Grid Computing for Real World Applications

Suresh Marru

Indiana University

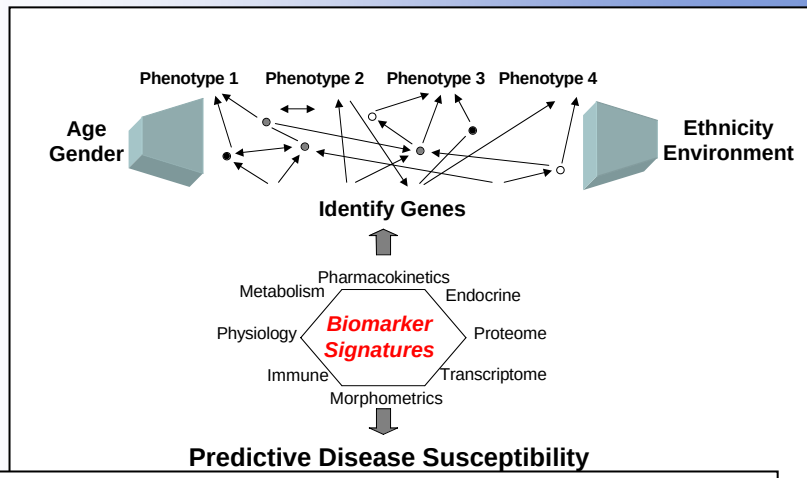
(smarru@cs.indiana.edu)

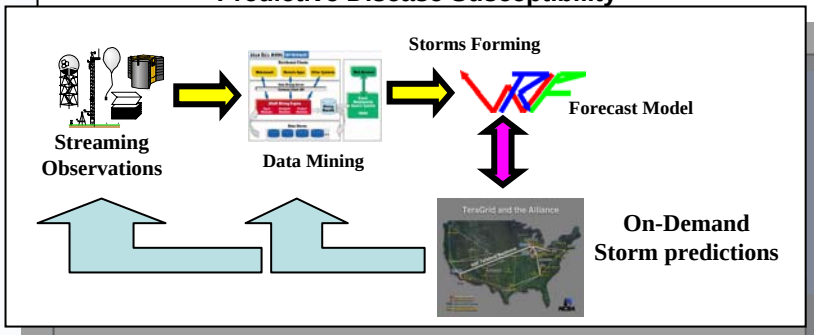
5th October 2005
OSCER Symposium @ OU

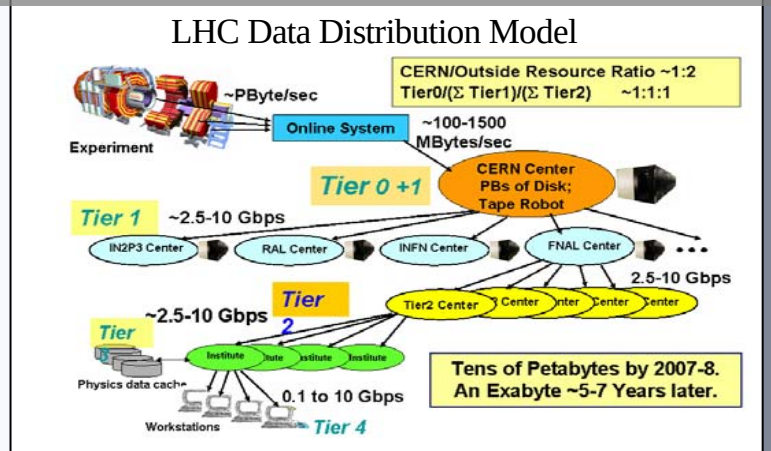
Motivation: Scientific Challenges

The current and future generations of scientific problems are:

- Data Oriented
  - Increasingly stream based.
  - Often need petabyte archives.
- In need of on-demand computing resources
- Conducted by geographically distributed teams of specialists
  - Users do not want to expend too much time learning new technologies.







Solution

Adapt Grid Computing and solve every computing problem in this world.

Is this true? I wish it is, but not really ..then what ..

- Grid Technology bridges the gap between the applications and the infrastructure.

Fine, but what the heck is grid computing? follow along to find out …

Introduction to Grid

Grid Computing enables

- sharing,
- selection and
- aggregation

of a wide variety of geographically distributed resources

including

- supercomputers,
- storage systems,
- data sources and
- specialized devices

owned by different organizations

for solving large-scale resource intensive problems in

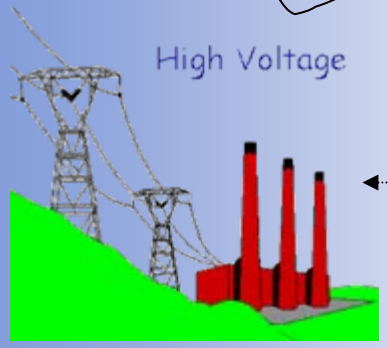
science, engineering, and commerce.

Power Grid Analogy

Users

Interface

Grid Portal



Infrastructure

Gateway Services

MyProxy

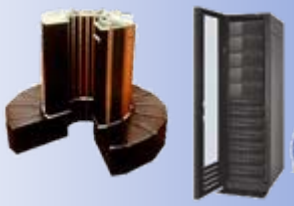
Grid Middleware

Globus Server API

Resources

Supercomputers

Networks



Storage Devices

Instrumentation

Key Features of Grid Computing

- Provides a secure infrastructure for computing on a distributed computing environment
- Provides single sign-on feature by which a user can authenticate once and perform multiple computations over extended period of time
- Facilitates inter-domain access mechanisms
- Better portability (code can run on many kinds of computers) and exportability (move files from one computer to another)

Grid Security

- Grid Certificates:
- Needed for using the Grid
- Used to provide a set of privileges of one resource to another
- Provide the features of dynamic delegation, dynamic entities and repeated authentication
- Standard PKI infrastructure is used for validation

Challenges of using Grid computing

- The concept of grid is promising but users have to cope up with ..
  - ❖ Emerging technology
  - ❖ Evolving standards
  - ❖ Frequent new versions of middleware with little or no backwards compatibility
  - ❖ Users have to learn the technology to use it.

Alternatively, use grid-enabled science portals

Science Portals

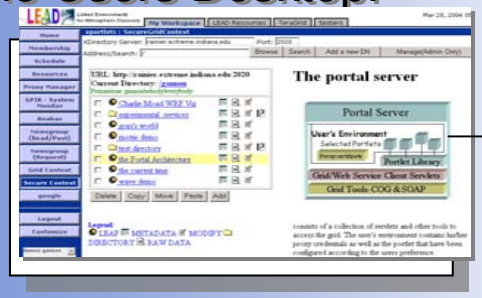
- The goals of a Science Portal are
  - To give a community of researchers easy access to the tools, data and computational power needed to solve today’s scientific and engineering problems.
  - To do this in a discipline specific language that is common to the target community.
  - To hide any underlying Grid technology.

Portal Science Capabilities

- Data Access is the most important
  - Allow the user community access to important shared data resources
    - Visualize it, publish it, download it, curate it.
  - Data Discovery
    - Searchable metadata directories.
- Web access to important tools
  - Web-form interfaces to allow users to run important community codes
  - Webstart access to common java-based tools
  - Limited shell access - perhaps to a VM
- Workflow tools
  - Allow users to combine community codes into workflows managed by the portal back-end resources.

The Architecture of Gateway Services

The Users Desktop.



Grid Portal Server

Gateway Services

Proxy Certificate Server / vault

Application Deployment

Application Workflow

User Metadata Catalog

Application Events

Resource Broker

App. Resource catalogs

Replica Mgmt

Core Grid Services

Security Services

Information Services

Self Management

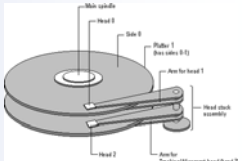
Resource Management

Execution Management

Data Services

OGSA-like Layer

Physical Resource Layer

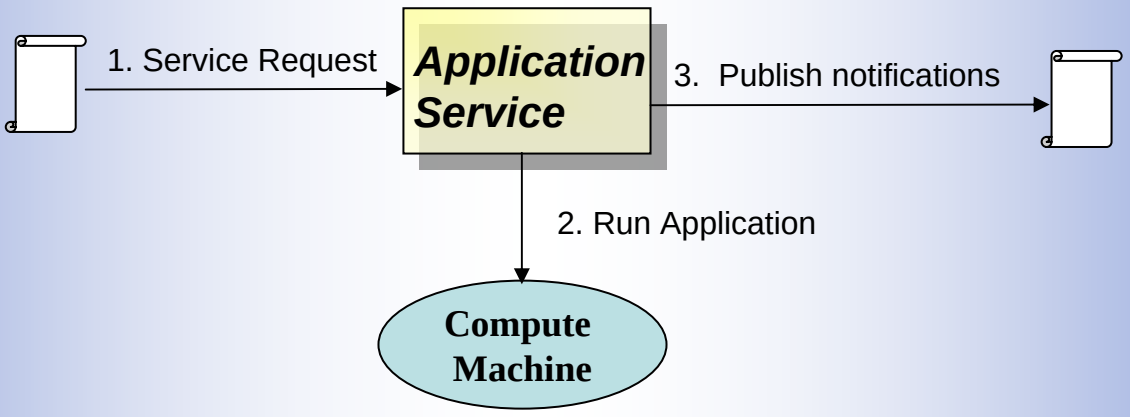


Service Architecture

- The Foundation of the gateway science portal software is based on the concept of “services” and “service oriented architectures.”

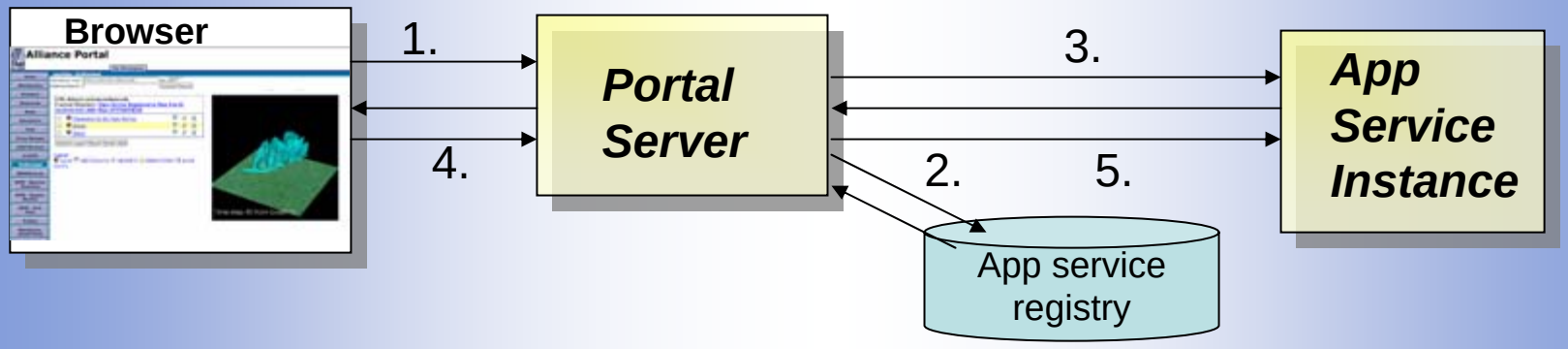
What's a service anyway?

- A "web server" that runs an application for you.
  - You send it requests (XML documents) and it processes the information and send replies (notifications) when it is done.



The Portal - Service interaction

- Each application is deployed as a service which can be invoked by the portal or another service.
  - 1. User looks up & selects application services from portal.
  - 2. Portal locates service instance.
  - 3. Service is contacted and replies with a interface description
  - 4. Server displays the interface and user fills it out.
  - 5. Server create ws request and sends it to the app service.



What do we do with Applications?

- Service-oriented **applications**
  - Wrap applications as services
  - Compose applications into workflows
  - Execute applications on remote resources on behalf of user in a secured manner

What a User Gains By Using Grid and Portals

- As a direct user
- – Can easily
- Execute jobs at one or more remote sites
- Move data between sites
- All with single sign-on security
- As a user of a grid enabled application
- Will not see the grid
- Will see an application whose development was
- eased with grid functions or grid-based web services
- Ease of development should result in more
- applications or faster availability of applications

What Application Developers Gain by Using Grids and Portals

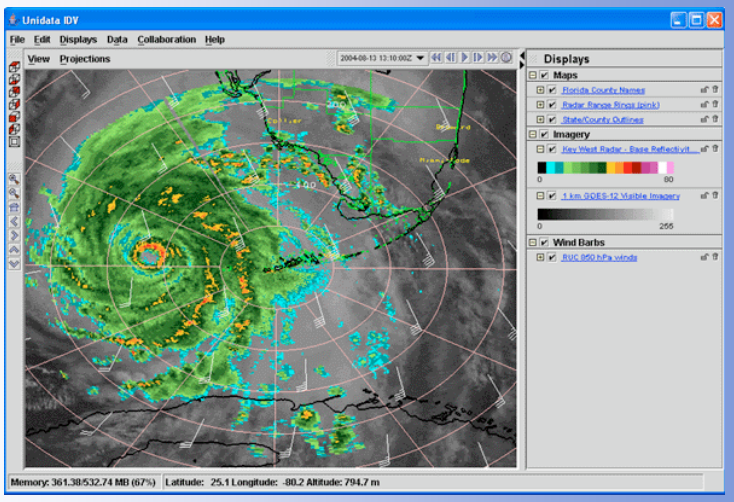
- Application web services can be built by re-using
- capabilities provided by existing grid-enabled
- Web services.
- Applications can also be built by using grid
- functions
- Grid functions/services handle distributed
- management of tasks and data
- – Developer can focus on logic of application and not
- logic of distributed interaction

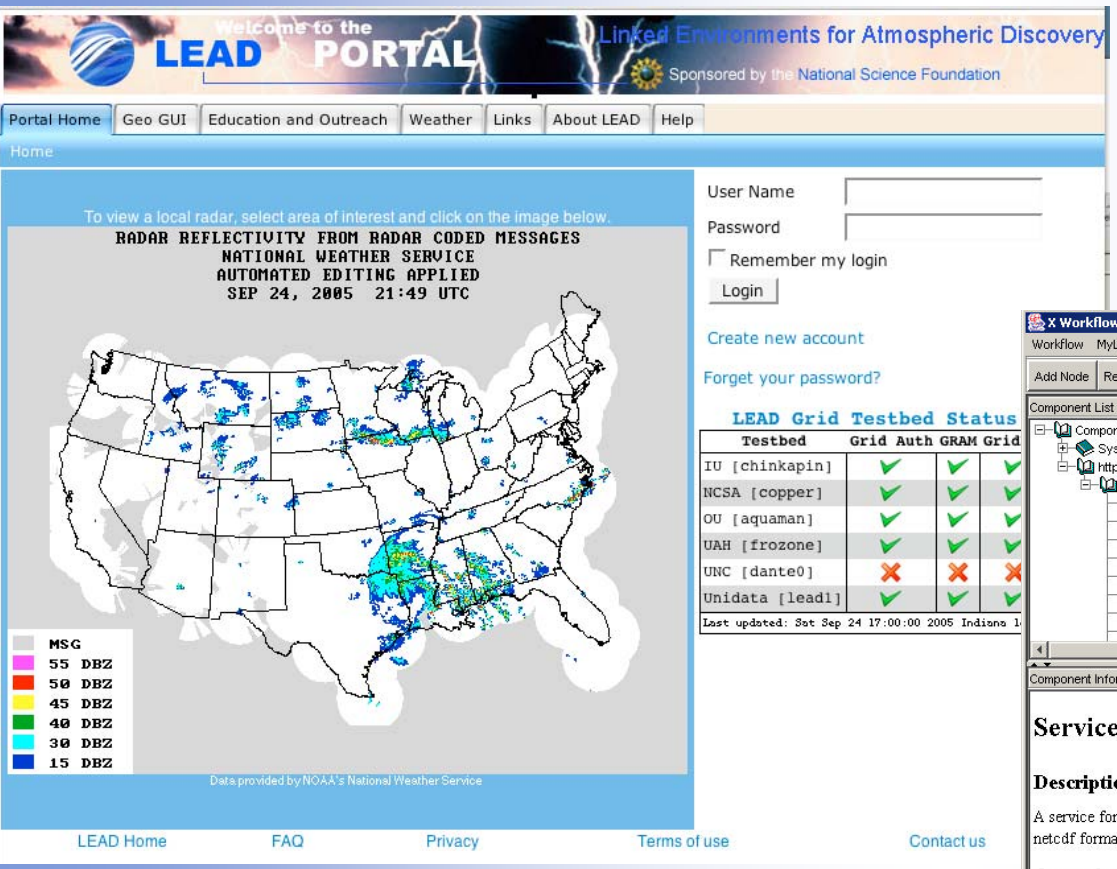
Example Gateway LEAD – Linked Environments for Atmospheric Discovery

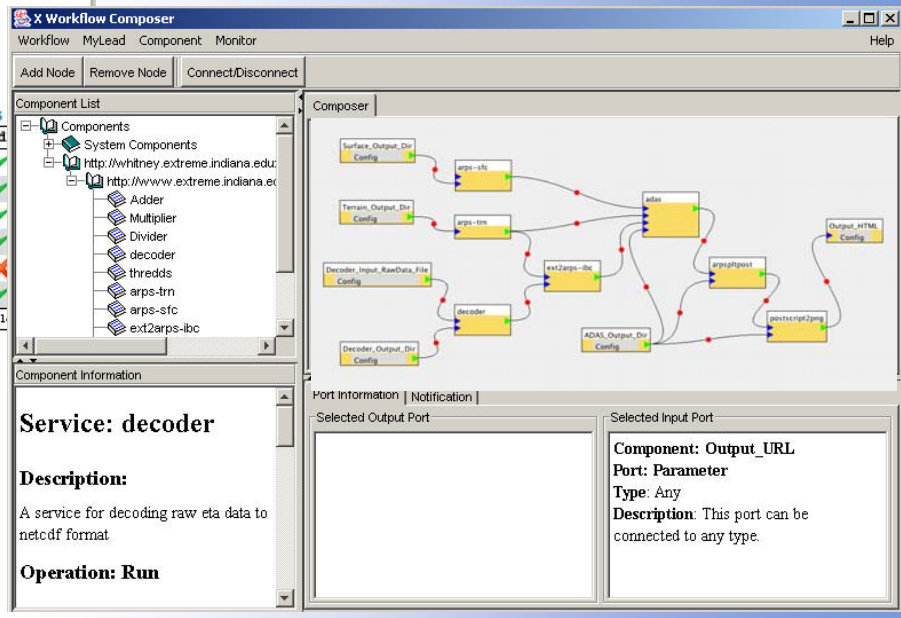
(Mesoscale Meteorology)

NSF LEAD project - making the tools that are needed to make accurate predictions of tornados and hurricanes.

- Data exploration and Grid workflow







Example Gateway - LEAD
(Mesoscale Meteorology)

University Corporation for Atmospheric Research

Colorado State University

HOWARD UNIVERSITY

INDIANA UNIVERSITY
Quality Education. Lifetime Opportunities.

LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

UNC CHAPEL HILL

UAH The University of Alabama in Huntsville

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

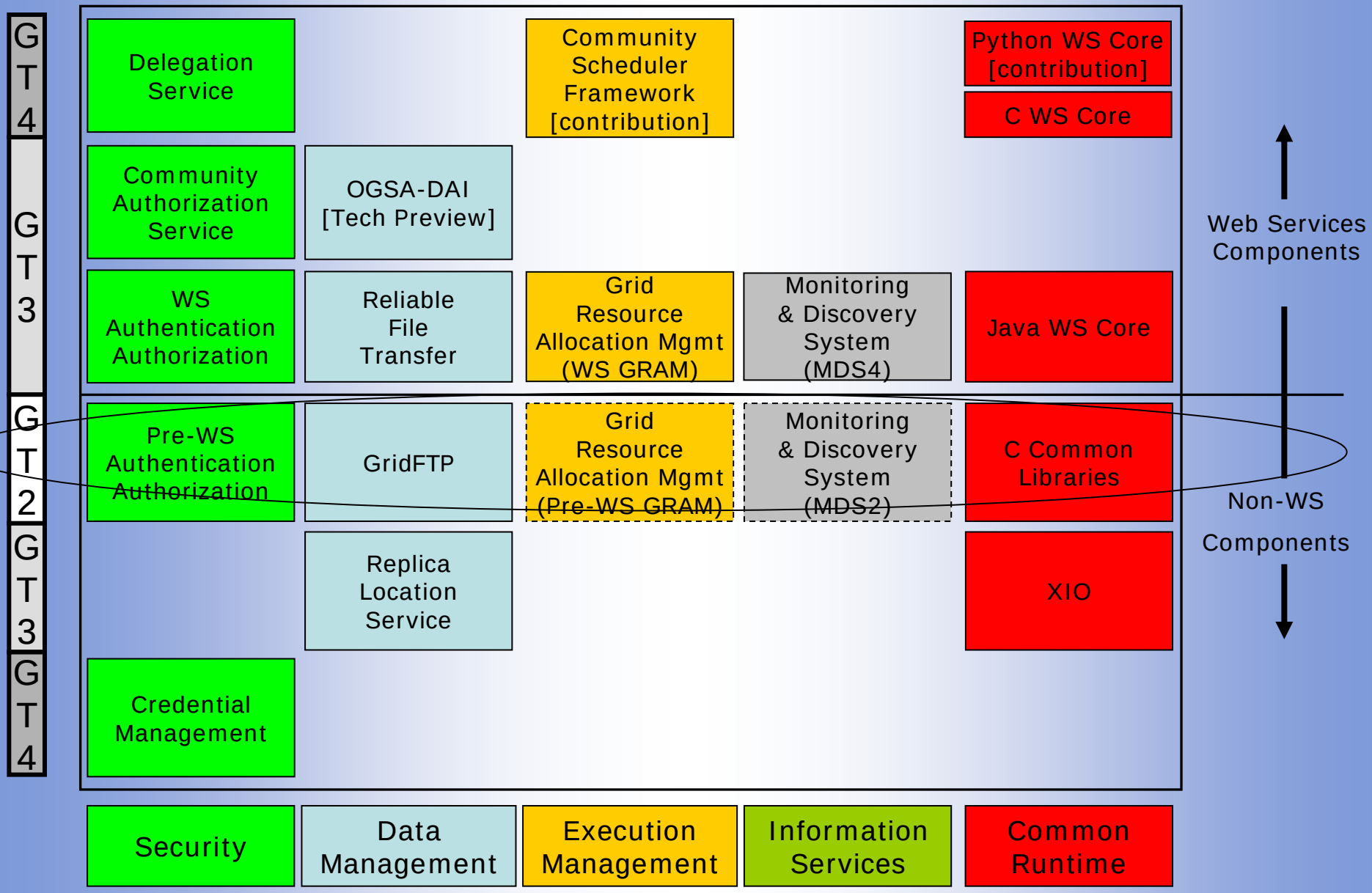
MILLERSVILLE UNIVERSITY Ville

the university of OKLAHOMA

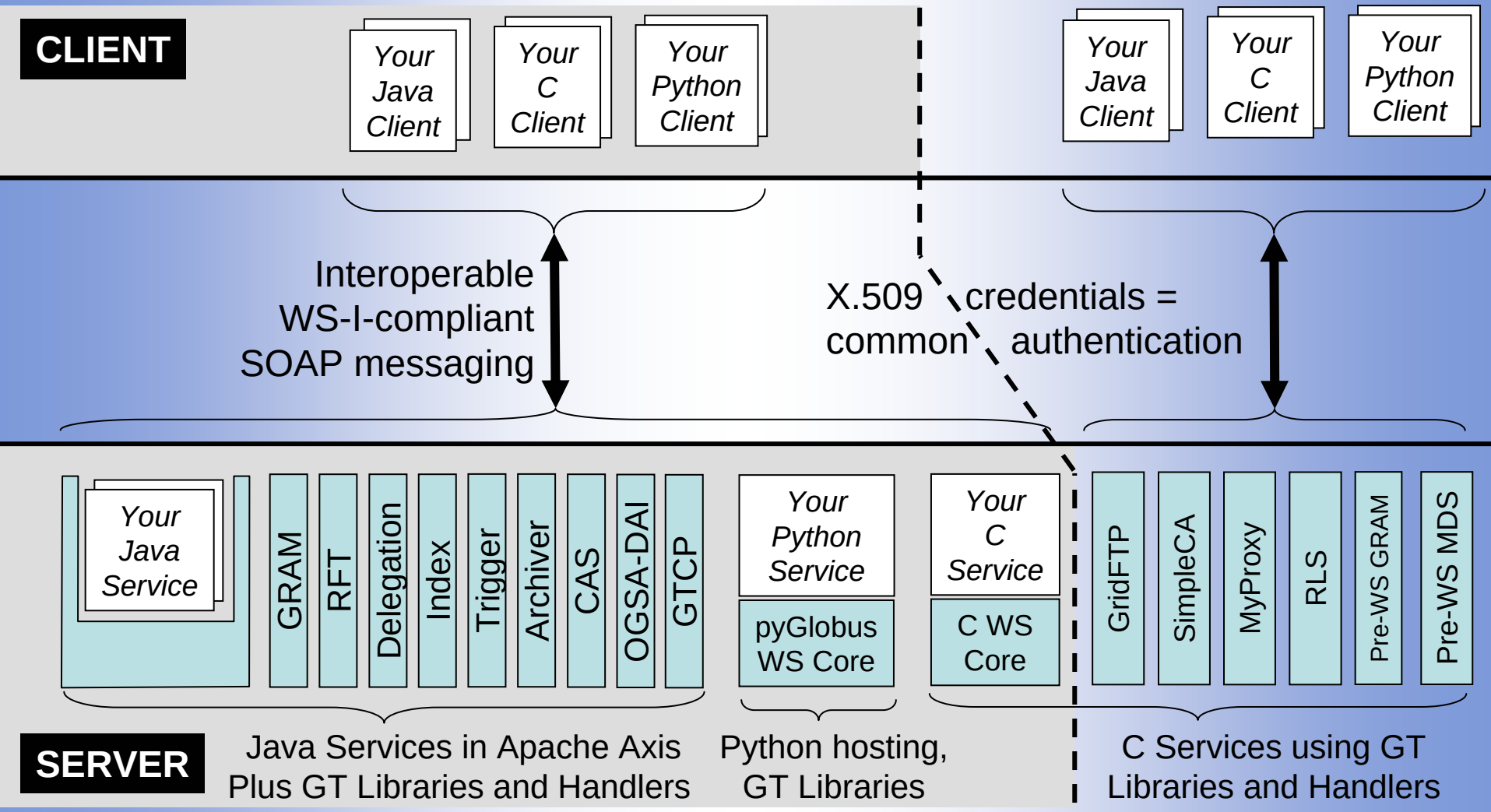
LEAD utilizes grid tools adopting a strategy to deal with middleware issues

- Stick with pre-ws globus version (Globus 2.4) and slowly transition to GT 4
- Develop software on self-controlled test grid constituting of machines distributed at various partner institutions
- port the tested version on to teragrid – production grid resources by working closely with the resource providers

Pre-WS Globus components are still supported



GT4 Components

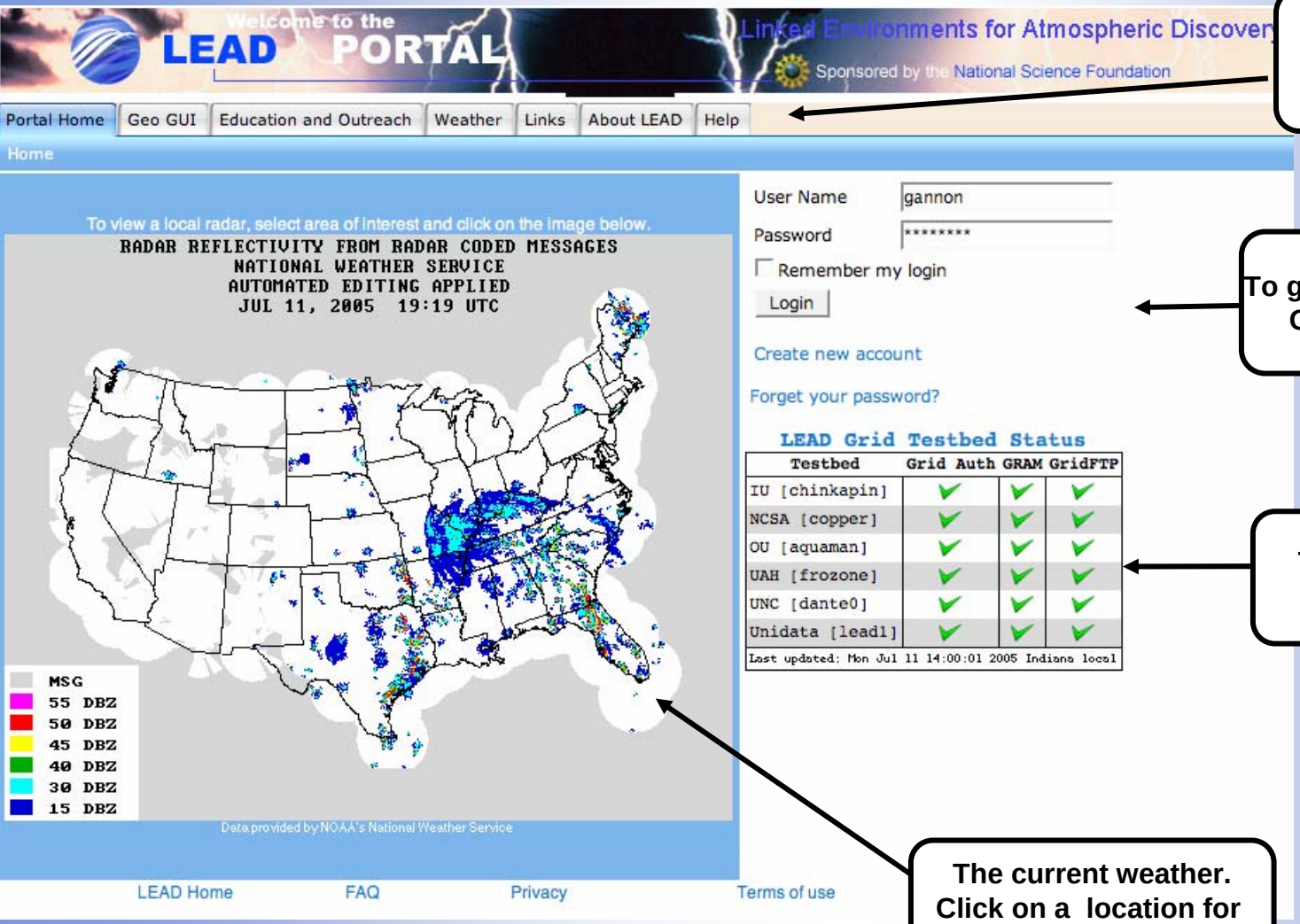


LEAD Test-bed Grid

DEMO

- LEAD Portal demo

The top level view



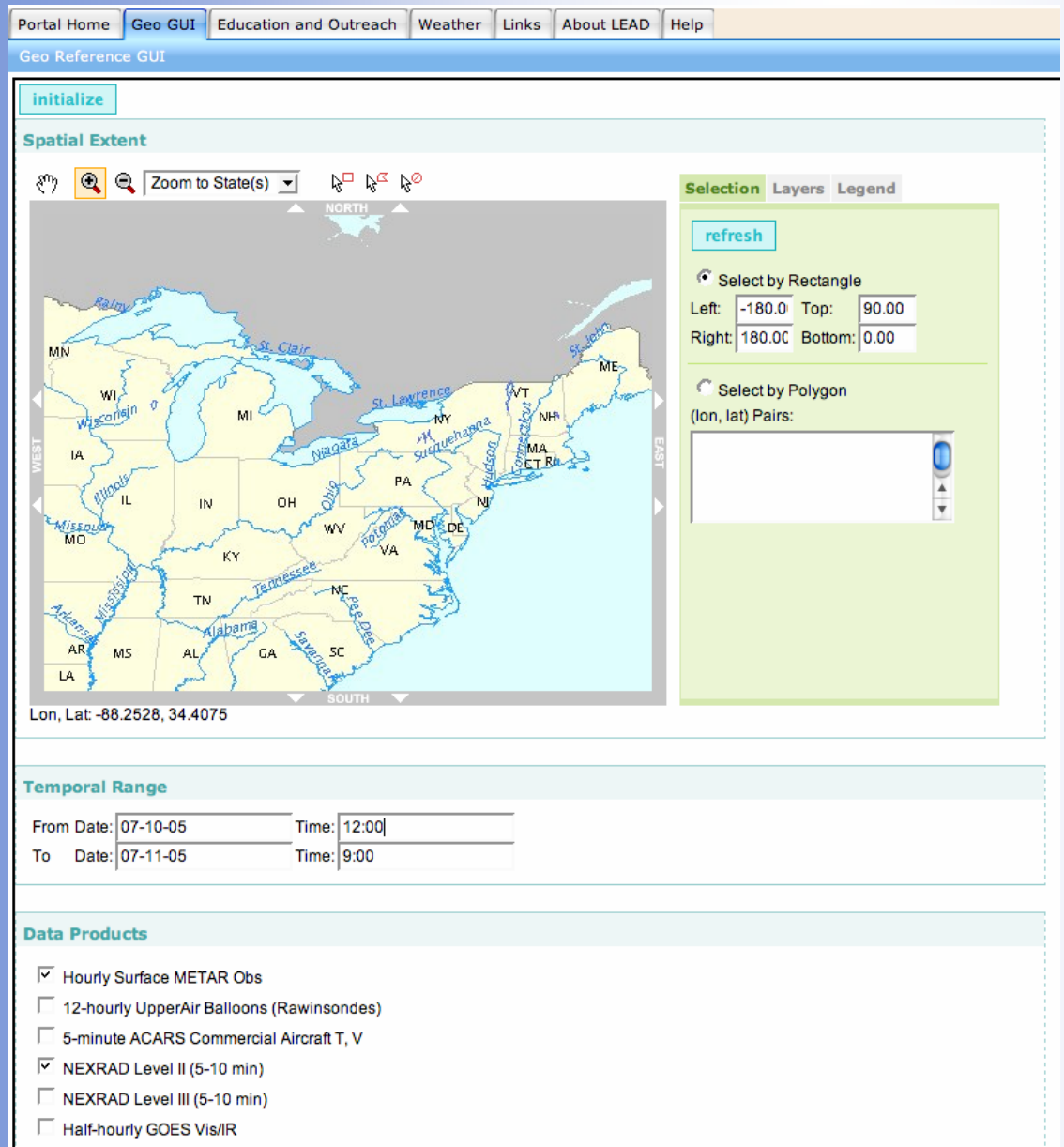
Top Level tabs to public Tools and information.

To get to your stuff, log in here Or create a new account

The current testbed status

The current weather. Click on a location for more data

GEO Reference GUI Prototype



- Use mouse to drag a region of interest.
- Fill in the data requirements
- The tool, when finished will gather the data for you.

Educational Resources

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

Education Home

LEAD Education and Outreach

Home | LEAD Learning Communities | Shared Workspace | News LEADers | LEAD Educator Workshops | Evaluation Rubric

Modules

LEAD Main Home

Lead E&O Vision

IDV Beginner's Tutorial

IDV Webstart Download

IDV Users Guide

Links to Related Material

LEAD is funded by the National Science Foundation

LEAD-TO-LEARN

EXPLORING TEMPERATURE CONTRASTS THROUGH VISUALIZATIONS

Purpose

To introduce students to visualization as a tool for scientific problem solving using temperature data from the NCEP Eta Model as an example.

Overview

In this learning activity, students will use the Integrated Data Viewer (IDV) tool to visualize temperature data from a numerical model output in order to explore the relationship between temperature contrasts, weather patterns, and geographical features so that important patterns become evident. The IDV will allow the student to visualize how temperature contrasts vary across North America, as well as how they cl

Student Outcomes

By exploring a LEAD
patterns in the model

Students can analyze
visualization as a tool

Science Concepts (N

General

Visual models help us

Earth Sciences

Visualizations help or
heights above the Ear

Geography

Geographical visualiz

Science Inquiry Abiliti

Identify answerable q
Design and conduct a
Use appropriate tools
Develop and construc

TEMPERATURE WORKSHEET-1.doc

LEAD-TO-LEARN

EXPLORING TEMPERATURE CONTRASTS THROUGH VISUALIZATIONS

WORK SHEET

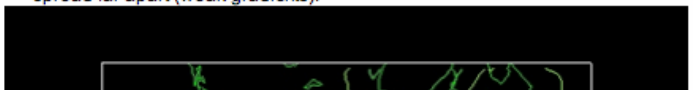
In this activity you will visualize temperature data from a numerical model and use it to explore the relationship between temperature contrasts, weather patterns, and geographical features.

PART A: Generating the visualization

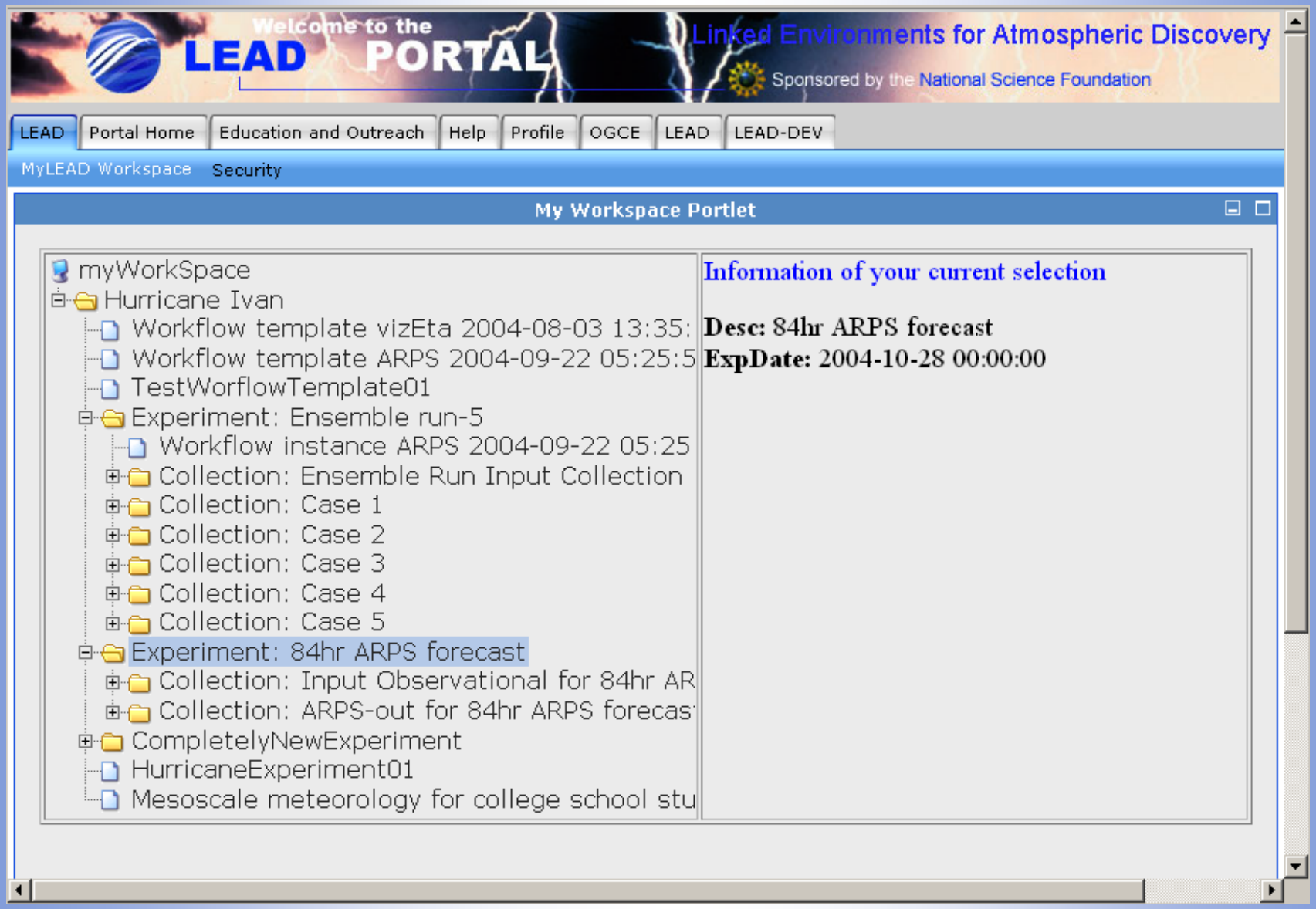
a. Launch IDV
b. Four windows should appear: 1) The Display Window, 2) The Data Selector Window, 3) Parameter Control Window, and the User's Guide.

PART B: Orienting and Problem Solving with IDV

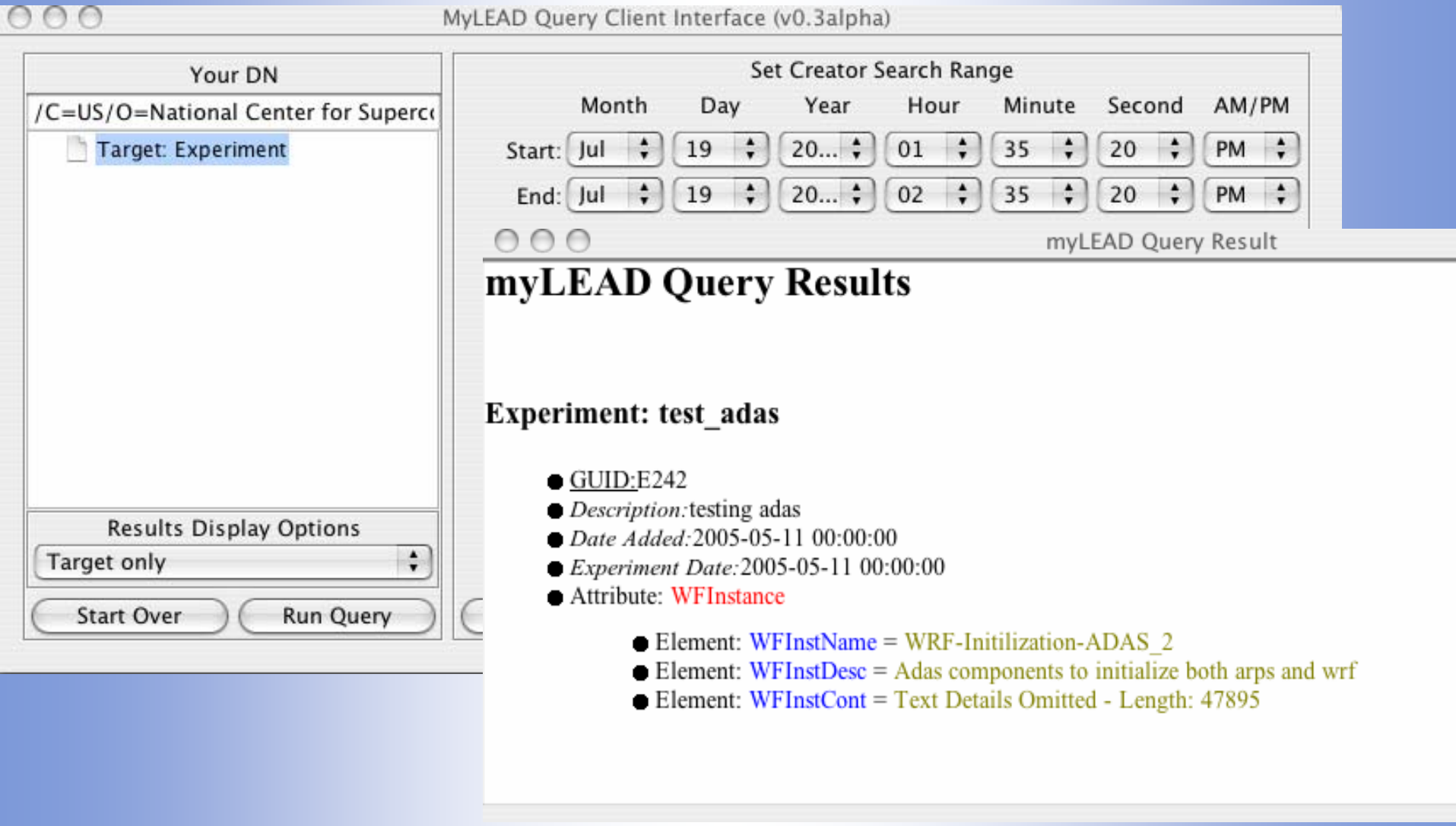
a. You will notice that there are regions where the temperature contours are spaced closely together (strong gradients), and others where they are spread far apart (weak gradients).



Log in and see your MyLEAD Space

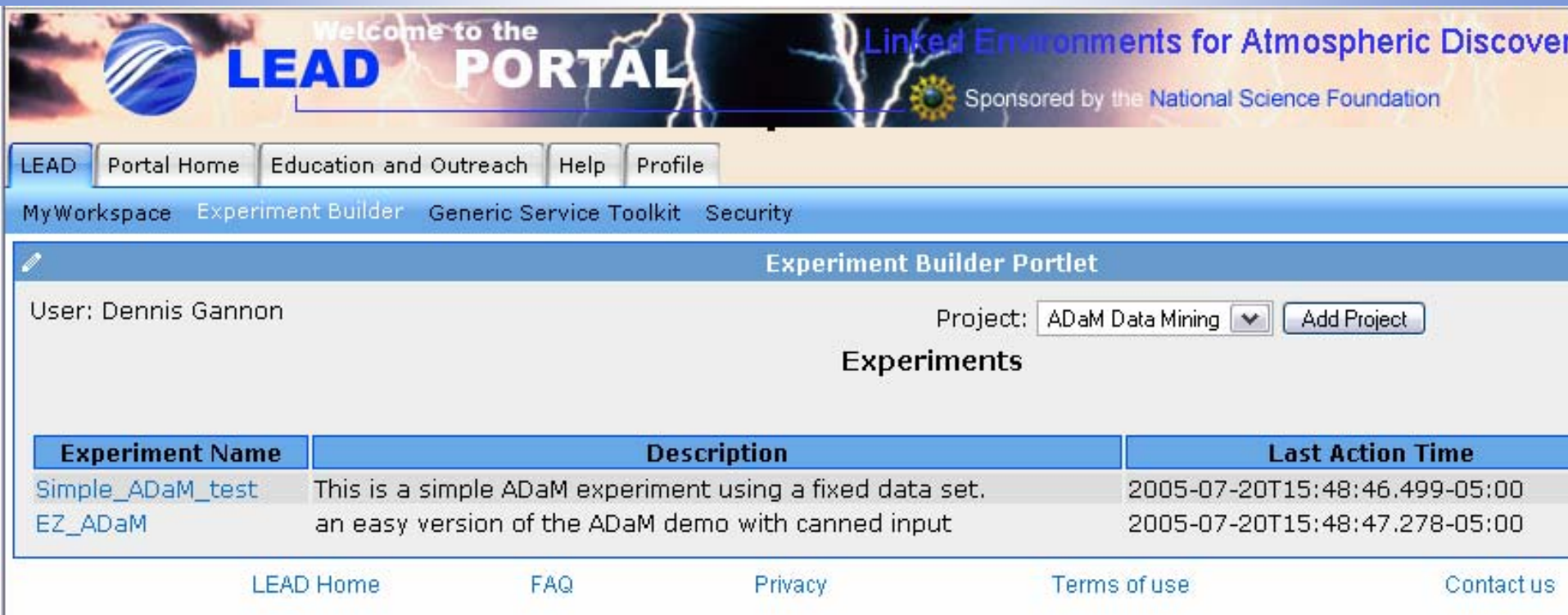


Searching MyLEAD



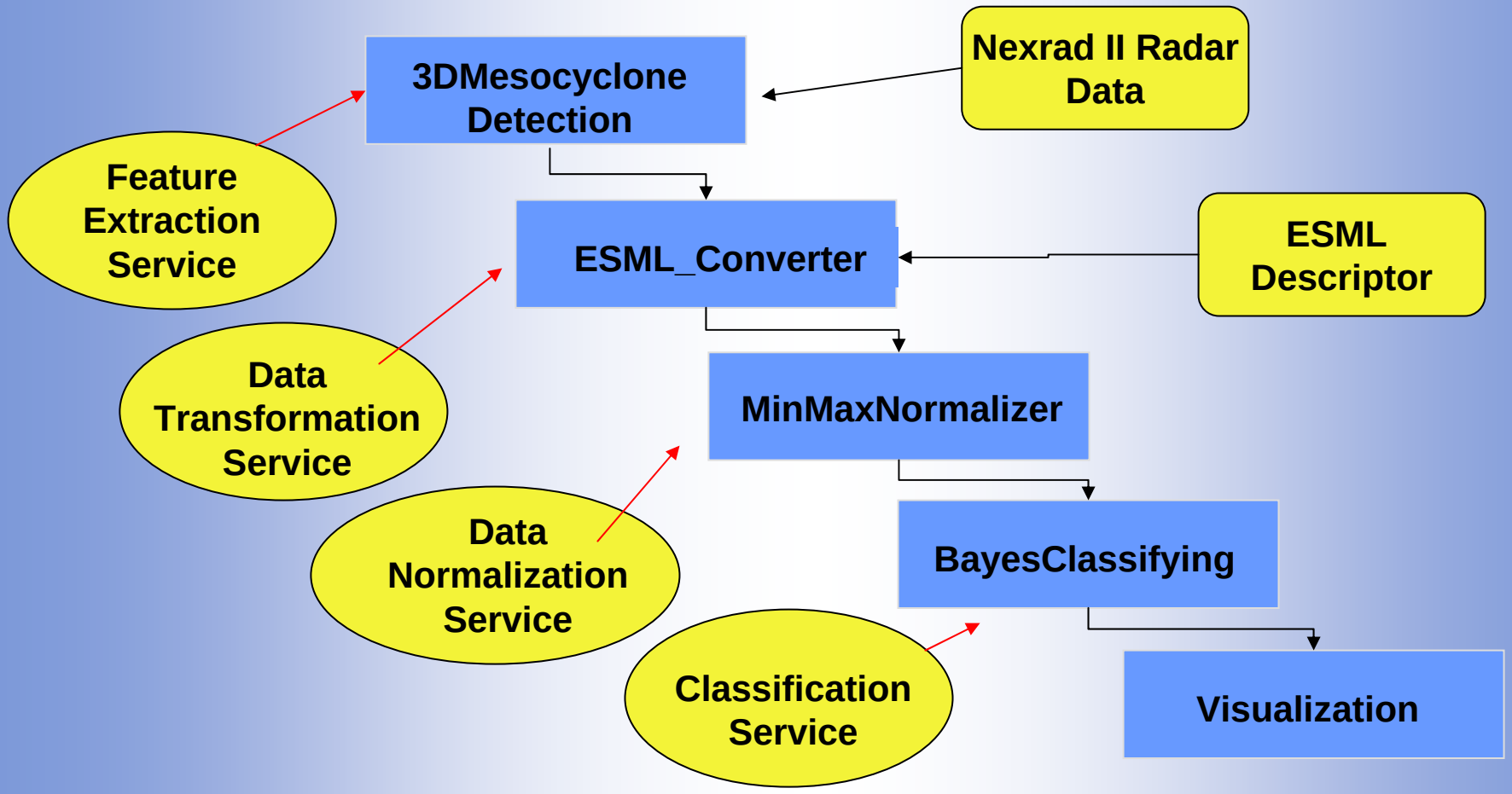
The Experiment Builder

- To review your previous experiments and create new ones
- Experiments are organized into projects
  - You can select an old one to look at,
  - Or create a new project or experiment.
  - Let's do a new experiment! (click "new")



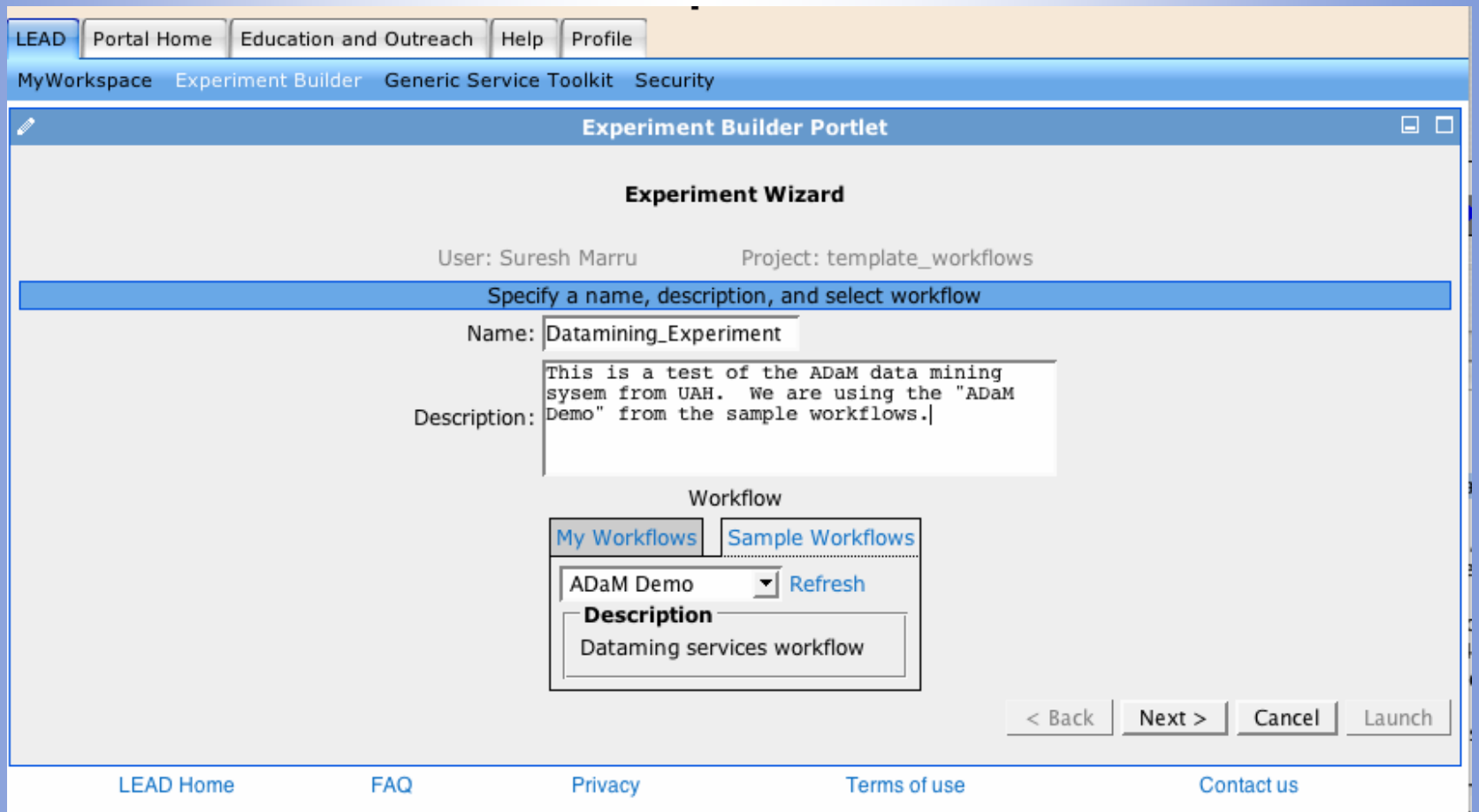
Creating a workflow for Data Mining

- Use ADaM services from UAH



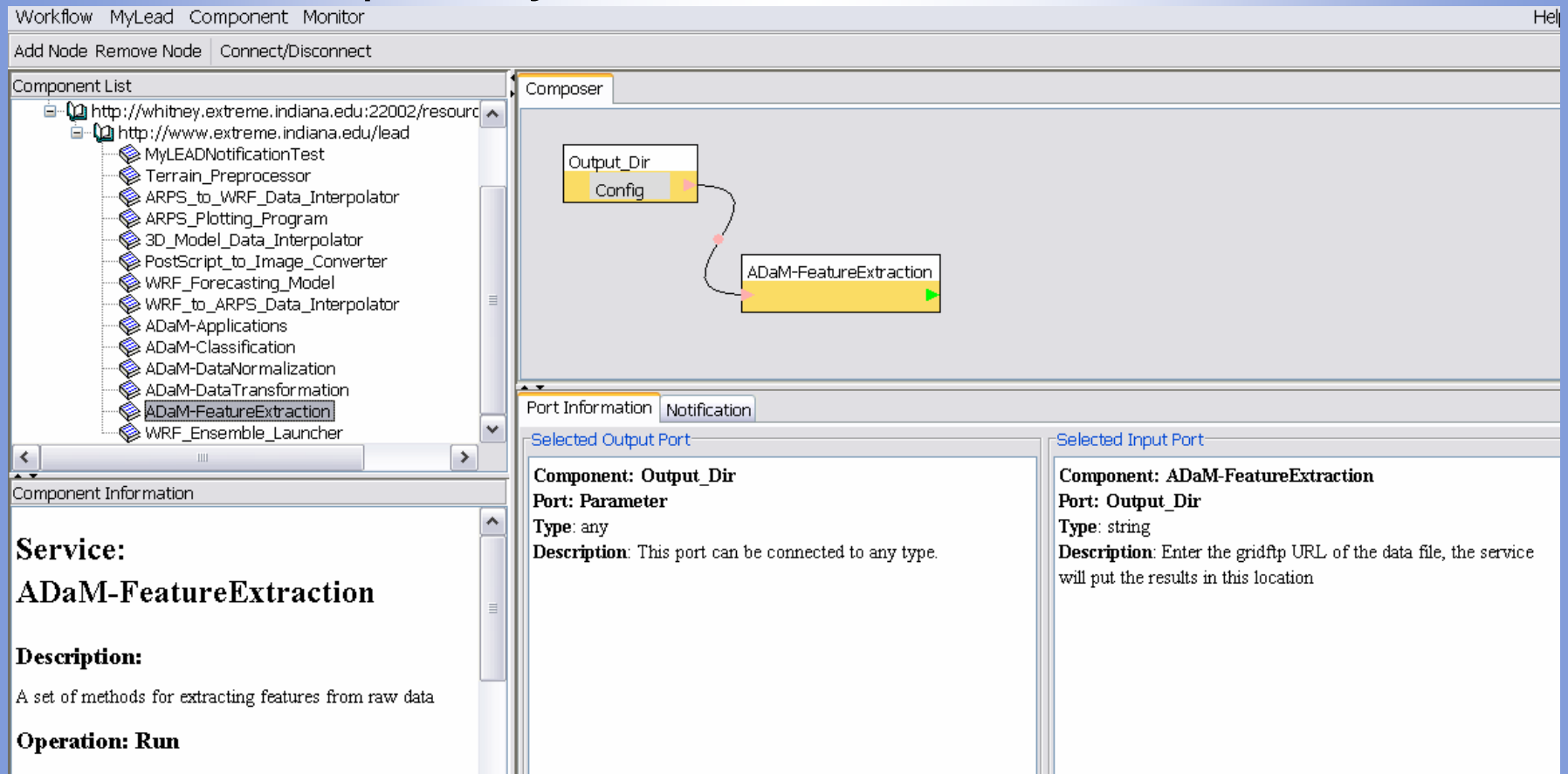
Provide a name and description

- Next select an application from the dropdown list or create a new workflow.
- Once we have selected the app, we push “next” to add data.



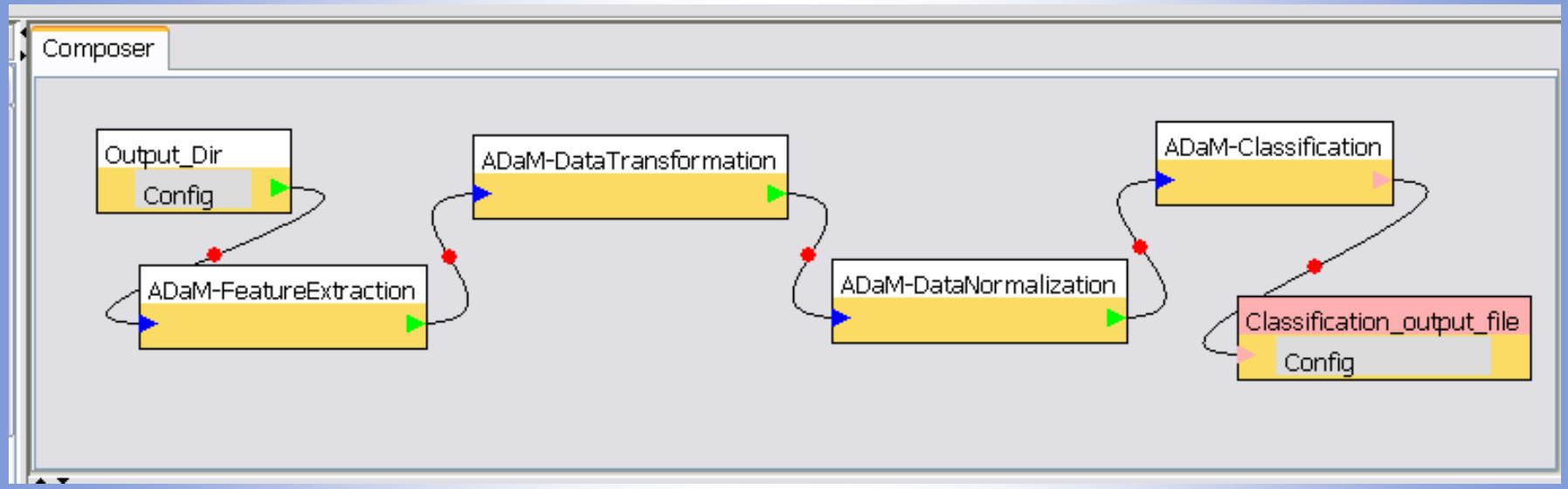
Composing the Workflow

- Graphical Composer
  - Standard drop-and-drag composer model (like Kepler and others)
  - Compiles Python or PBEL code

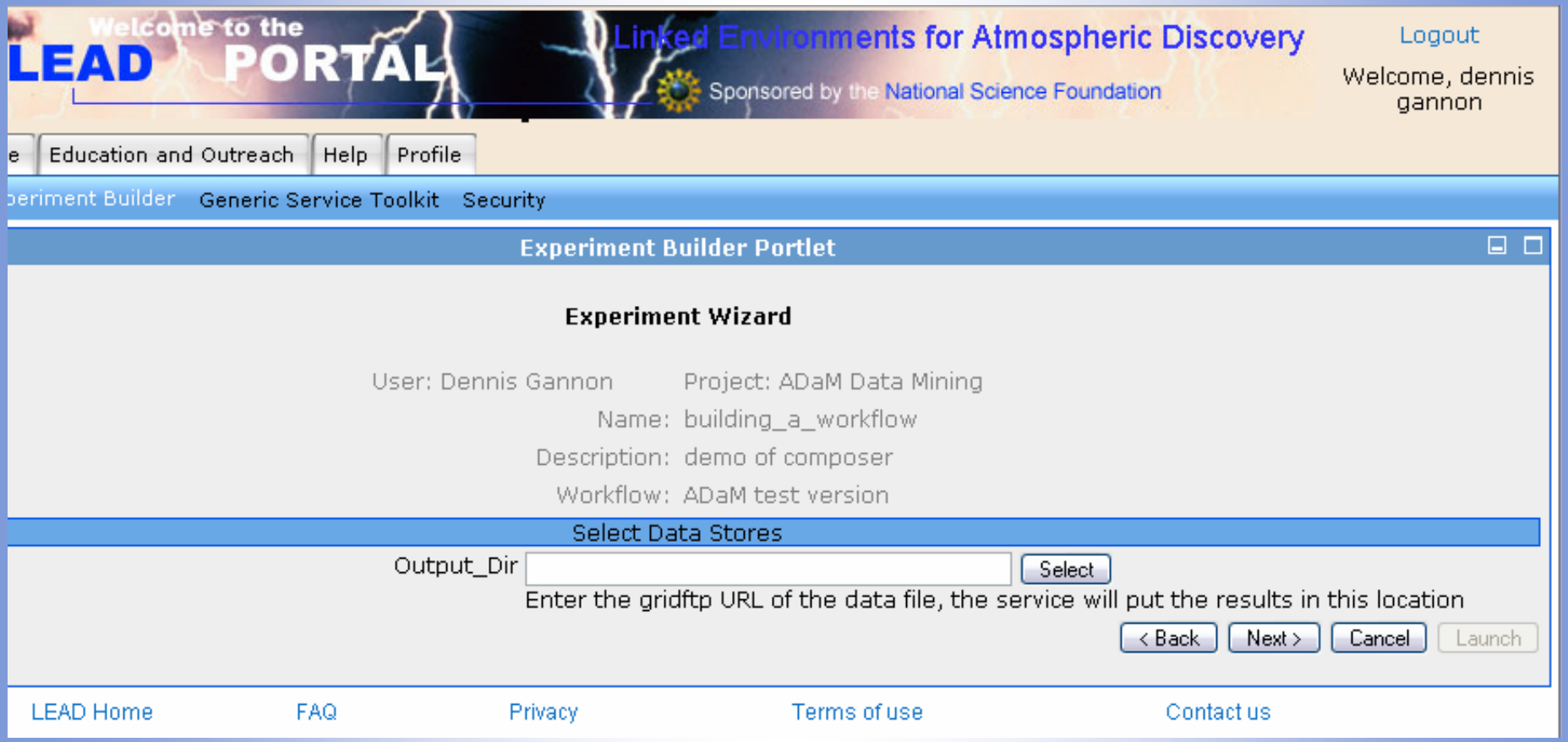


Final Workflow

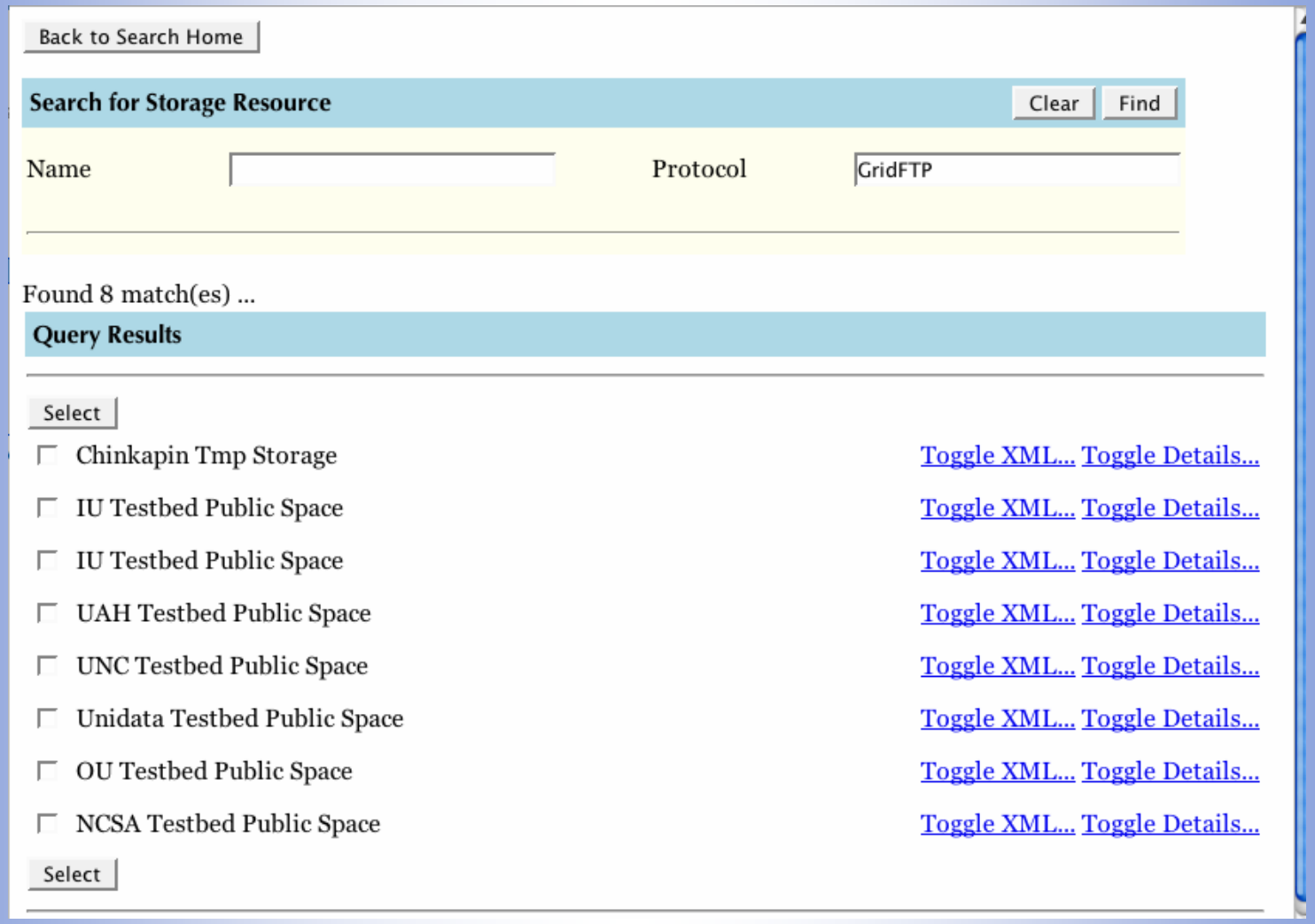
- Save it back to my lead
- Next we must bind the inputs to the workflow



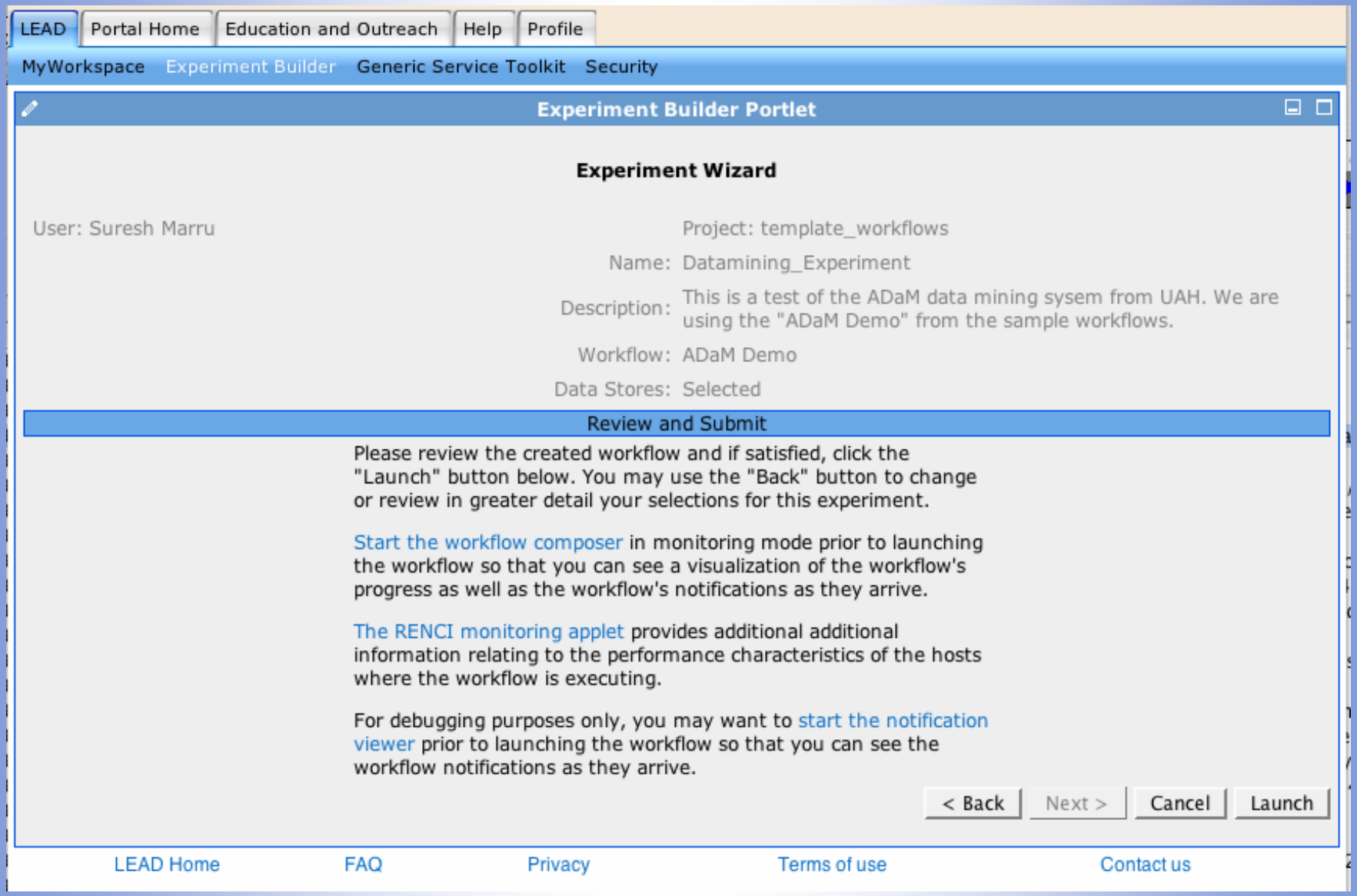
Wizard understand the workflow requirements



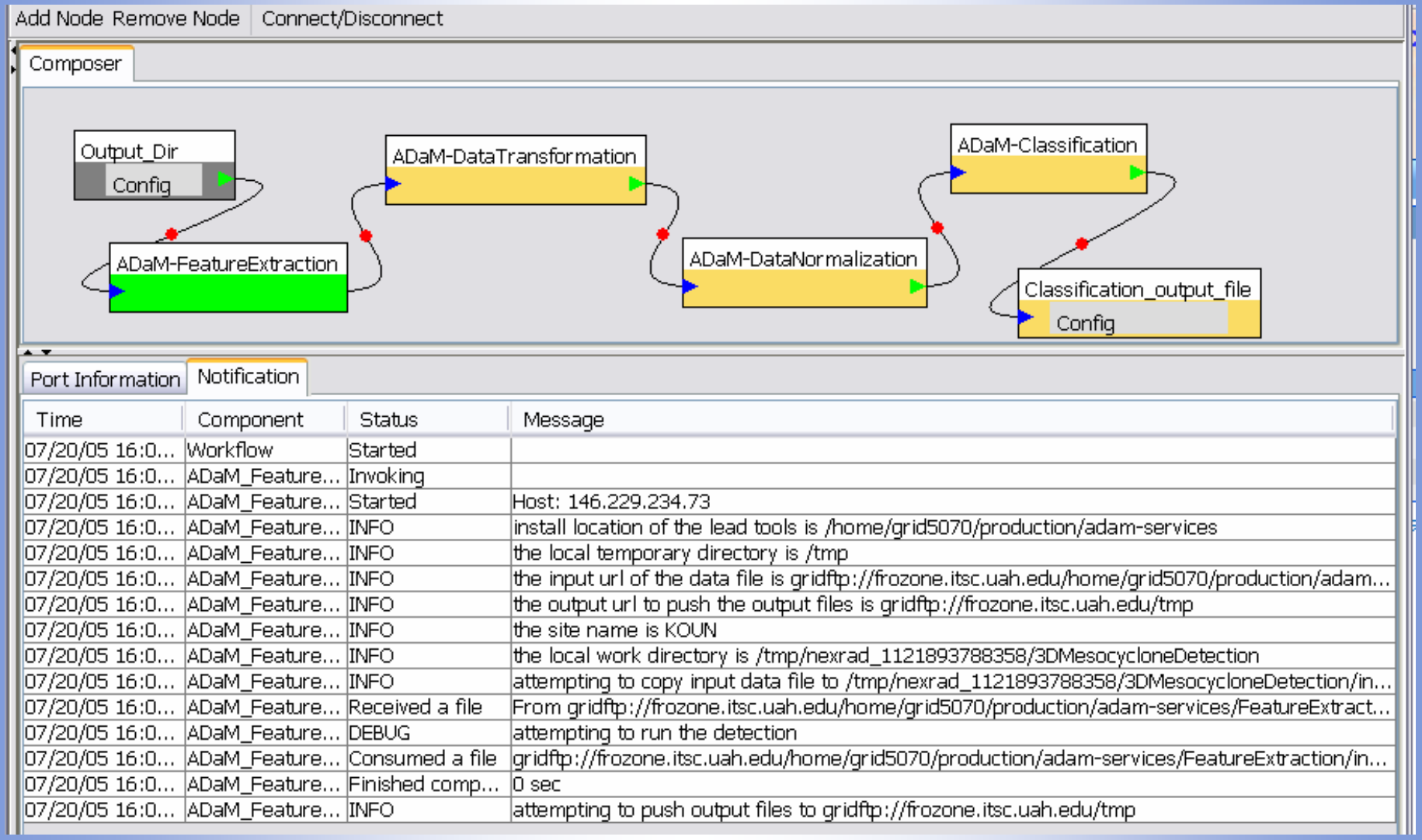
Select an output location



Submitting the workflow



Monitor results in real time



Time	Component	Status	Message
07/20/05 16:0...	Workflow	Started	
07/20/05 16:0...	ADaM_Feature...	Invoking	
07/20/05 16:0...	ADaM_Feature...	Started	Host: 146.229.234.73
07/20/05 16:0...	ADaM_Feature...	INFO	install location of the lead tools is /home/grid5070/production/adam-services
07/20/05 16:0...	ADaM_Feature...	INFO	the local temporary directory is /tmp
07/20/05 16:0...	ADaM_Feature...	INFO	the input url of the data file is gridftp://frozone.itsc.uah.edu/home/grid5070/production/adam...
07/20/05 16:0...	ADaM_Feature...	INFO	the output url to push the output files is gridftp://frozone.itsc.uah.edu/tmp
07/20/05 16:0...	ADaM_Feature...	INFO	the site name is KOUN
07/20/05 16:0...	ADaM_Feature...	INFO	the local work directory is /tmp/nexrad_1121893788358/3DMesocycloneDetection
07/20/05 16:0...	ADaM_Feature...	INFO	attempting to copy input data file to /tmp/nexrad_1121893788358/3DMesocycloneDetection/in...
07/20/05 16:0...	ADaM_Feature...	Received a file	From gridftp://frozone.itsc.uah.edu/home/grid5070/production/adam-services/FeatureExtract...
07/20/05 16:0...	ADaM_Feature...	DEBUG	attempting to run the detection
07/20/05 16:0...	ADaM_Feature...	Consumed a file	gridftp://frozone.itsc.uah.edu/home/grid5070/production/adam-services/FeatureExtraction/in...
07/20/05 16:0...	ADaM_Feature...	Finished comp...	0 sec
07/20/05 16:0...	ADaM_Feature...	INFO	attempting to push output files to gridftp://frozone.itsc.uah.edu/tmp

Check it out in MyLEAD

Experiment Builder Portlet

User: Suresh Marru

Project: demo-demo-run

Experiment Status Page

Experiment Details

Name: Datamining-demo

Description: Datamining demo

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

Status: INFORMATION

Workflow

Name: ADaM Demo

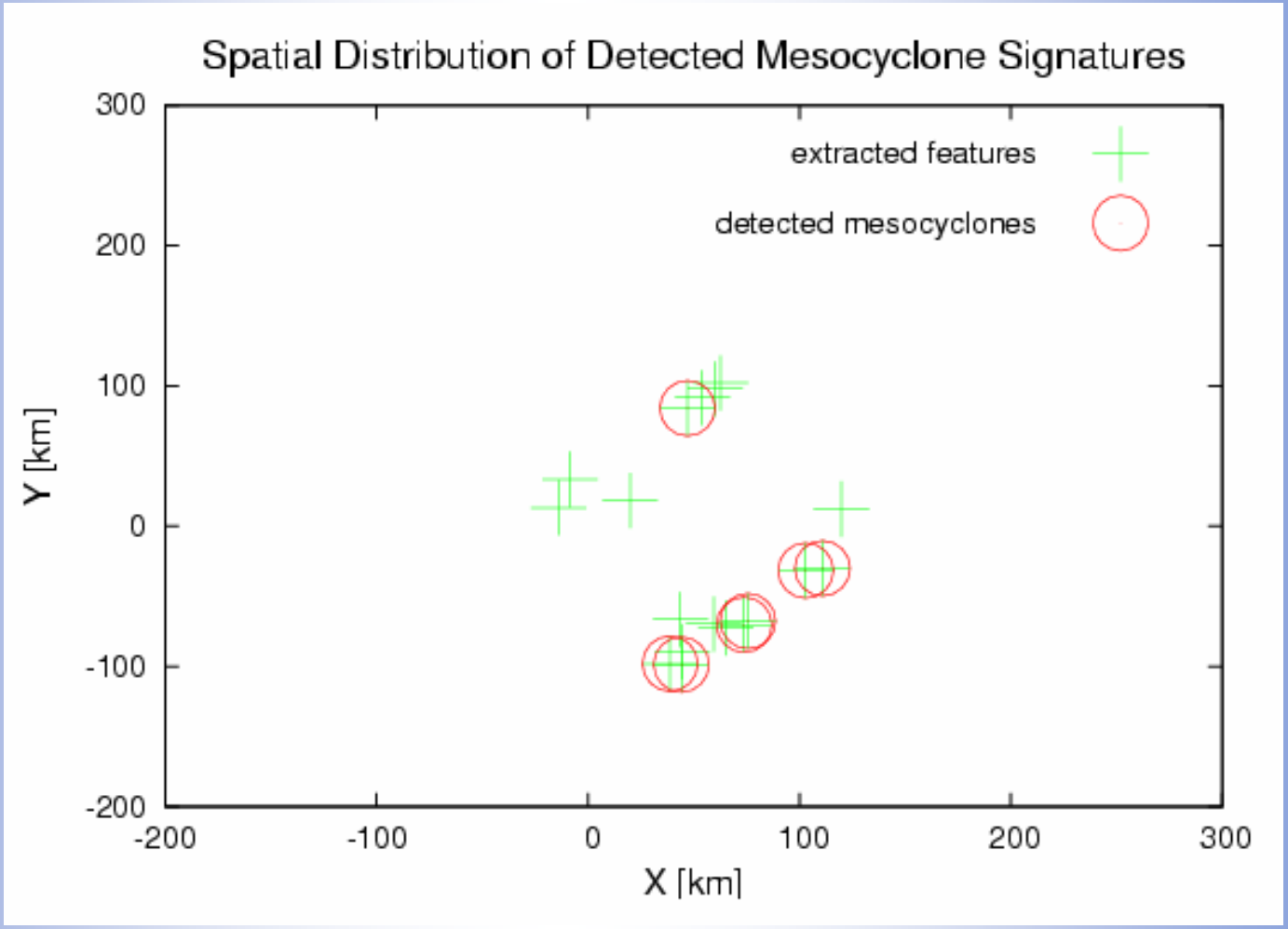
Description: Dataming services workflow

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

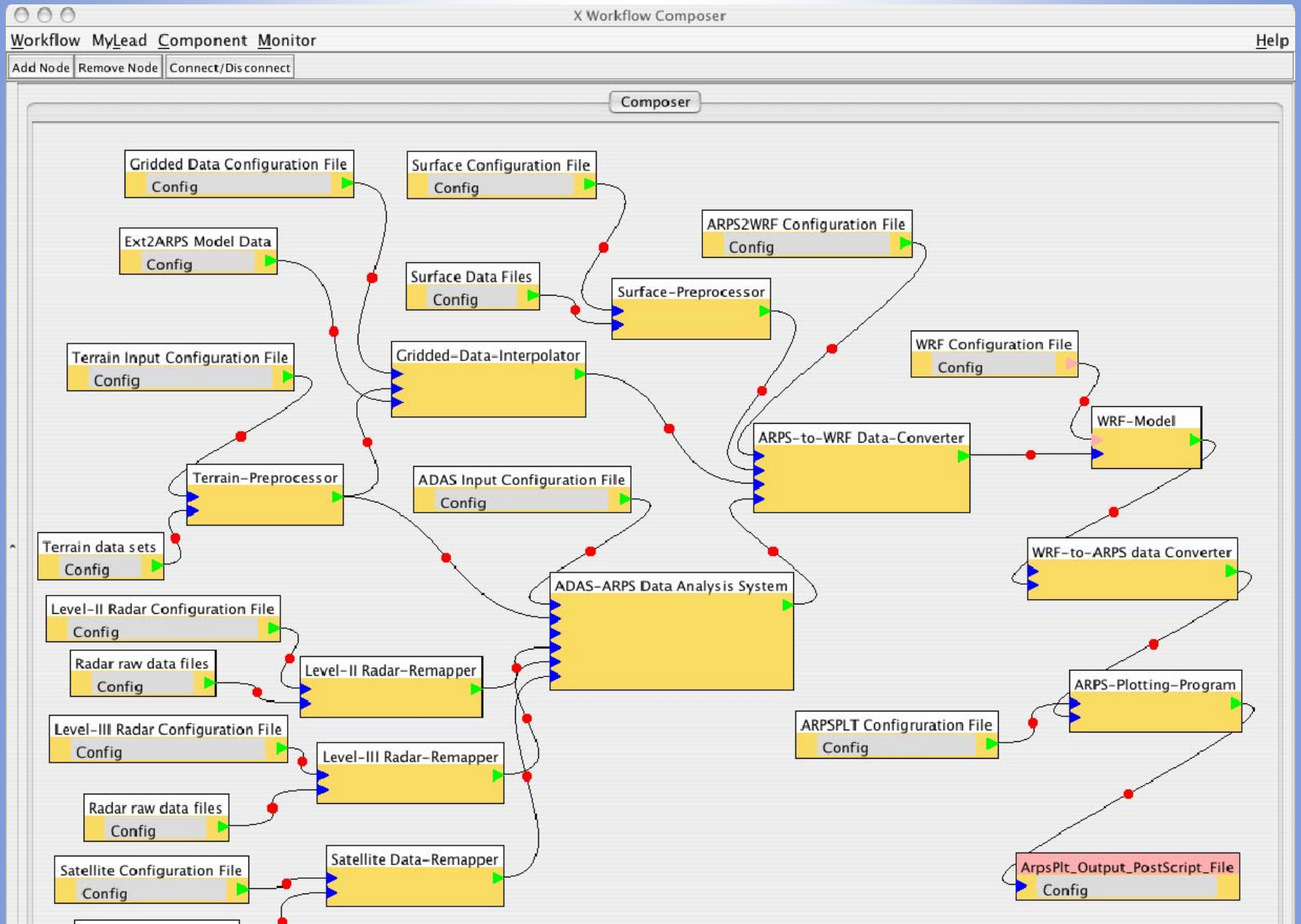
Notifications

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

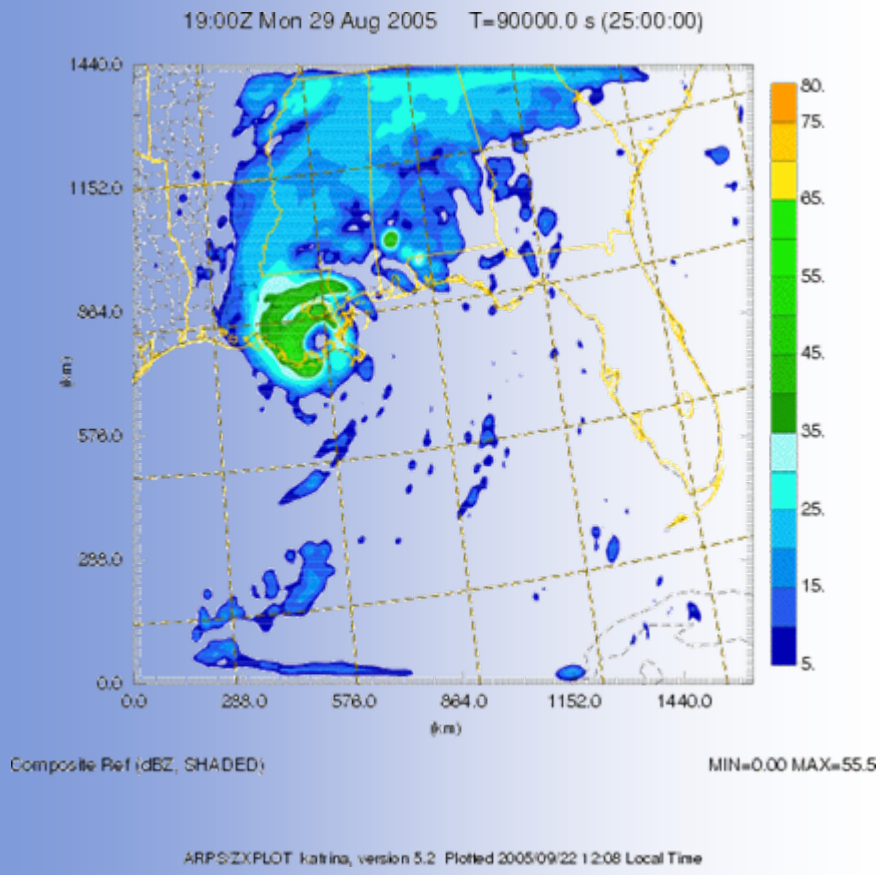
Click on the output file to see visualization

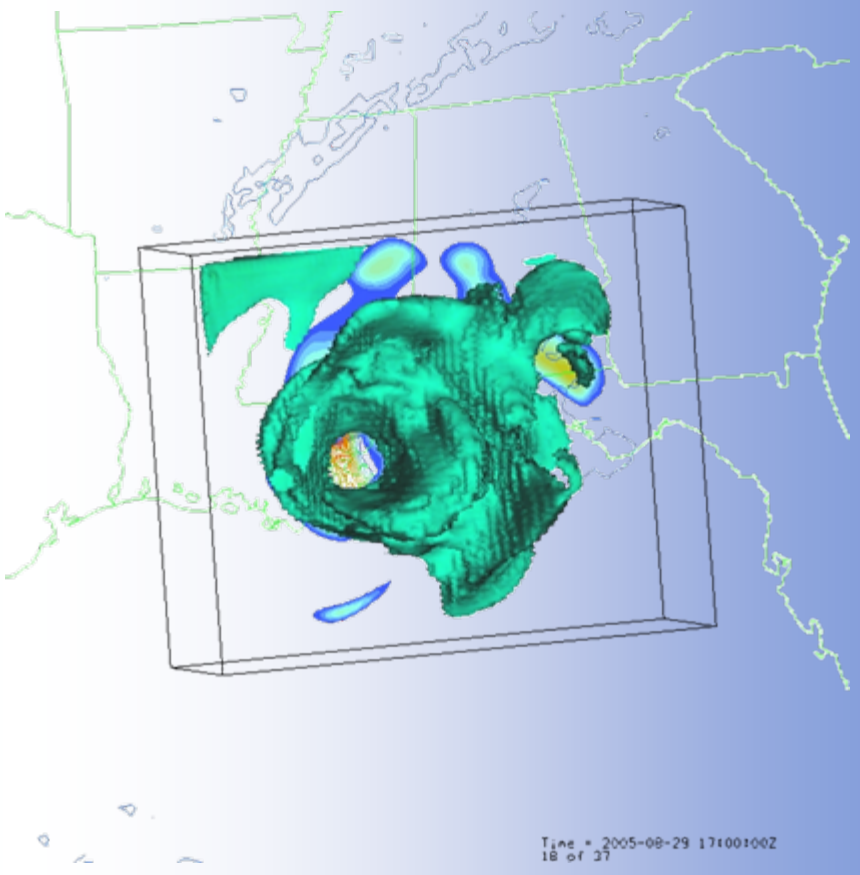


Large workflows can be composed



Output from the Weather Workflow





Acknowledgements

Slide Courtesy:

- Dr. Dennis Gannon
- Globus Website (http://globus.org)

? ? ?

Questions ?