
Cyberinfrastructure for Data Driven Science Research and Education: Application to Mesoscale Meteorology

Beth Plale and Dennis Gannon
Department of Computer Science
School of Informatics
Indiana University



Linked Environments for Atmospheric Discovery



The LEAD Project



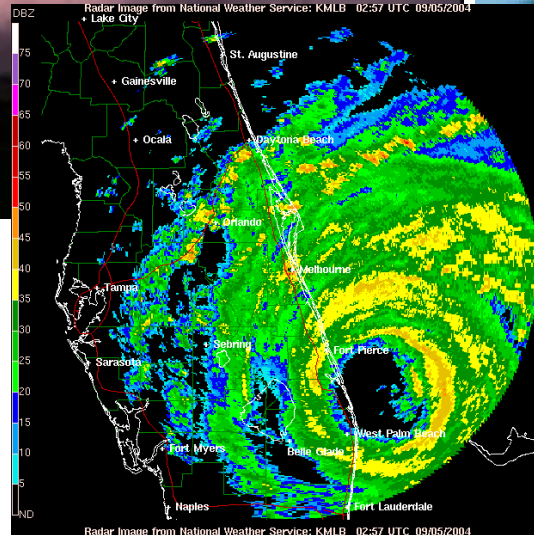
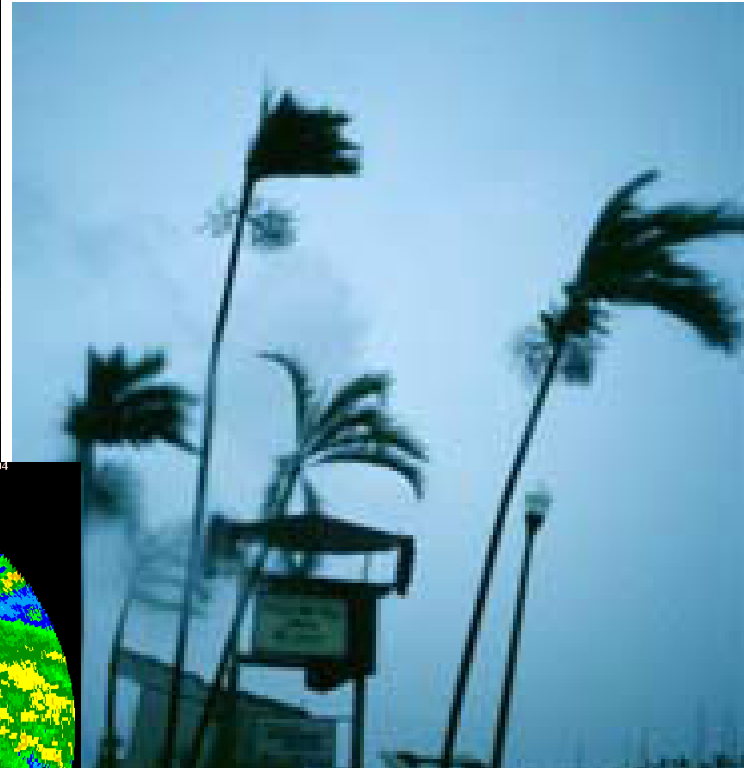
L I N K E D
E N V I R O N M E N T S
F O R A T M O S P H E R I C
D I S C O V E R Y



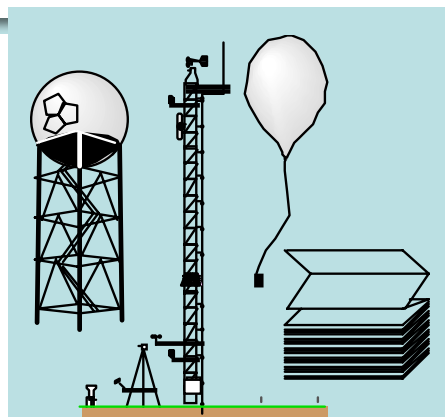
Linked **E**nvironments for **A**tmospheric **D**iscovery



What is Mesoscale weather?

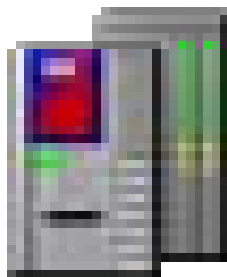


Traditional Methodology



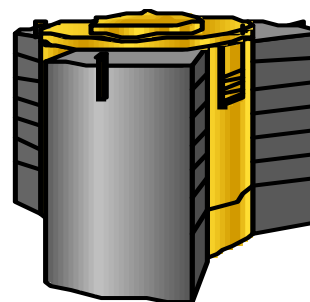
STATIC OBSERVATIONS

Radar Data
Mobile Mesonets
Surface Observations
Upper-Air Balloons
Commercial Aircraft
Geostationary and Polar Orbiting Satellite
Wind Profilers
GPS Satellites



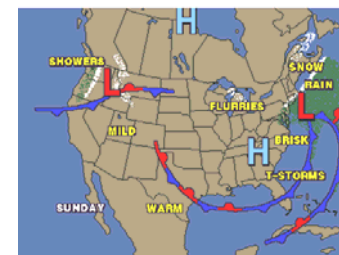
Analysis/Assimilation

Quality Control
Retrieval of Unobserved Quantities
Creation of Gridded Fields



Prediction/Detection

PCs to Teraflop Systems



Product Generation, Display, Dissemination



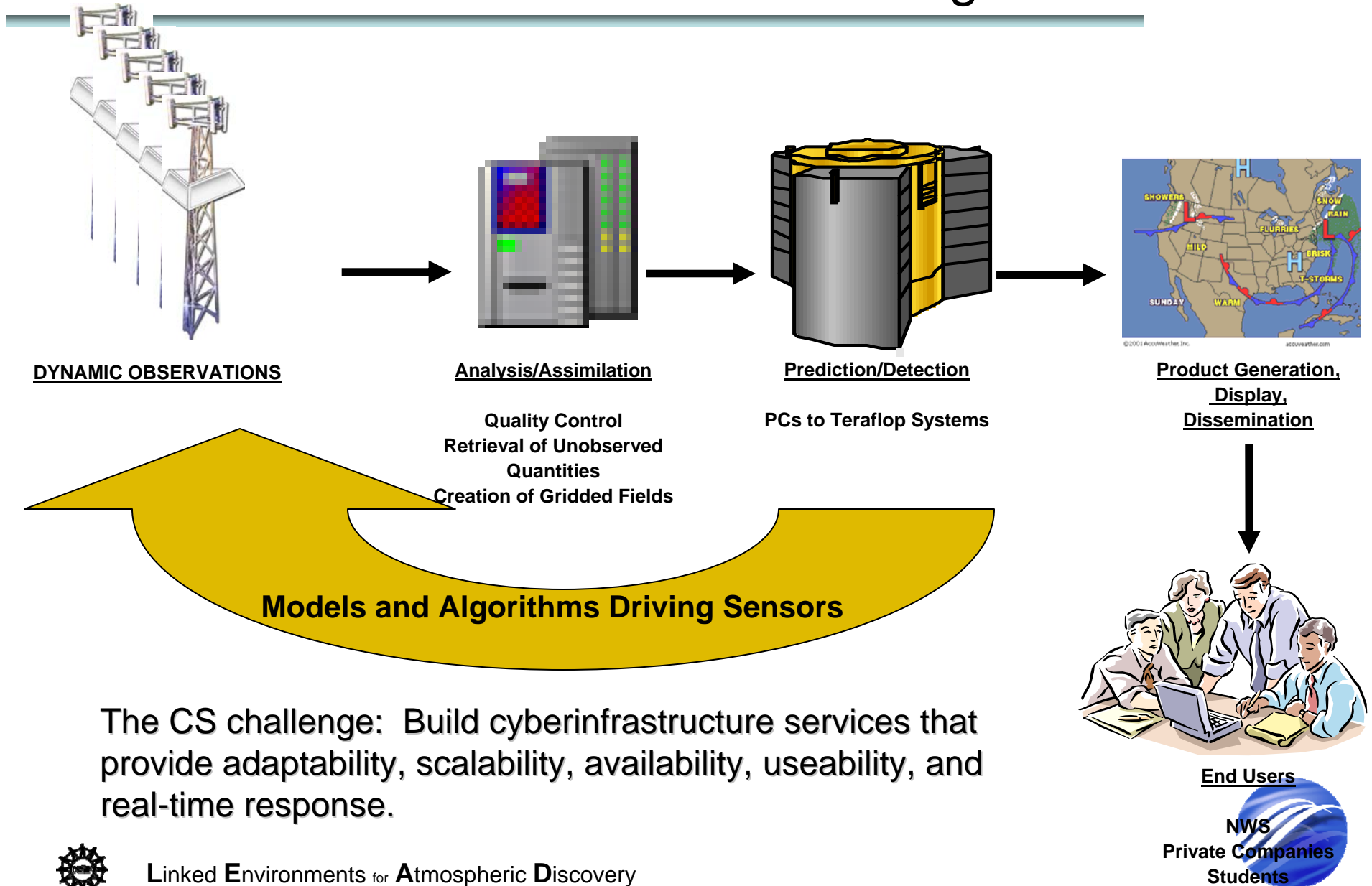
End Users

NWS
Private Companies
Students

**The Process is Entirely Serial and Static (Pre-Scheduled):
No Response to the Weather!**



The LEAD Vision: A Paradigm Shift

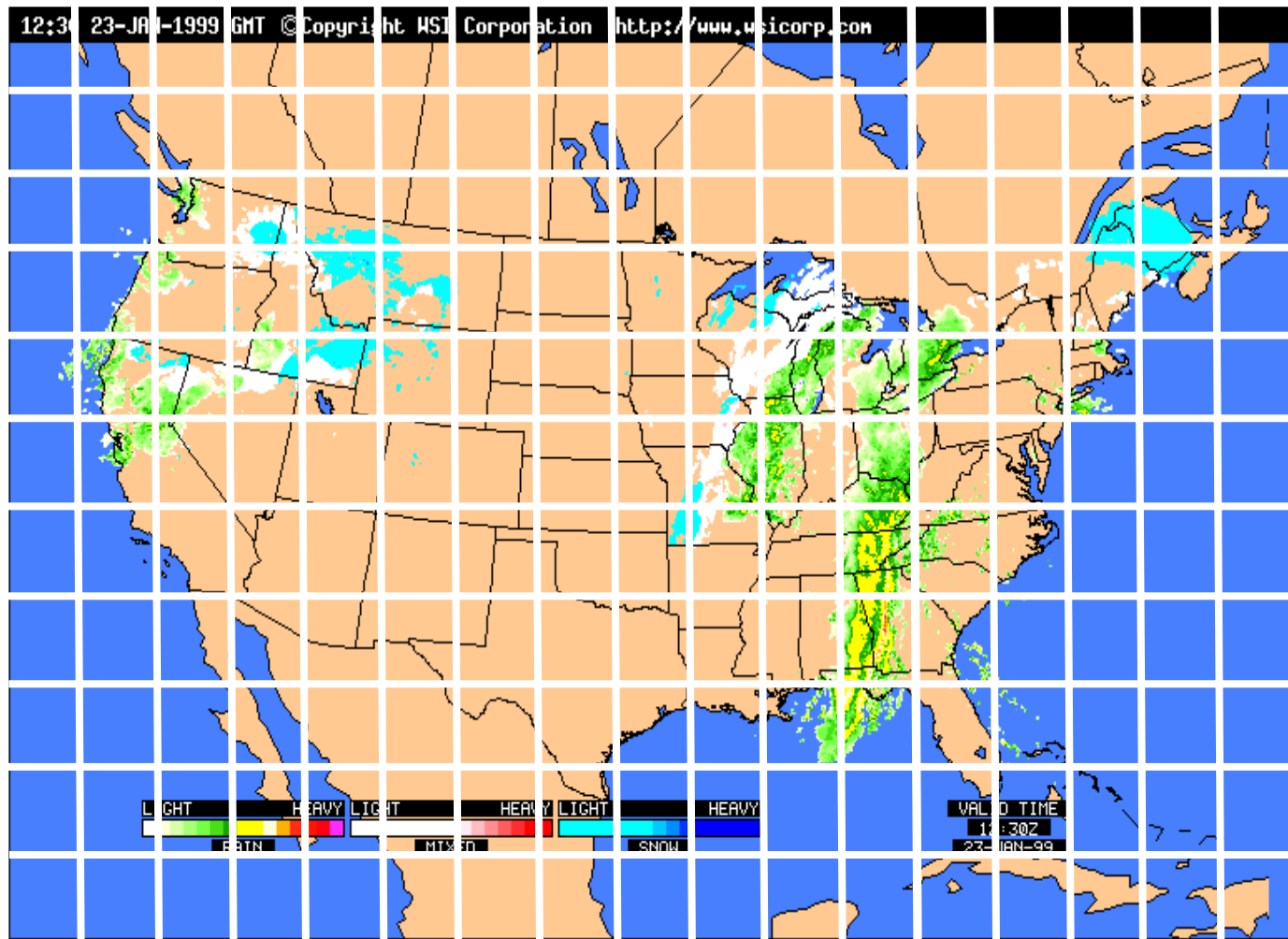


Change the Paradigm

- To make fundamental advances we need:
 - Adaptivity in computational model.
- But also Cyberinfrastructure to:
 - Execute complex scenarios in response to weather events
 - Stream processing, triggers
 - Close loop with the instruments.
 - Acquire computational resources on demand.
 - Need supercomputer-scale resources
 - Invoked in response to weather events
 - Deal with data deluge
 - User can no longer manage his/her own experiment products

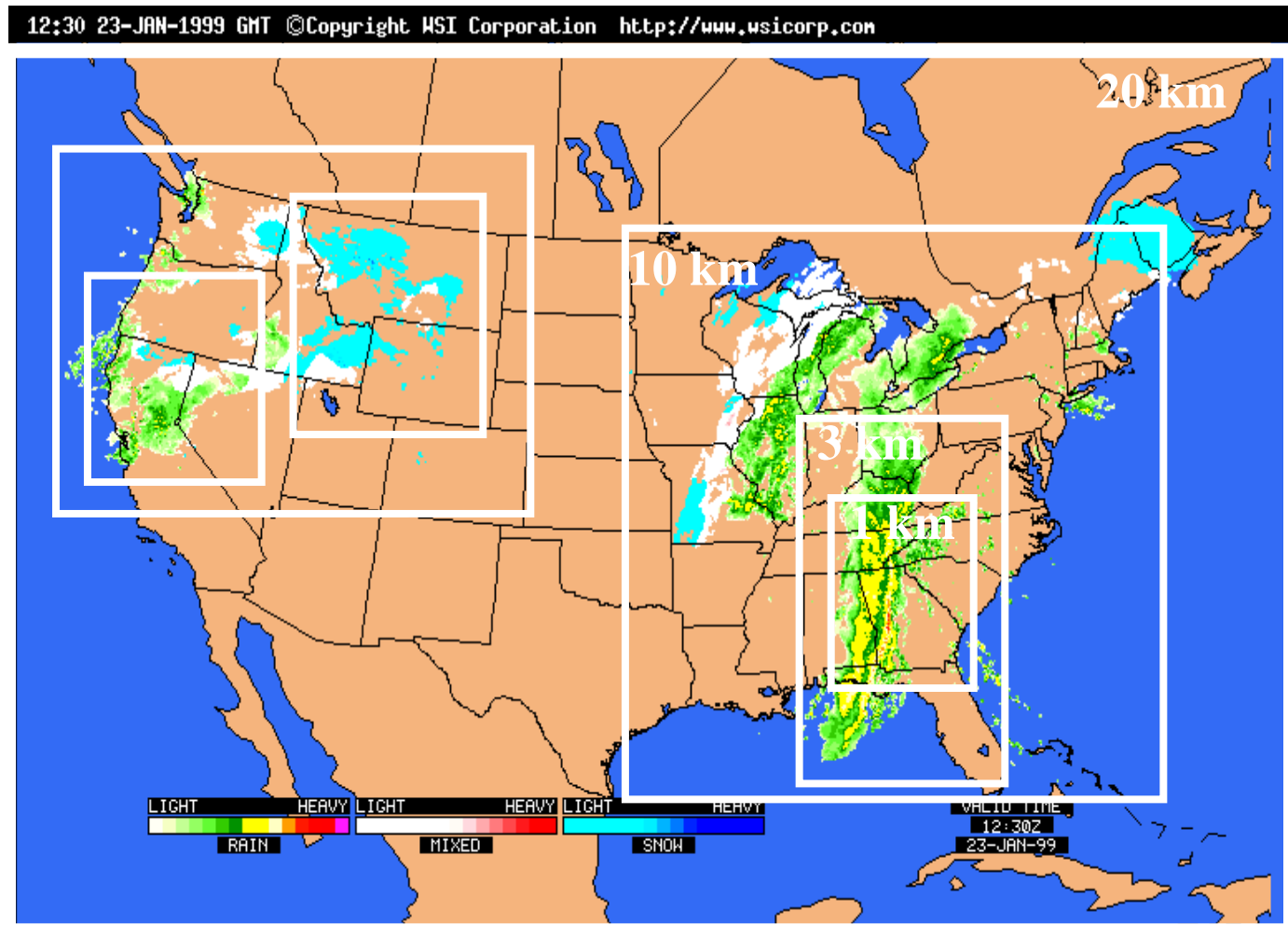


One Paradigm Shift Scenario: Model Grids Fixed

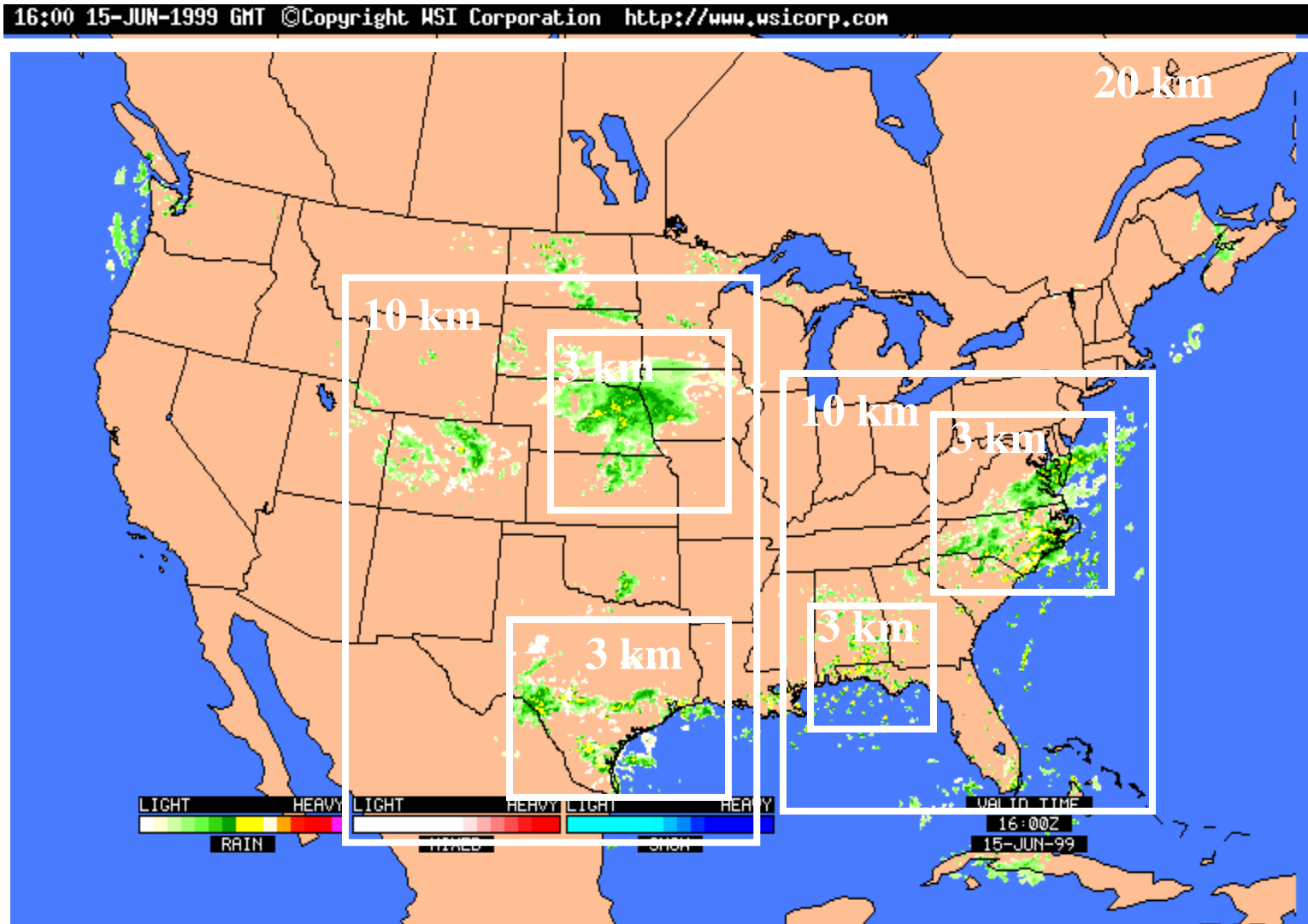


Model Dynamic Adaptivity

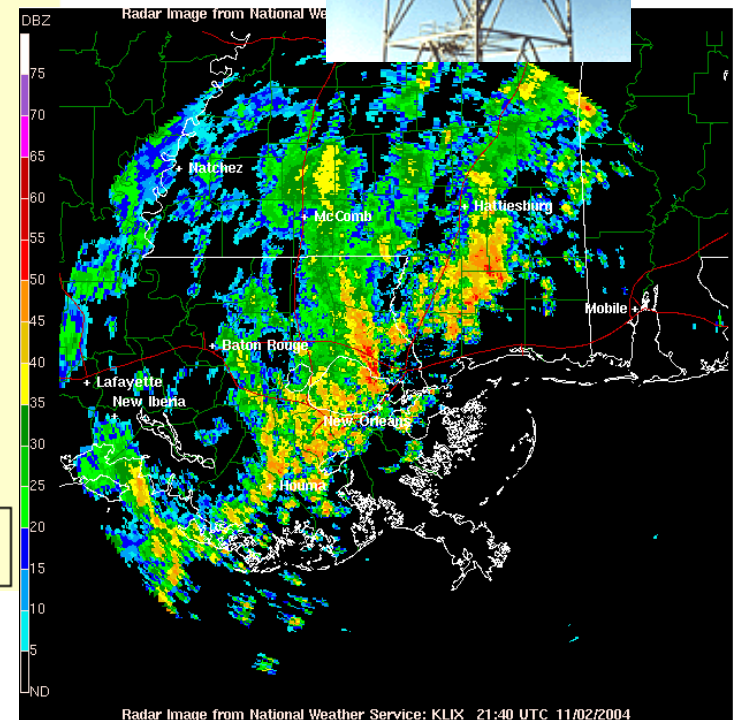
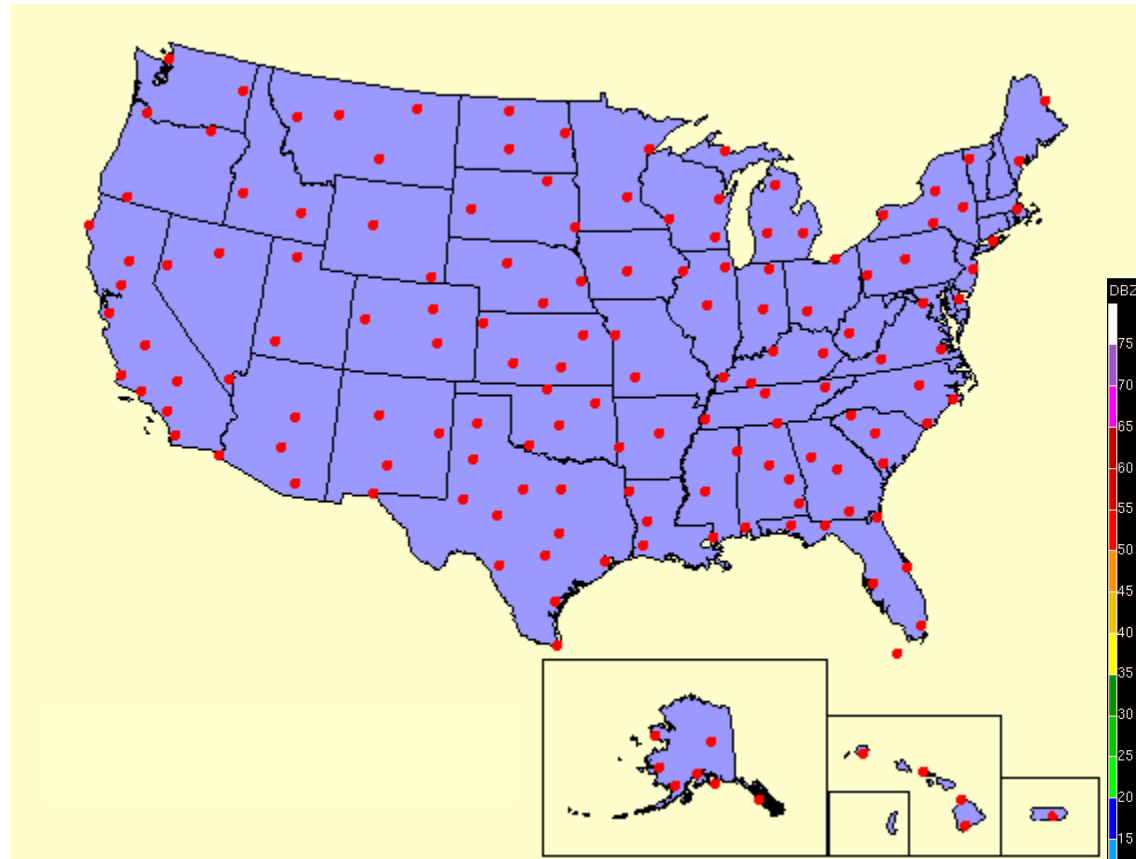
$$t = t_0$$



$$t = t_0 + 2 \text{ Hours}$$



Second Paradigm Shift: CASA NETRAD adaptive Doppler Radars.



Third Paradigm Shift: Experiment as Control- or Data-flow graph

Experiment Builder Portlet

User: Suresh Marru

Project: demo-demo-run

Experiment Status Page

Experiment Details

Name: Datamining-demo

Last Notification Time: 2005-07-19T14:25:32.590-05:00

Description: Datamining demo

Status: INFORMATION

Workflow

Name: ADaM Demo

Description: Datamining services workflow

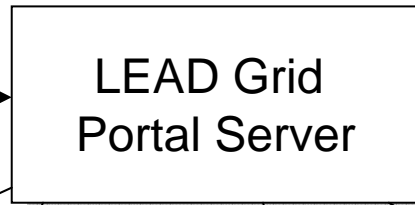
Parameter	Value
Output_Dir	gridftp://frozone.itsc.uah.edu/tmp

Notifications

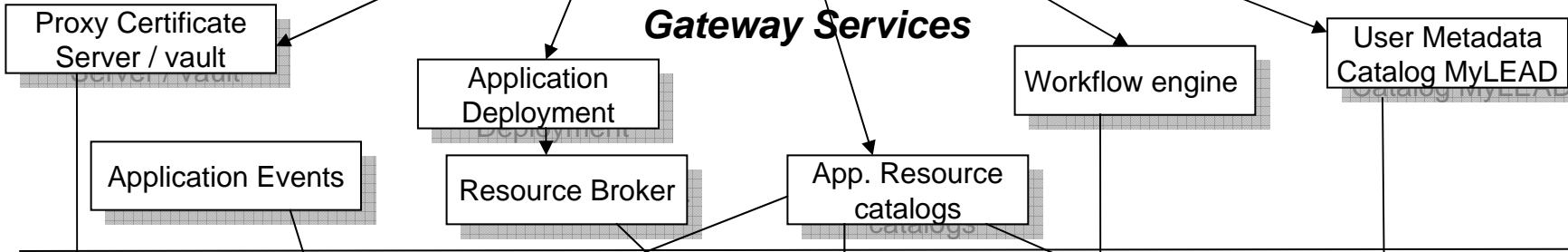
Message	Status	Timestamp
INVOKE SERVICE START	INFORMATION	2005-07-19T14:25:32
APPLICATION START	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: install location of the lead tools is /home/grid5070/production/adam-services	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: the local temporary directory is /tmp	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: the input url of the data file is gridftp://frozone.itsc.uah.edu/home/grid5070/production/adam-services/FeatureExtraction/input.dat	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: the output url to push the output files is gridftp://frozone.itsc.uah.edu/tmp	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: the site name is KOUN	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: the local work directory is /tmp/nexrad_1121800848241/3DMesocycloneDetection	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: attempting to copy input data file to /tmp/nexrad_1121800848241/3DMesocycloneDetection/input.data	INFORMATION	2005-07-19T14:25:32
FILE RECEIVE DURATION Duration(Millis): 1452 FILE UUID: testUUID	INFORMATION	2005-07-19T14:25:32
FILE CONSUMED FILE UUID: testUUID	INFORMATION	2005-07-19T14:25:32
LEVEL: INFO MESSAGE: attempting to push output files to gridftp://frozone.itsc.uah.edu/tmp	INFORMATION	2005-07-19T14:25:32
FILE SEND DURATION FILE SIZE (Bytes): 1025 FILE UUID: testUUID	INFORMATION	2005-07-19T14:25:32
PUBLISH URL: http://chinkapin.cs.indiana.edu/data-output/input.data_3Dmesocyclone_1121800848241.dat	INFORMATION	2005-07-19T14:25:32
PUBLISH URL: http://chinkapin.cs.indiana.edu/data-output/detection_1121800848241.output	INFORMATION	2005-07-19T14:25:32

The Architecture

The User's Desktop



Gateway Services

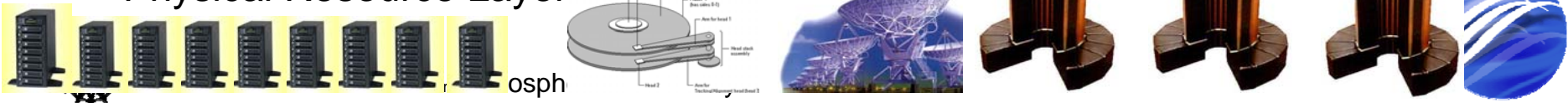


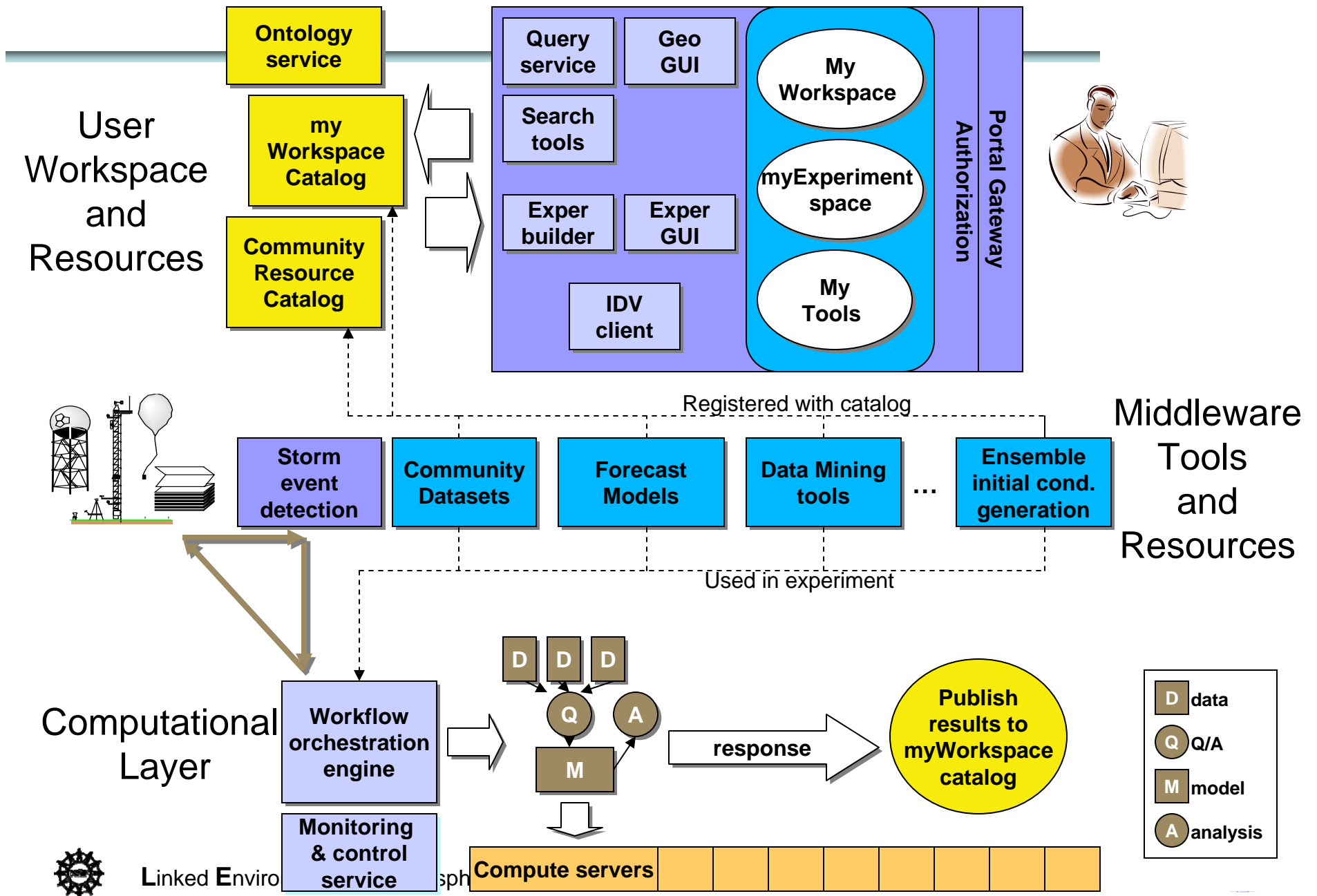
Core Grid Services



OGSA-like Layer

Physical Resource Layer





The LEAD Portal

Provides access to the LEAD resources

Weather and LEAD testbed status

Welcome to the **LEAD PORTAL** Linked Environments for Atmospheric Discovery
Sponsored by the National Science Foundation

Portal Home | Geo GUI | Education and Outreach | Weather | Links | About LEAD | Help

To view a local radar, select area of interest and click on the image below.

**RADAR REFLECTIVITY FROM RADAR CODED MESSAGES
NATIONAL WEATHER SERVICE
AUTOMATED EDITING APPLIED
SEP 24, 2005 21:49 UTC**

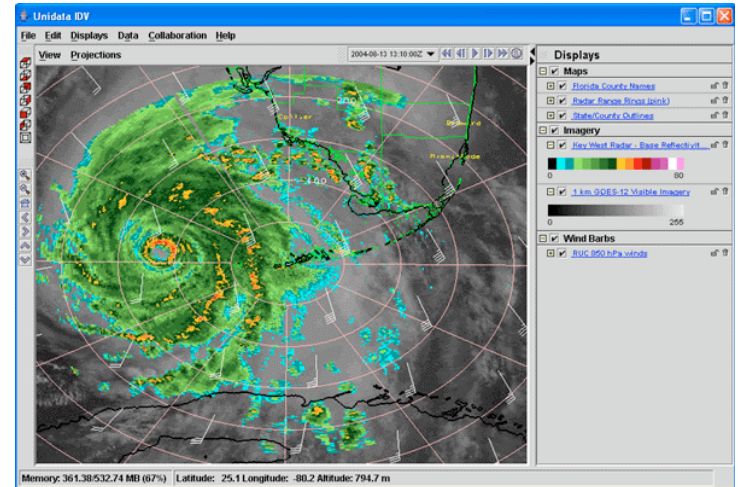
Testbed	Grid Auth	GRAM	Grid
IU [chinkapin]	✓	✓	✓
NCSA [copper]	✓	✓	✓
OU [aquaman]	✓	✓	✓
UAH [frozone]	✓	✓	✓
UNC [dante0]	✗	✗	✗
Unidata [lead1]	✓	✓	✓

Last updated: Sat Sep 24 17:00:00 2005 Indiana 1

Legend: MSG, 55 DBZ, 50 DBZ, 45 DBZ, 40 DBZ, 30 DBZ, 15 DBZ

LEAD Home | FAQ | Privacy | Terms of use | Contact us

Linked Environments for Atmospheric Discovery



Viz tool

Workflow tool

Workflow Composer

Workflow MyLead Component Monitor Help

Add Node Remove Node Connect/Disconnect

Component List

- System Components
- http://whitney.extreme.indiana.edu
- http://www.extreme.indiana.edu
- Adder
- Multiplier
- Divider
- decoder
- threads
- arps-trn
- arps-sfc
- ext2arps-ibc

Component Information

Service: decoder

Description:
A service for decoding raw eta data to netcdf format

Operation: Run

Port Information | Notification |

Selected Output Port

Selected Input Port

Component: Output_URL
Port: Parameter
Type: Any
Description: This port can be connected to any type.

LEAD Portal

Welcome to the **LEAD PORTAL** **Linked Environments for Atmospheric Discovery**
 Sponsored by the National Science Foundation

Portal Home **Geo GUI** Education and Outreach Weather Links About LEAD Help

Geo Reference GUI

Data Products

- Surface METAR Obs
- UpperAir Balloon Soundings (Rawinsondes)
- 5-minute ACARS
- NEXRAD Level II
- NEXRAD Level III
- GOES Imagery
- NCEP Eta Forecast Model

Data Categories:

- Temperature
- Wind
- Water Vapor/Moisture

Spatial Extent initialize

Lon, Lat: -82.0917, 37.3621

Temporal Range

From

Date:

Time:

To

Date:

Time:

or

Lon, Lat: -82.0917, 37.3621 -- x, y: 383, 266

Select community data products for import to workspace or use in experiment



Workflows: Execution of Complex Experiments

LEAD requires ability to construct workflows that are

- Data Driven
 - Weather data streams define nature of computation
- Persistent and Agile
 - Data mining of data stream, detects “interesting” feature, event triggers workflow scenario that has been waiting for months.
- Adaptive
 - In response to weather: weather changes.
 - Nature of workflow may have to change on-the-fly.
 - Resources
 - More may be needed, sometimes they become unavailable.
 - Need to be self-aware



The workflow composer

- User designs, then compiler generates GBPEL

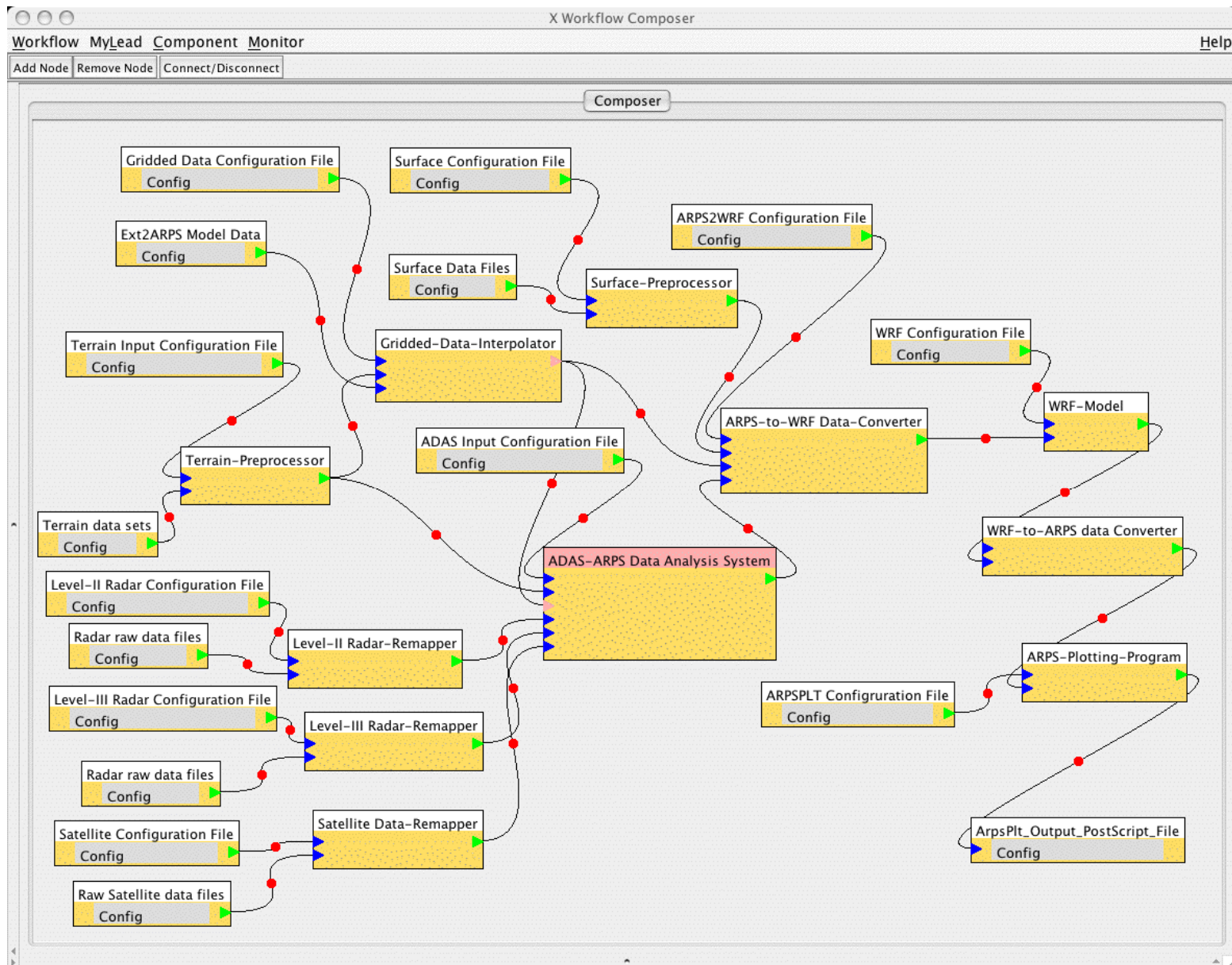
The screenshot displays the X Workflow Composer application window. The title bar reads "X Workflow Composer" and includes standard window controls. Below the title bar is a menu bar with "Workflow", "MyLead", "Component", "Monitor", and "Help". A toolbar contains "Add Node", "Remove Node", and "Connect/Disconnect" buttons.

The main workspace is divided into several panels:

- Component List:** A tree view showing a hierarchy of components. Under "System Components", there are two sub-folders for "http://whitney.extreme.indiana.edu:" and "http://www.extreme.indiana.ec". The second folder contains a list of components: Adder, Multiplier, Divider, decoder, thredds, arps-trn, arps-sfc, and ext2arps-ibc.
- Composer:** A central workspace showing a workflow diagram. It features three "Config" nodes (yellow boxes with a green arrow) and one "decoder" node (pink box with a green arrow). The nodes are connected by lines with red dots at the connection points, indicating a flow from the two input "Config" nodes to the "decoder" node, and then to the "Output_URL" node.
- Component Information:** A panel displaying details for the selected "decoder" component. It includes:
 - Service:** decoder
 - Description:** A service for decoding raw eta data to netcdf format
 - Operation:** Run
- Port Information:** A panel with two tabs: "Port Information" and "Notification". The "Port Information" tab is active, showing:
 - Selected Output Port:** (Empty)
 - Selected Input Port:** Component: Output_URL, Port: Parameter, Type: Any, Description: This port can be connected to any type.



Assimilation-forecast workflow



Sample App Services

ARPS Data Analysis System – ADAS

Description: Generates 3D gridded analysis of the current atmosphere by combining the observed information from NEXRAD radars, wind profilers, satellites, surface observation networks and aircrafts with a background field created by external model data interpolator.

Input: A configuration file and processed observational data from various data preprocessors.

Output: ARPS history formatted analyzed data.

ARPS to WRF data converter – ARPS2WRF

Description: Ingests data files in ARPS history format and generates WRF input and lateral boundary files.

Input: A configuration file, surface characteristic files and initialization files in ARPS history format.

Output: WRF input file, WRF lateral boundary file and WRF initialization files in NetCDF format.

WRF Model – WRF

Description: Performs storm, mesoscale and synoptic weather prediction by a non-hydrostatic, limited area model to study convection, baro-clinic waves, boundary layer turbulence, real-time weather phenomena.

Input: A configuration file, ADAS analysis data and later boundary conditions generated by external model interpolator.

Output: Weather forecast output data files.



More Sample LEAD Services

Terrain Data Preprocessor – ARPSTRN

Description: Performs analysis of terrain data and generates a terrain file by interpolating the data to the ARPS grid.

Input: A configuration file and a terrain data set.

Output: Processed terrain data file.

Surface Characteristics Preprocessor – ARPSSFC

Description: Prepares the surface characteristic data set for use in ARPS and generates surface characteristic files with soil types, vegetation types, leaf area index and surface roughness.

Input: A configuration file with soil type, vegetation type and vegetation fraction data files.

Output: Constructed set of surface and vegetation characteristic fields for the ARPS grid.

Gridded Data Interpolator – EXT2ARPS

Description: Extracts and interpolates pertinent fields from a National Weather Service model forecast dataset to an ARPS grid to provide an ADAS analysis background or initial conditions and boundary conditions for an ARPS/WRF forecast.

Input: A configuration file and gridded data from external models like NCEP ETA, RUC and AVN.

Output: Processed data files with fields on ARPS grid in standard ARPS history format and external boundary conditions format.

Radar Level-II Data Remapper – 88D2ARPS

Description: Converts raw NEXRAD Level II radar data in polar coordinates to Cartesian coordinates and remaps the data to the ARPS grid.

Input: A configuration file and raw WSR-88D Level-II radar data files.

Output: Re-mapped data files.

Radar Level-III Data Remapper – NIDS2ARPS

Description: Converts WSR-88D Level-III raw velocity and reflectivity data and remaps it onto a sigma-Z Cartesian ARPS grid.

Input: A configuration file and raw WSR-88D Level-III radar data files.

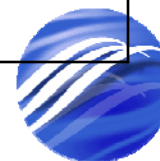
Output: Re-mapped data files.

Satellite Data Remapper – MCI2ARPS

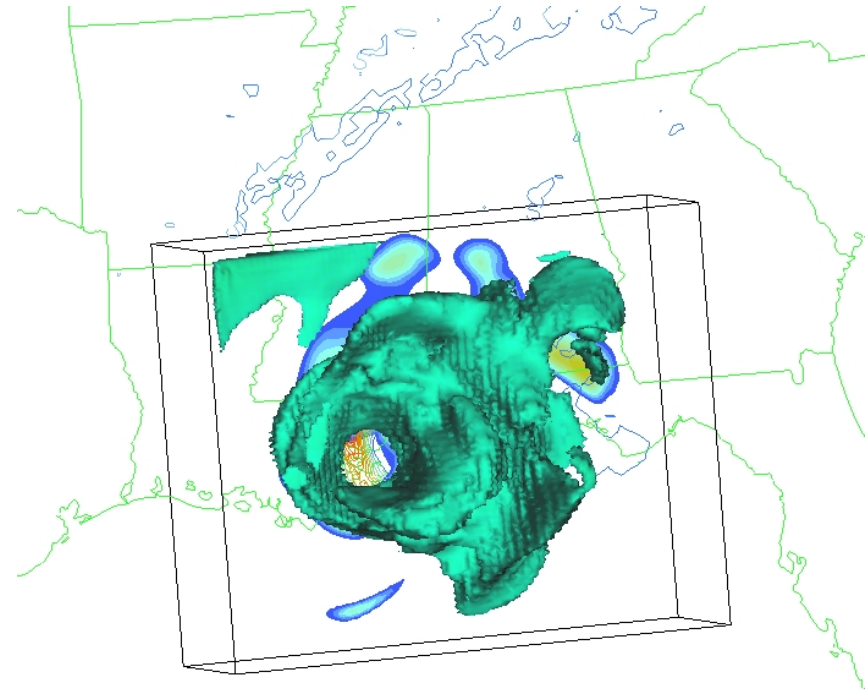
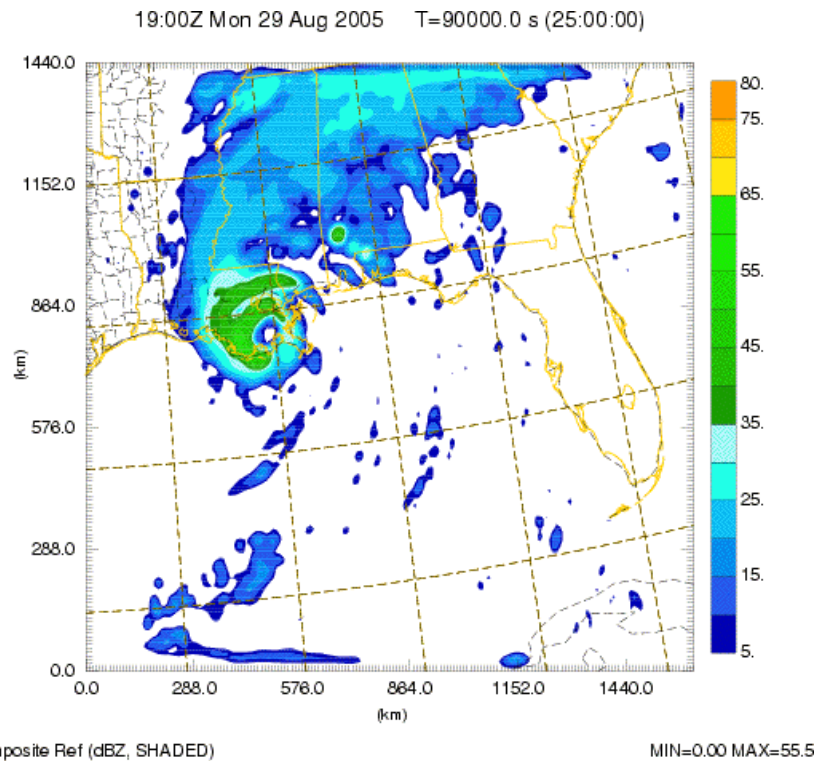
Description: Remaps observed McIDAS GVAR AREA satellite data from the satellite-observed pixels to the ARPS grid.

Input: A configuration file and observed data from satellites.

Output: Re-mapped data files.



Workflow applied to Katrina

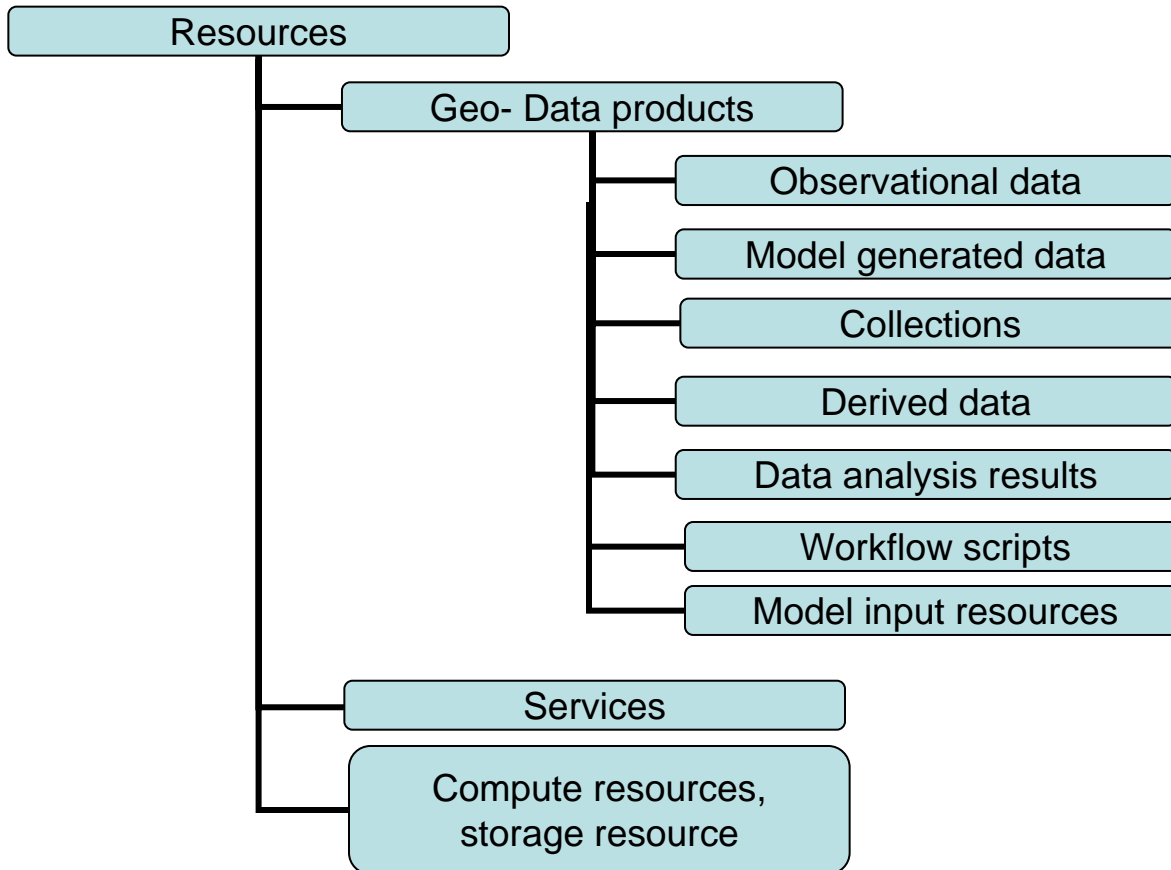


2D image generated by ARPS Plotting Service

3D Image generated by IDV



Managing Data in LEAD



Personal data products

-- user's experiment products, personal collections, scripts, input config params.

Community data products

-- data, mostly observational, made available to LEAD virtual community

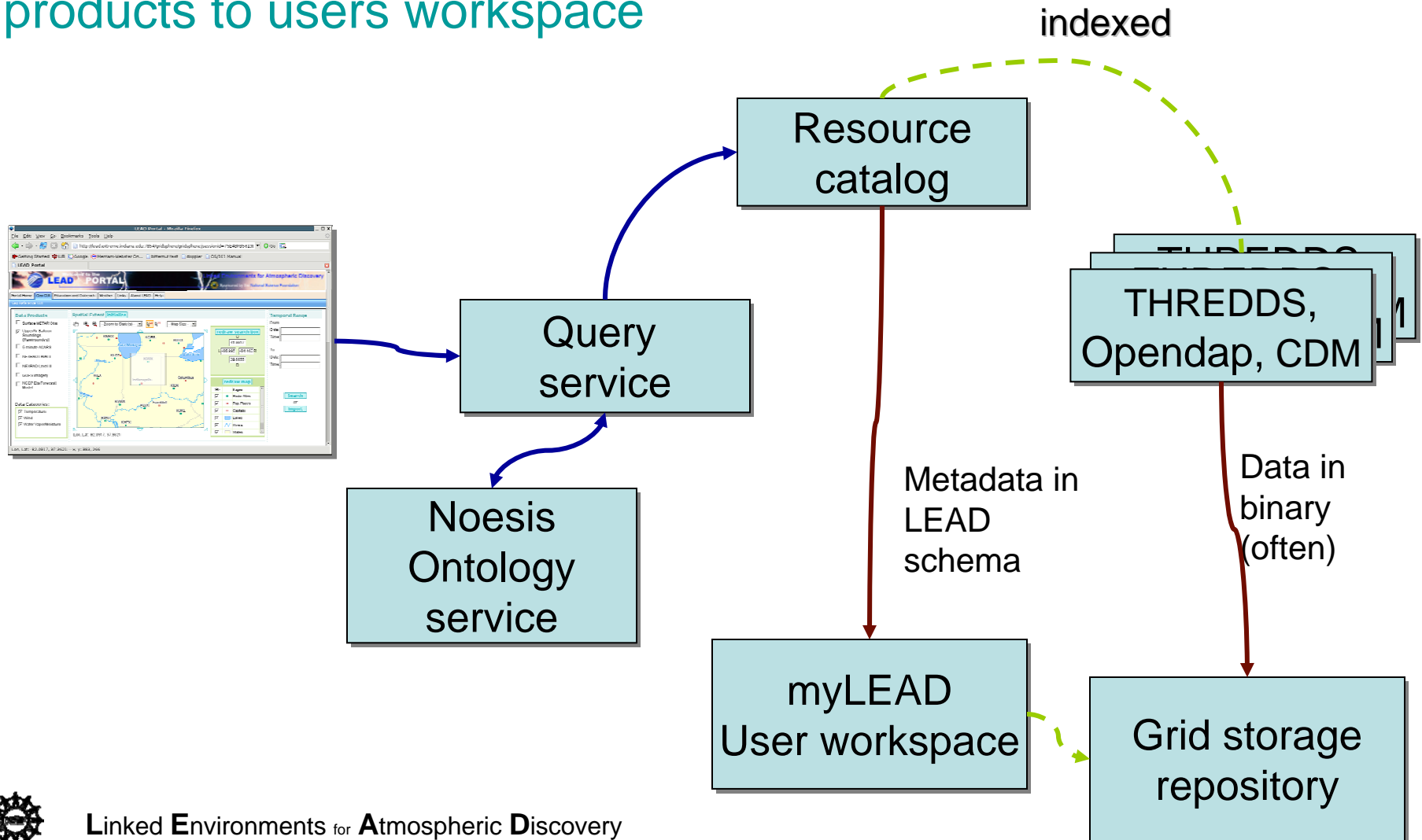
External products

-- new and future data sources.



LEAD Data Use Scenario

- Importing community data products to users workspace



Log in and see your MyLEAD Space

Welcome to the **LEAD PORTAL** Linked Environments for Atmospheric Discovery
Sponsored by the National Science Foundation

LEAD Portal Home Education and Outreach Help Profile OGCE LEAD LEAD-DEV

MyLEAD Workspace Security

My Workspace Portlet

- myWorkSpace
 - Hurricane Ivan
 - Workflow template vizEta 2004-08-03 13:35:
 - Workflow template ARPS 2004-09-22 05:25:5
 - TestWorkflowTemplate01
 - Experiment: Ensemble run-5
 - Workflow instance ARPS 2004-09-22 05:25
 - Collection: Ensemble Run Input Collection
 - Collection: Case 1
 - Collection: Case 2
 - Collection: Case 3
 - Collection: Case 4
 - Collection: Case 5
 - Experiment: 84hr ARPS forecast
 - Collection: Input Observational for 84hr AR
 - Collection: ARPS-out for 84hr ARPS forecas
 - CompletelyNewExperiment
 - HurricaneExperiment01
 - Mesoscale meteorology for college school stu

Information of your current selection

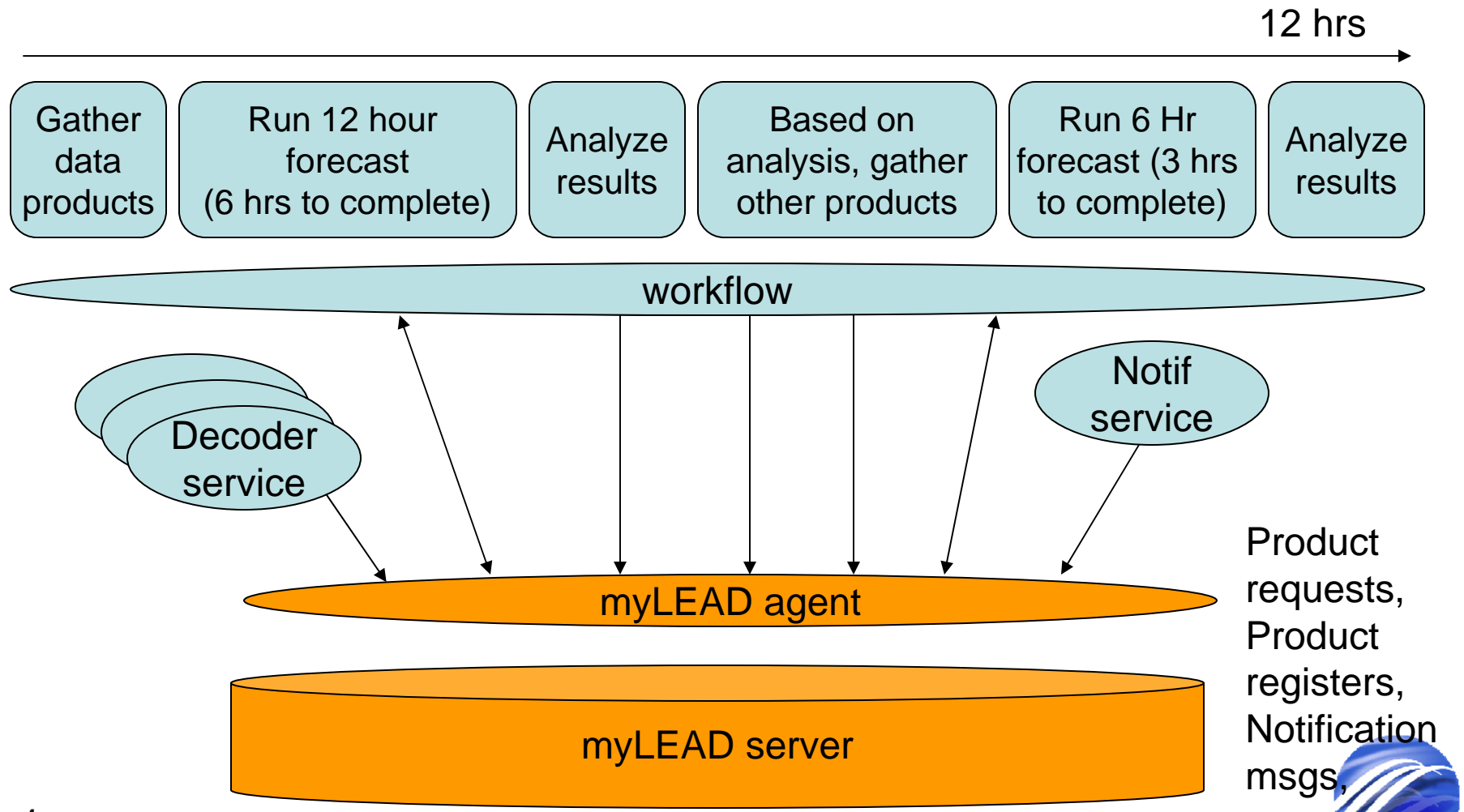
Desc: 84hr ARPS forecast
ExpDate: 2004-10-28 00:00:00

User's Workspace (myLEAD)

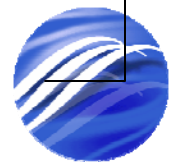
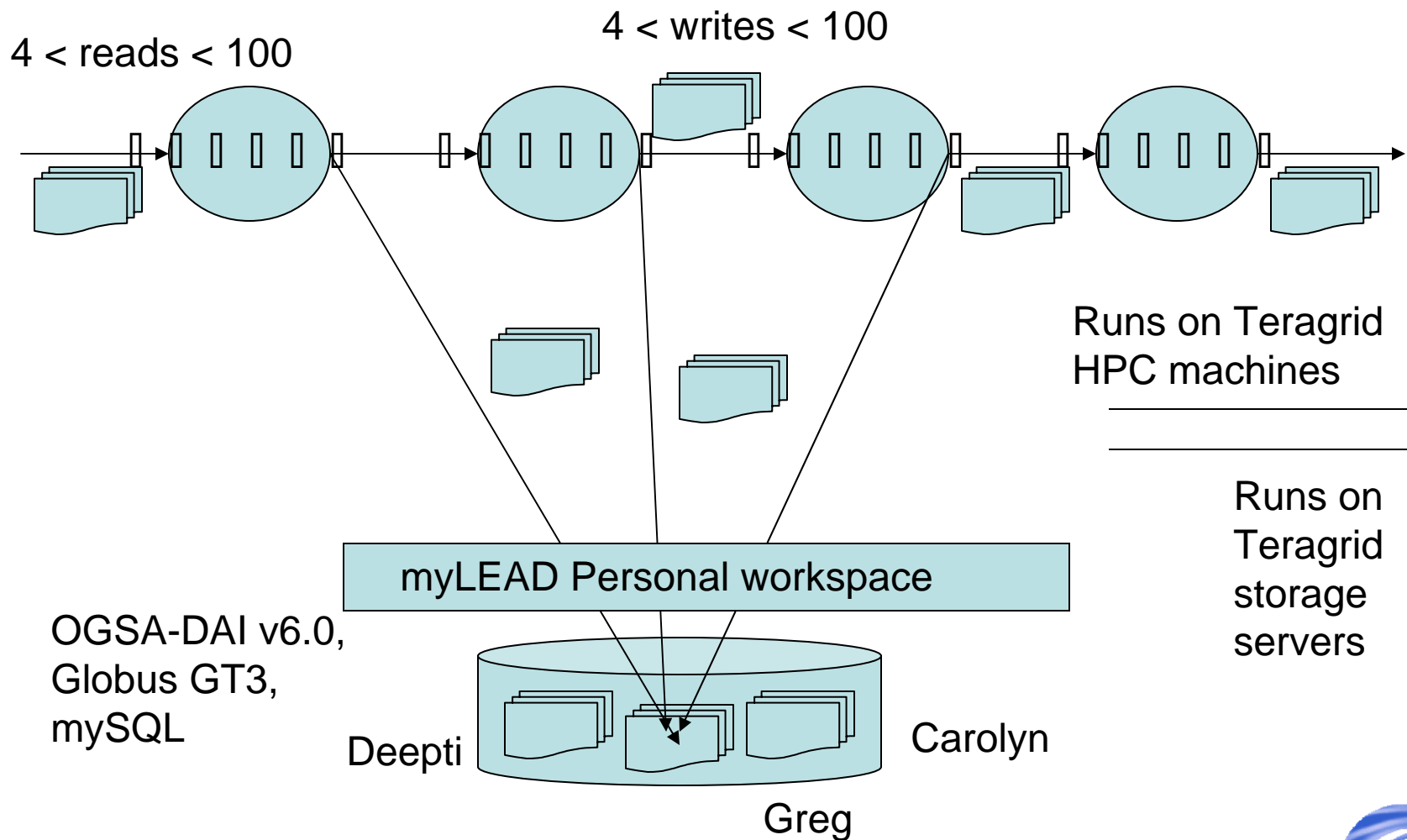
- Metadata catalog of user's data products
- User's storage on LEAD grid
- Agent actively archives data products:
 - **Derived data products** - data products result of processing original raw data
 - **Temporally changing data products** - data continuously changing through
 - regular additions streamed into archive
 - Ad hoc actions taken by content creators, or
 - In conjunction with workflow processes.
- Approach: general, reusable data model; open source database (mySQL); standardized metadata schemas (XML); service-oriented architecture (SOAP, WSDL, gridFTP, x509 certificates)



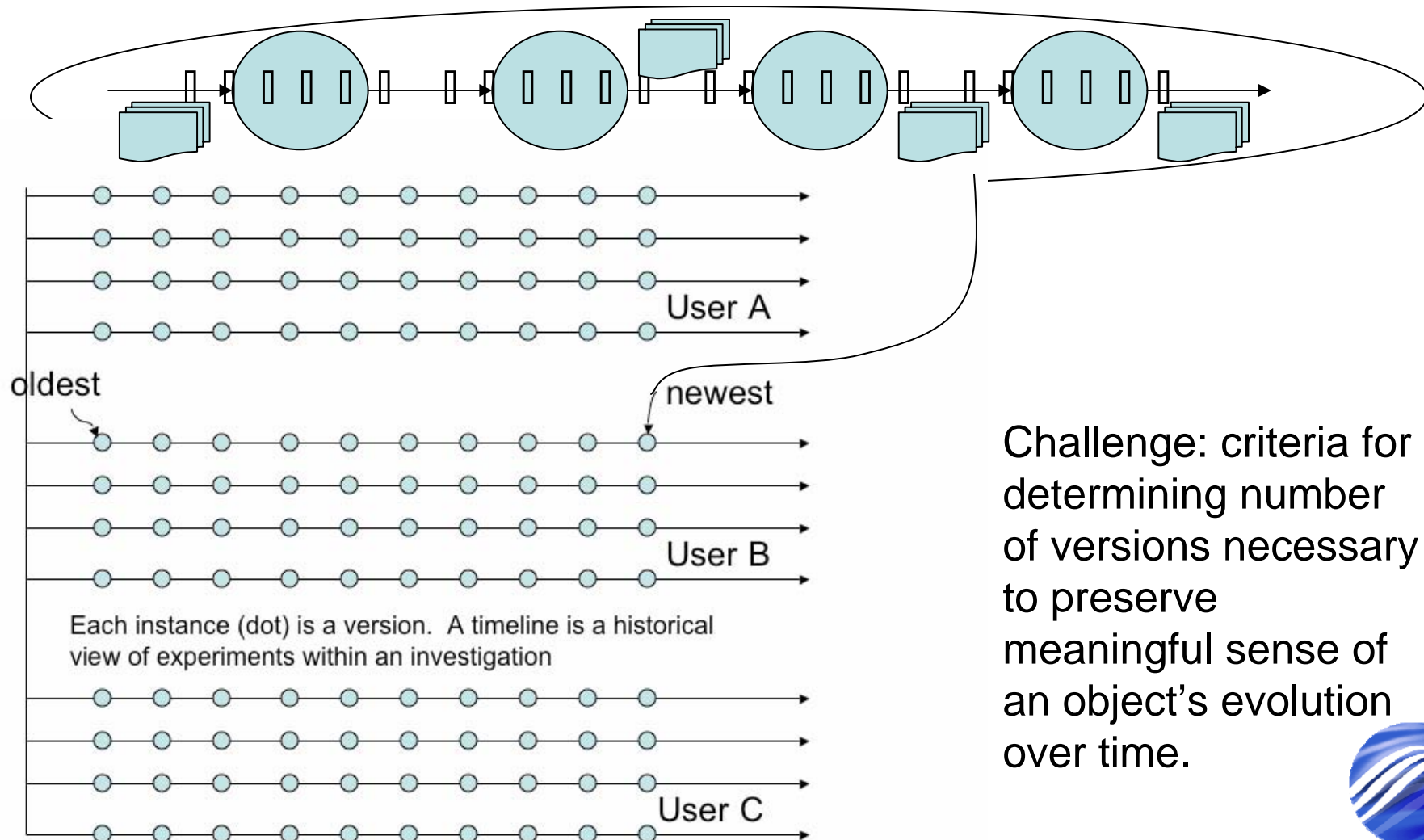
Creating structure in user's archive that models their investigation steps



Archiving **derived** and temporally changing data products



Archiving derived and **temporally changing** data products



Challenge: criteria for determining number of versions necessary to preserve meaningful sense of an object's evolution over time.



Conclusions

- LEAD has demanding set of requirements: architecture is response
 - Many in science community applying service-oriented architecture
- LEAD unique in fusion of data management and workflow based on
 - Tight coupling based on very fine grained event tracking and metadata based on entire experiments.
- Adaptivity requirements require rethinking of workflow, stream processing, ...

