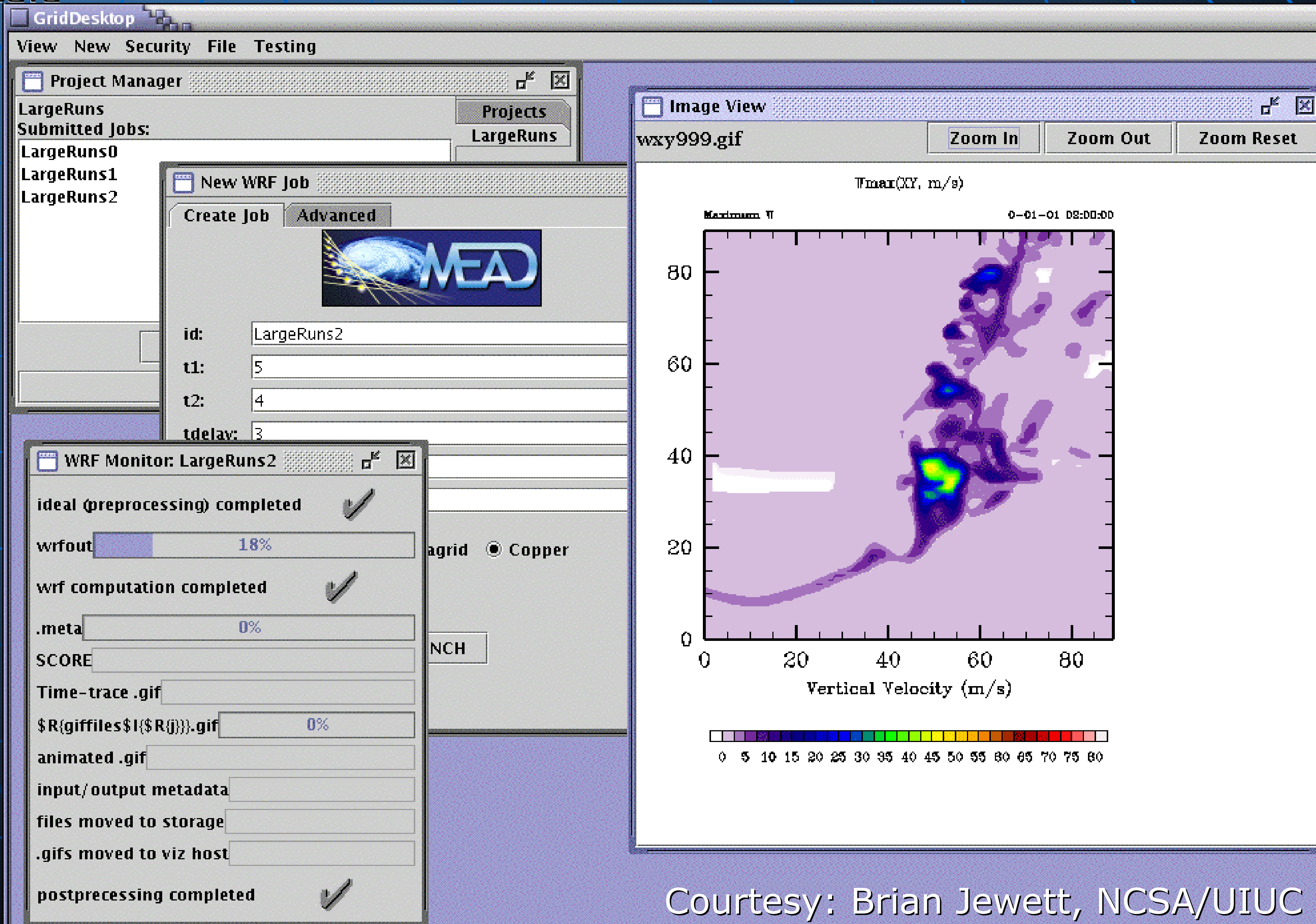
Globus

- A project that is investigating how to build infrastructure for Grid computing
- Has developed an integrated toolkit for Grid services
- Globus services include :
  - Resource allocation and process management
  - Communication services
  - Distributed access to structure and state information
  - Authentication and security services
  - System monitoring
  - Remote data access
  - Construction, caching and location of executables

Workflow Orchestration

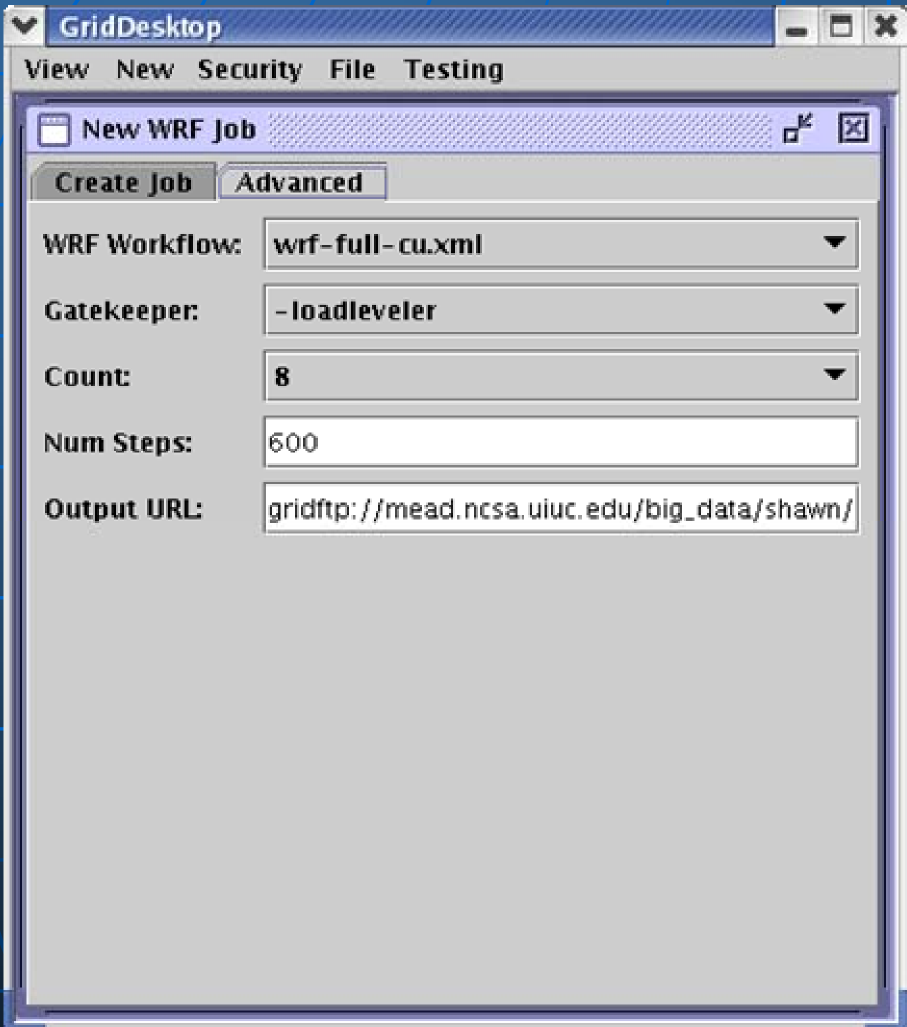


Workflow applied to storm modeling



Courtesy: Brian Jewett, NCSA/UIUC

Components of the Workflow

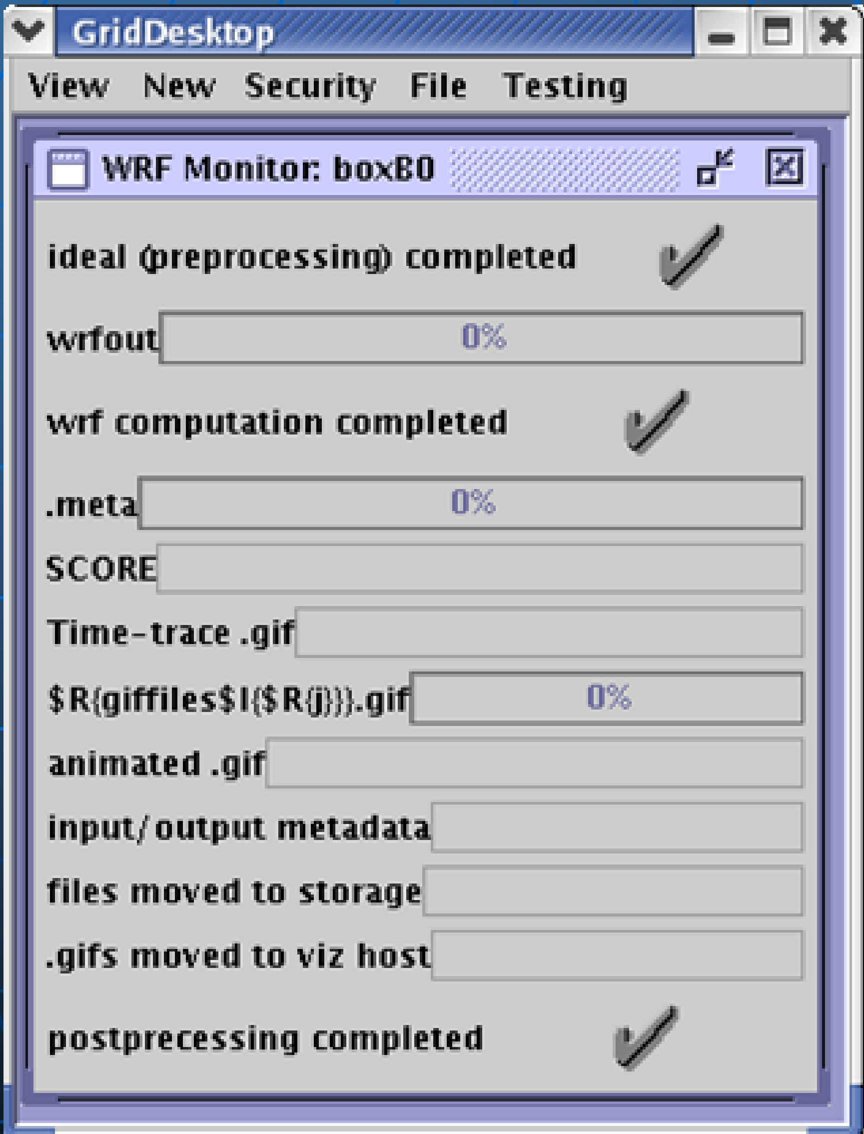


Job Launcher

- Specify platform
- Specify job parameters
- Run ID
- Initial storm cell
  - magnitude (temperature)
  - position
  - initiation time
- Additional options, including run length, time steps, etc.

Courtesy S. Hampton, A. Rossi / NCSA

Components of the Workflow



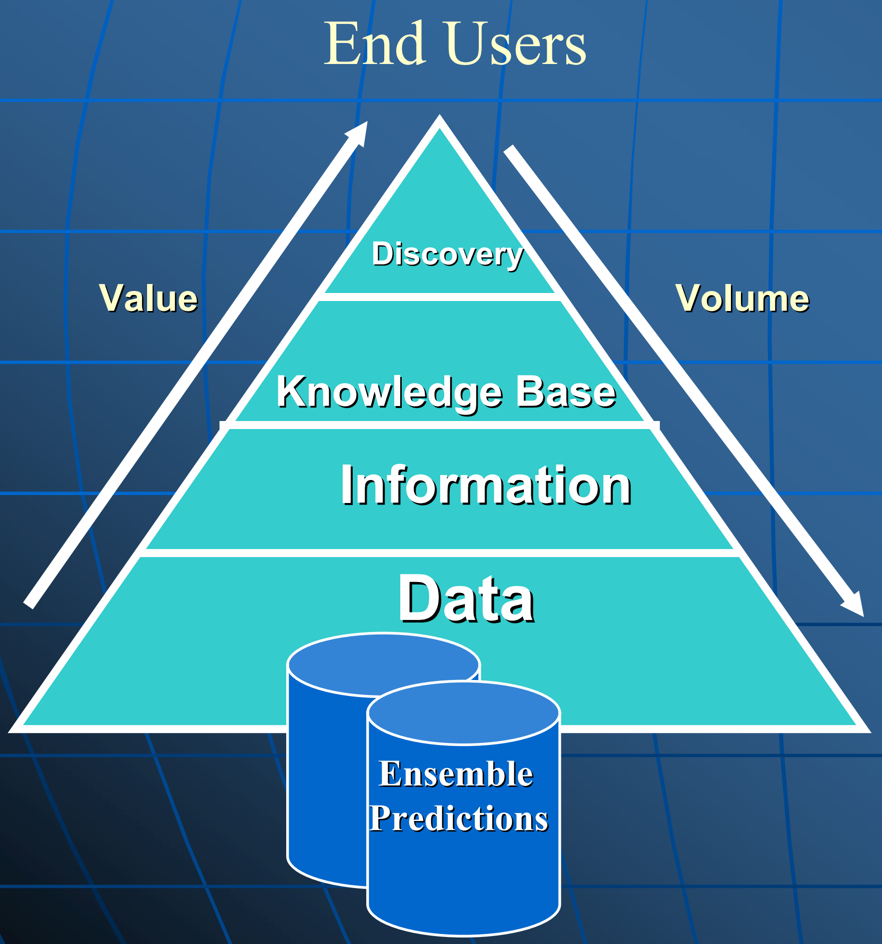
WRF Monitor

Shows state of remote job -

- Pre-processing
- WRF code execution
- Post-processing, including
  - Image (2D) generation
  - Scoring (statistics)
  - Time series data & plots
- Archival to mass store

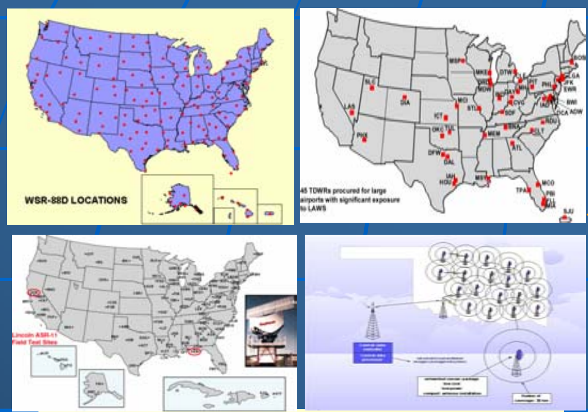
Courtesy S. Hampton, A. Rossi / NCSA

Data Mining and Knowledge Discovery



- In a world awash with data, we are starving for knowledge.
  - E.g., ensemble predictions
- Need scientific data mining approaches to knowledge management
- Key: Leveraging data to make BETTER decisions

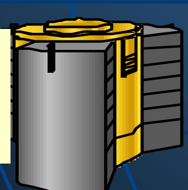
Mining/Detection in LEAD



NEXRAD, TDWR, FAA, NETRAD Radars

Other Observations

Forecast Model Output



Data Assimilation System

High-Resolution, Physically Consistent Gridded Fields of all Meteorological Variables

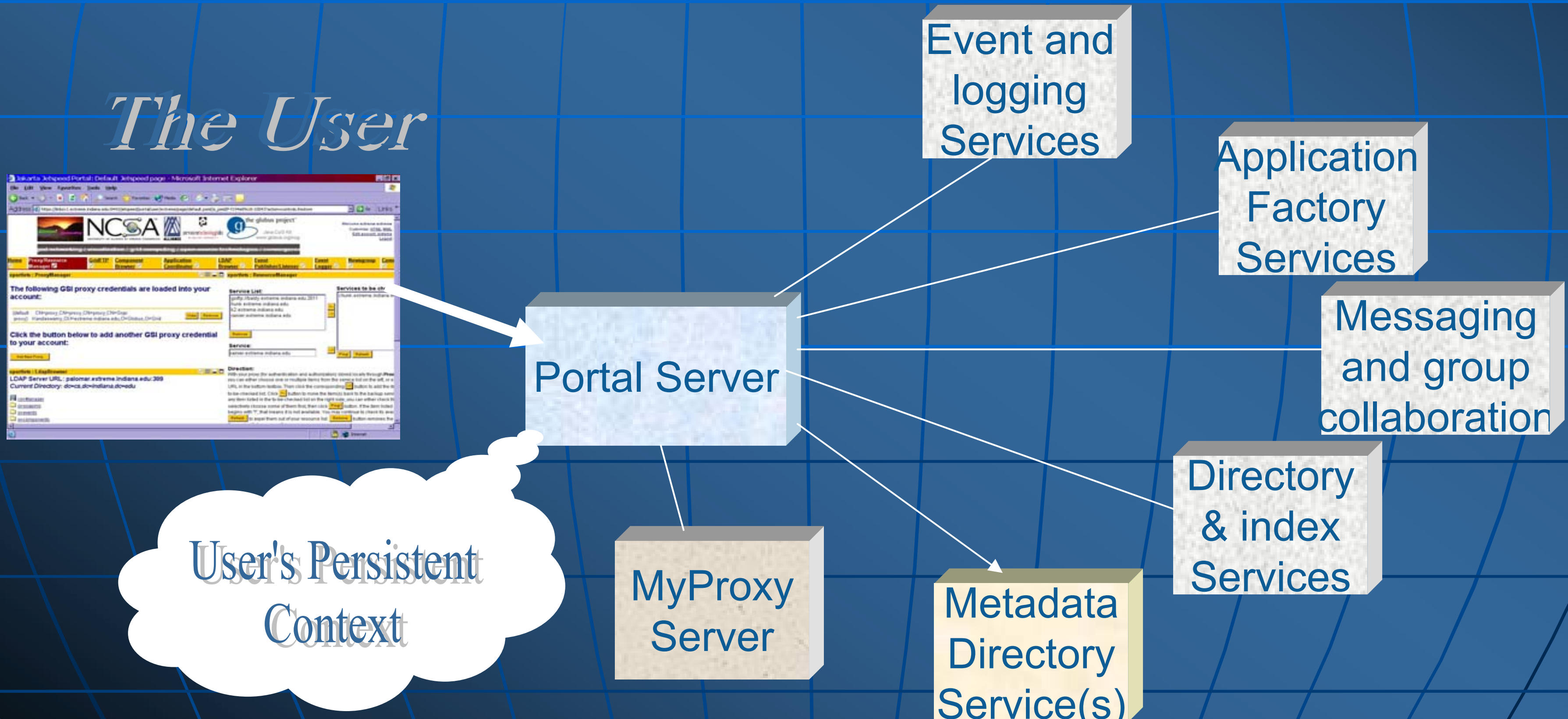
Forecast Models

Data Mining Engines

Features and Relationships

LEAD Portal: The Big Picture

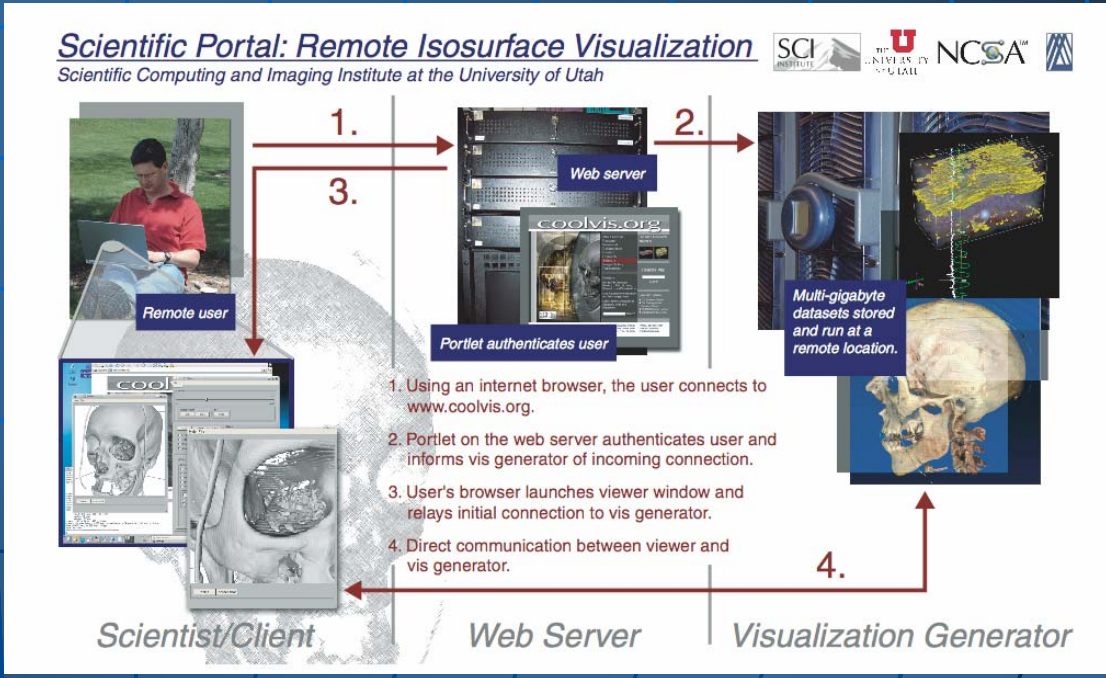
- The portal is the user's entry point to Grid and Web services and their orchestration



Courtesy: Dennis Gannon, IU

LEAD Portal: Basic Elements

- Management of user proxy certificates
- Remote file transport via GridFTP
- News/Message systems for collaborations
- Event/Logging service
- Personal directory of services, metadata and annotations.
- Access to LDAP services
- Link to specialized application factories
- Tool for performance testing
- Shared collaboration tools
  - Including shared Powerpoint
- Access and control of desktop Access Grid



Courtesy: Dennis Gannon, IU

Synergy with Other Grid and Non-Grid Projects

- LEAD will leverage, where possible, tools, technologies and services developed by many other ATM projects, including
  - Earth System Grid
  - MEAD
  - NASA Information Power Grid
  - WRF, ARPS/ADAS,…
  - OPeNDAP
  - THREDDS
  - MADIS
  - NOMADS
  - CRAFT
  - VGEE
  - And other projects…

LEAD Contact Information

- LEAD PI: Prof. Kelvin Droegemeier, kkd@ou.edu
- LEAD/UCAR PI: Mohan Ramamurthy, mohan@ucar.edu
- Project Coordinator: Terri Leyton, tleyton@ou.edu

http://lead.ou.edu/