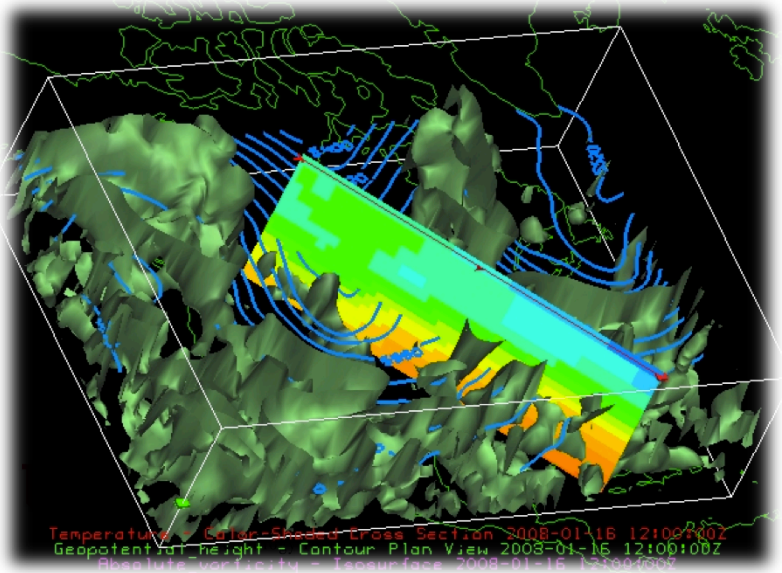
LEAD II hybrid workflows for timely weather products



Beth Plale
Director, Data To Insight Center
School of Informatics and Computing
Indiana University

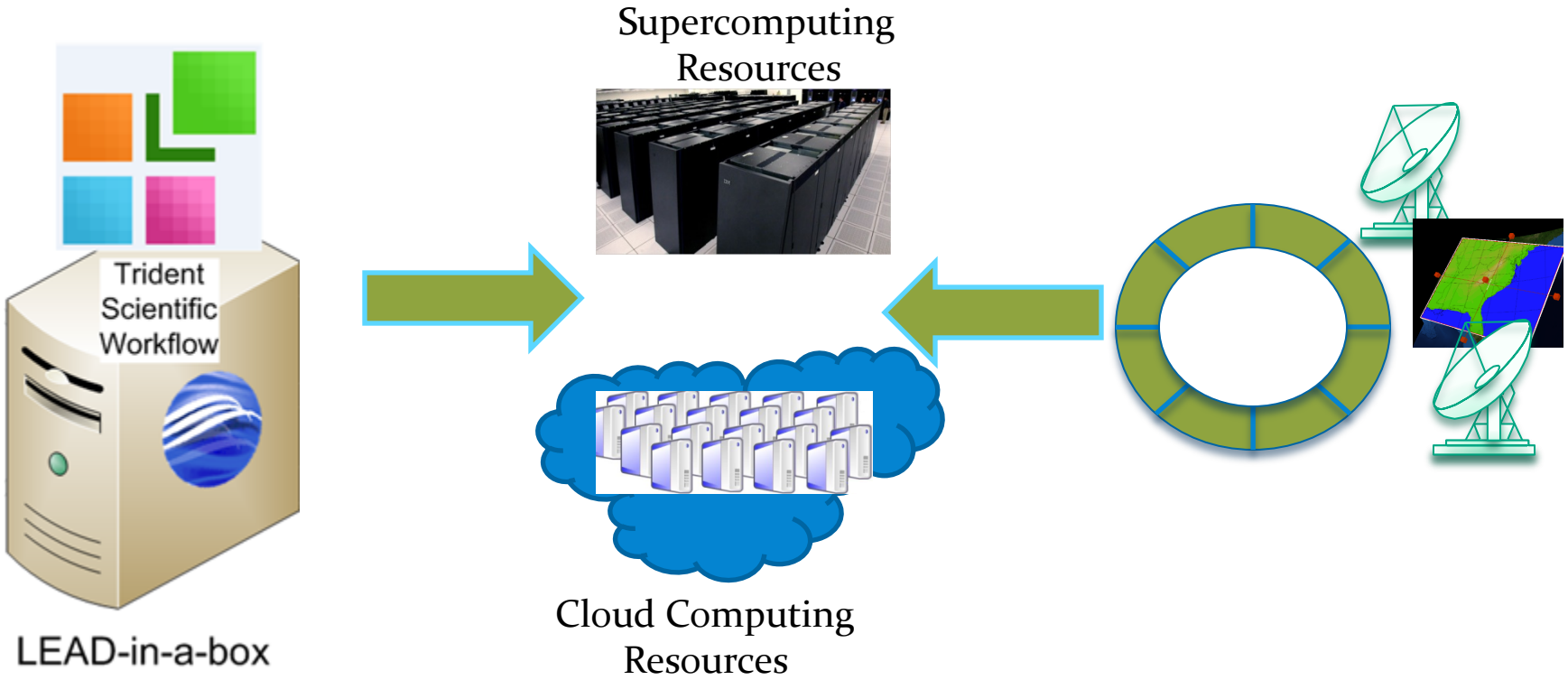
Scope: tools facilitating use of mesoscale meteorological data

- Mesoscale weather:
  - atmospheric phenomena such as thunderstorms, squall lines, fronts, and precipitation bands
  - Scales ranging in area from size of small city to size of state such as Iowa.
- Our effort, which we call LEAD II, draws off and extends Linked Environments for Atmospheric Discovery (LEAD) project 2003-2009

NRC report, 2008, "Observing Weather and Climate from Ground Up"

Goal: tools to enable immediate response to volumes of real time data

- Utilizing Trident Scientific Workflow Workbench



Current focus: Vortex 2

VORTEX 2 2009 • 2010

NSF NOAA

Welcome

• • • What is VORTEX2?

VORTEX2 is by far the largest and most ambitious effort ever made to understand tornadoes. We expect over 100 scientists and over 40 science and support vehicles to participate in this unique, fully nomadic, field program during its second and last field season, May/June 2010. The National Science Foundation (NSF) foundation and the National Oceanic and Atmospheric Administration (NOAA) are contributing over $10 million towards this effort. Participants will again be drawn from over a dozen universities, and several government and private organizations. International participants will be drawn from Italy, Netherlands, United Kingdom, Germany, Canada and Australia.

The basic questions driving VORTEX2 are simple to ask, but hard to answer:

The shared Mobile Atmospheric Research and Teaching radar heads toward the Texas Caprock in the panhandle.

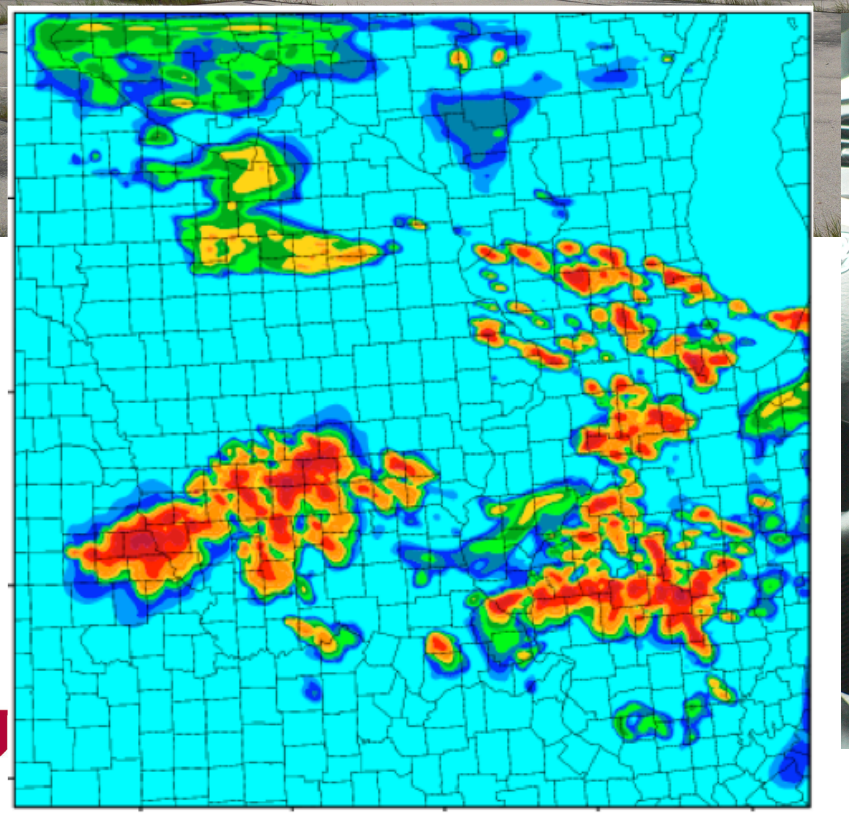
The challenge of observing

Fully nomadic field program studying tornadoes during May/June 2010

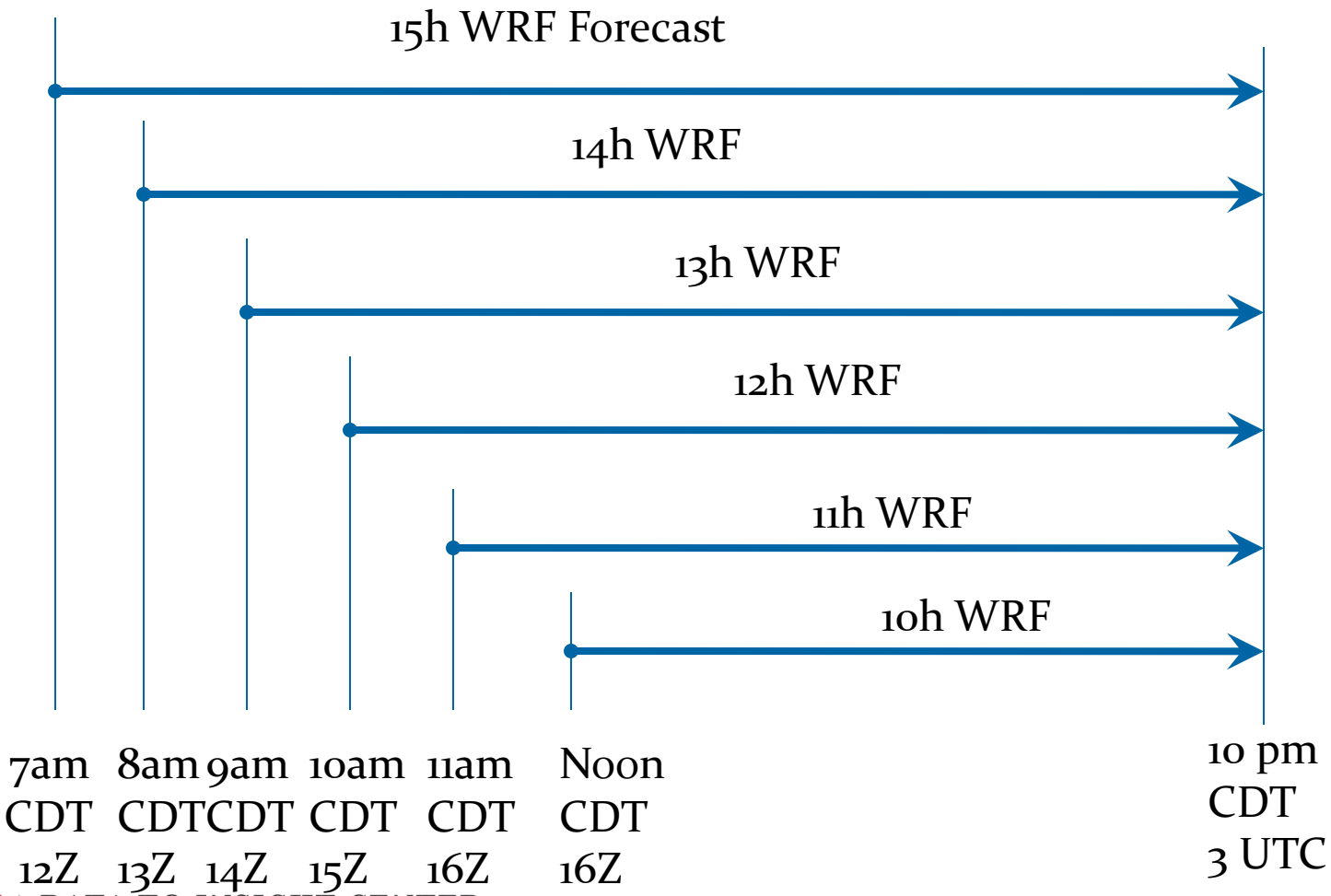
VORTEX2 2010

- May 1, 2010 to June 15, 2010
- ~6 weeks, 7-days per week
- Some down days when weather is calm
- With Oklahoma University (Keith Brewster) and Univ North Carolina Chapel Hill (Craig Mattocks)

Nomadic field effort to gather data about tornadoes



Six morning runs each of increasingly shorter term forecast

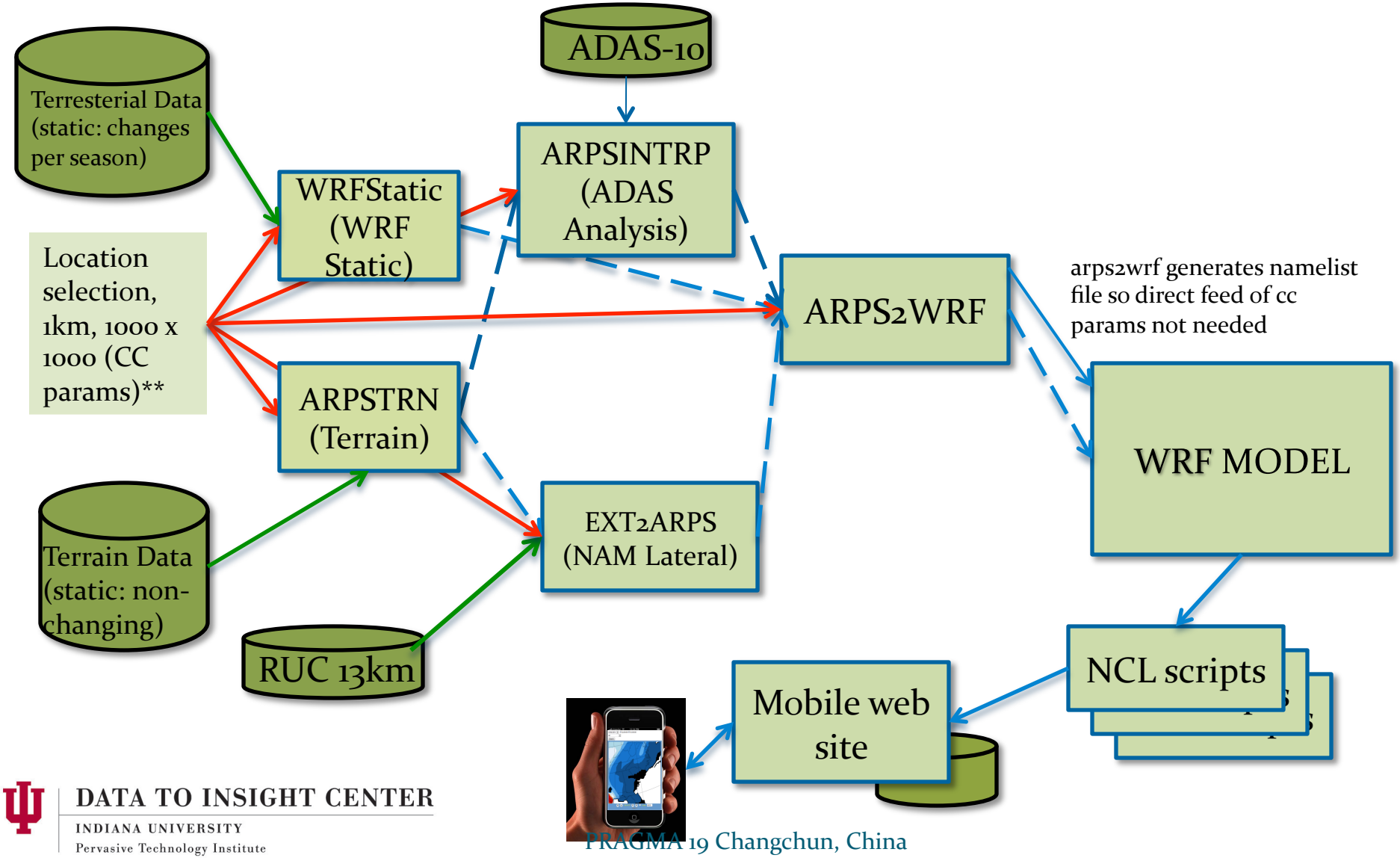


VORTEX2 2010 Workflow

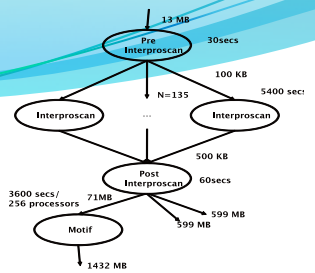
- User-Selected Domain Center
  - Our field guy will advise daily on area of interest
- 1000x1000 4-km WRF Domain
- Initialized with ADAS-10 and RUC-13
  - ADAS-10 produced by Oklahoma University;
  - RUC-13 product of NOAA/ NCEP operational weather prediction system running every hour out to at least 18h.
- Our edge? Oklahoma Univ's assimilation data gives access to radar data from the morning of and better cloud modeling.

The Vortex2 Workflow

ADAS-10 data:
gsiftp://gridftp.bigred.iu.teragrid.org:2812//N/dc/projects/lead/ldm/pub/other/lead/ADAS/10km/ad2010040621.nc



Workflow Categorization



I. Size

- Total Number of Tasks
- Number of Parallel Tasks – max number of parallel tasks (width)
- Longest Chain - number of tasks in longest chain
- Our workflows: not large on any of these dimensions.

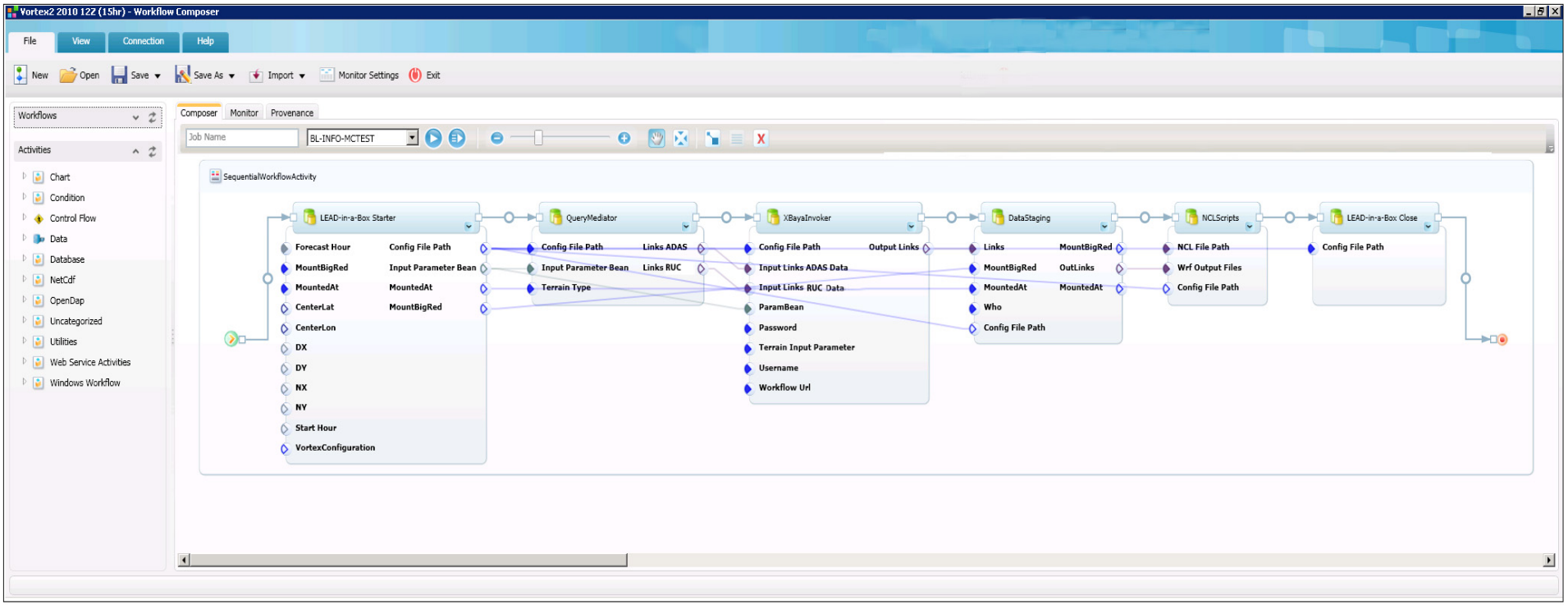
II. Resource Usage

- Max task processor width – max concurrent processors.
  - 16 on Windows box, >1000 on Big Red
- Total Computation time
  - Time budget is one hour; still working on optimal distribution.
- Data Sizes – sizes of inputs, outputs and intermediate data products
  - Ours is < 100 GB.

III. Structural pattern

- Sequential
- Parallel
- Parallel-split - one task's output feeds to multiple tasks.
- Parallel-merge - multiple tasks merge into one task.
- Parallel-merge-split - parallel-merge and parallel-split.
- Mesh - task dependencies are interleaved.
- Our workflows: Several nodes are parallel but merge within node before control transferred.

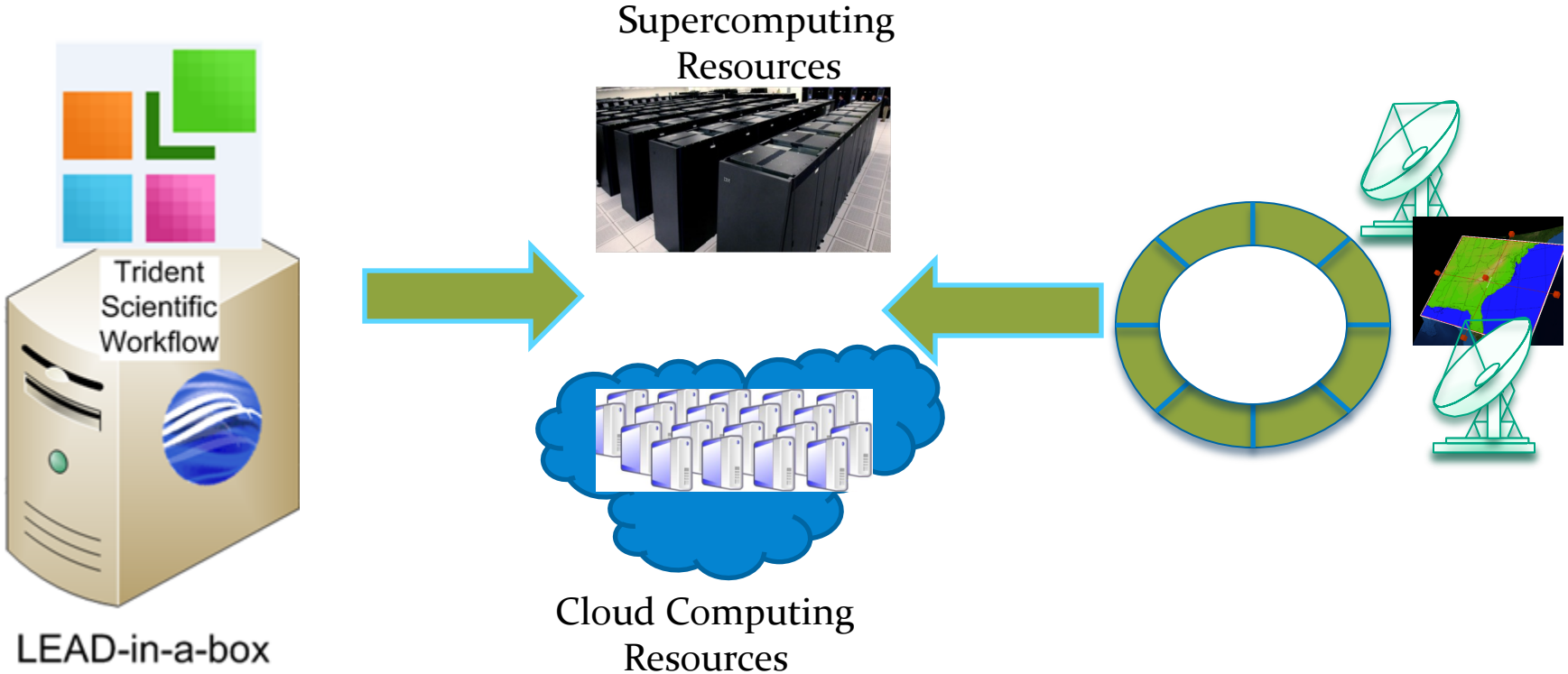
Workflows driven by Microsoft Trident Scientific Workflow Workbench



Research interest: ***Hybrid workflow model*** : workflows driven by Trident, but utilizing heterogeneous compute resources on back end

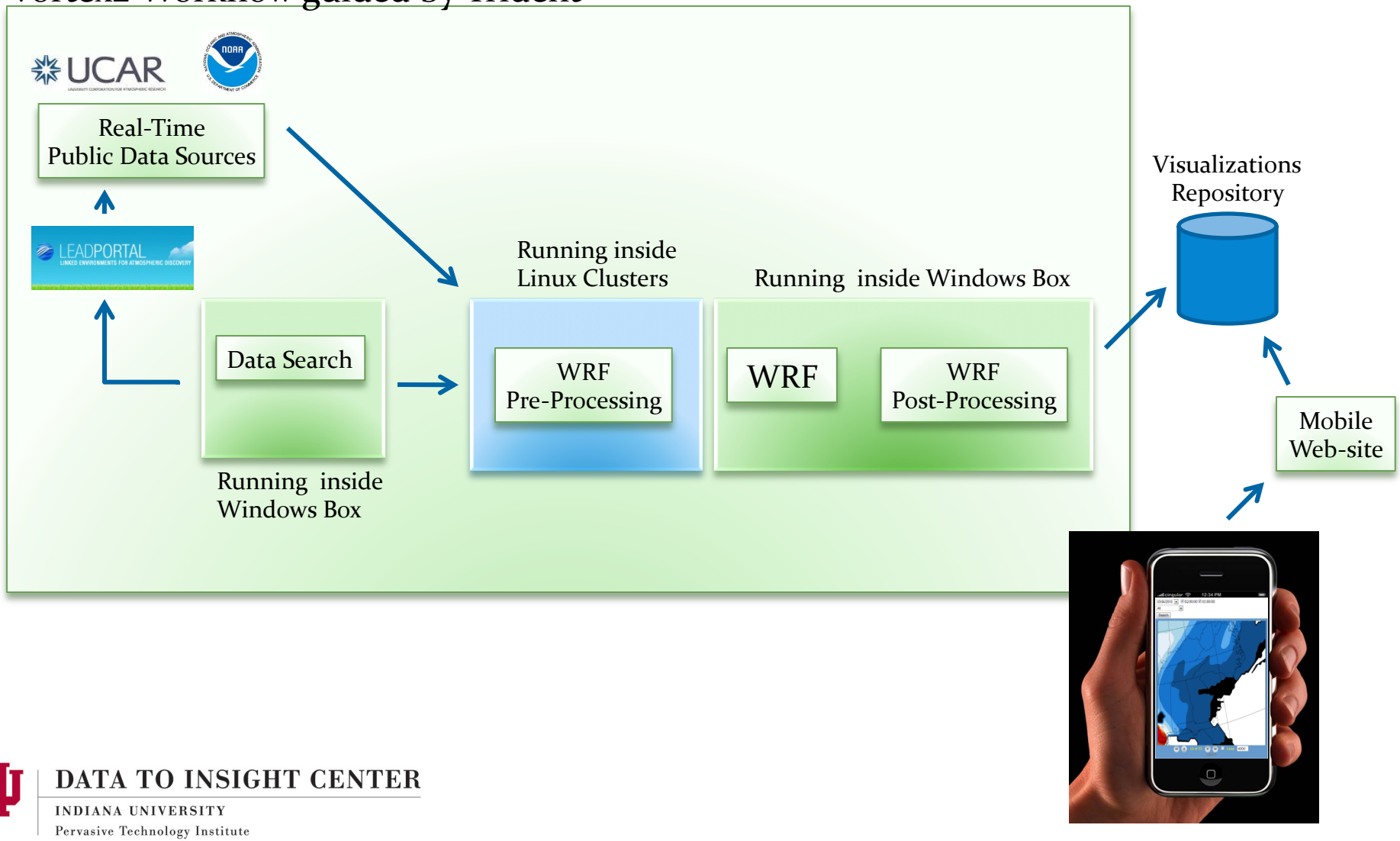
The High Level

- Trident activities control run of jobs on large-scale computing resources. Key: access to real time data sources

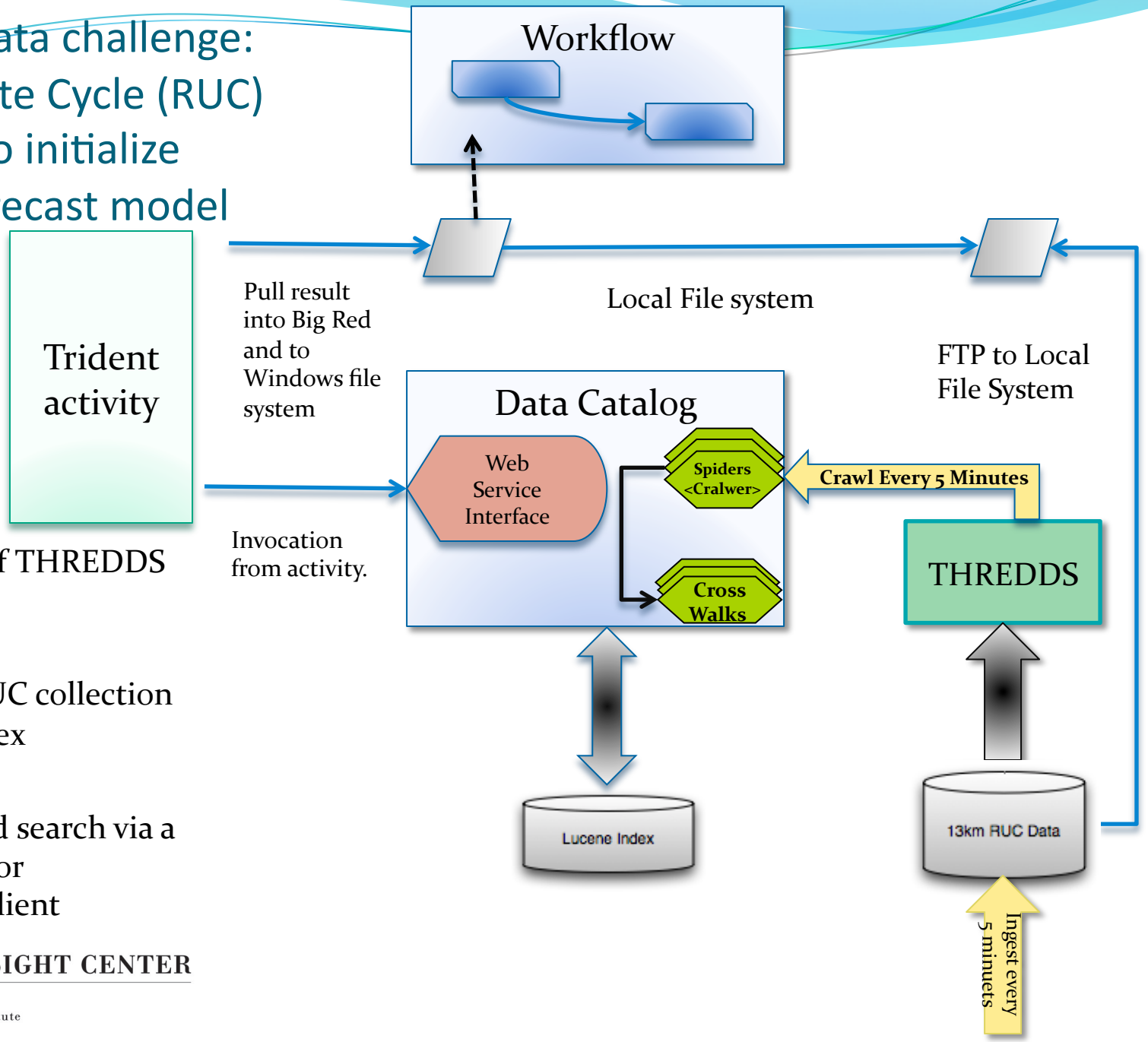


Vortex2 Experiment with Trident

Vortex2 Workflow guided by Trident



Real time data challenge: Rapid Update Cycle (RUC) data used to initialize weather forecast model

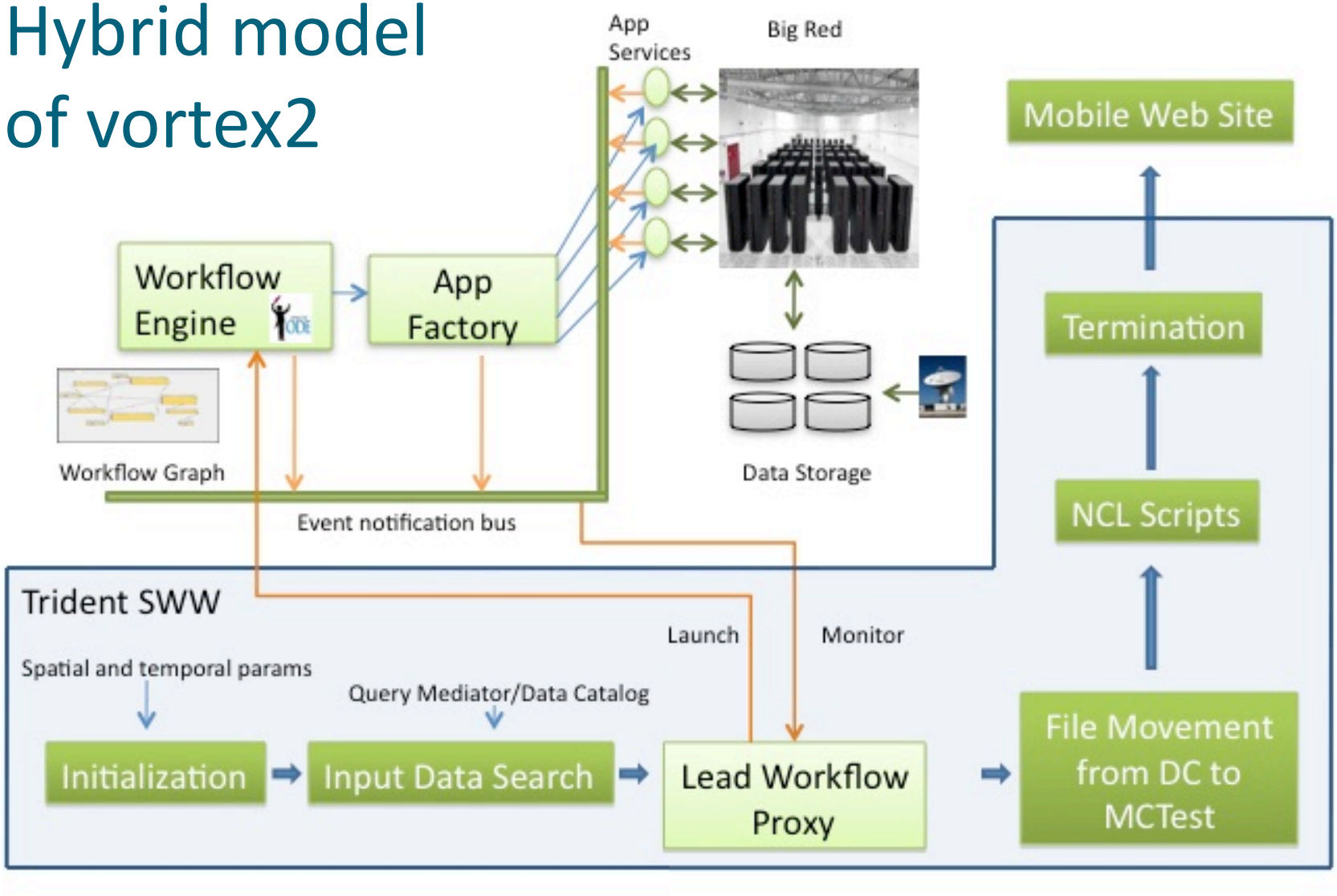


Network crawl of THREDDS (XML) catalogs

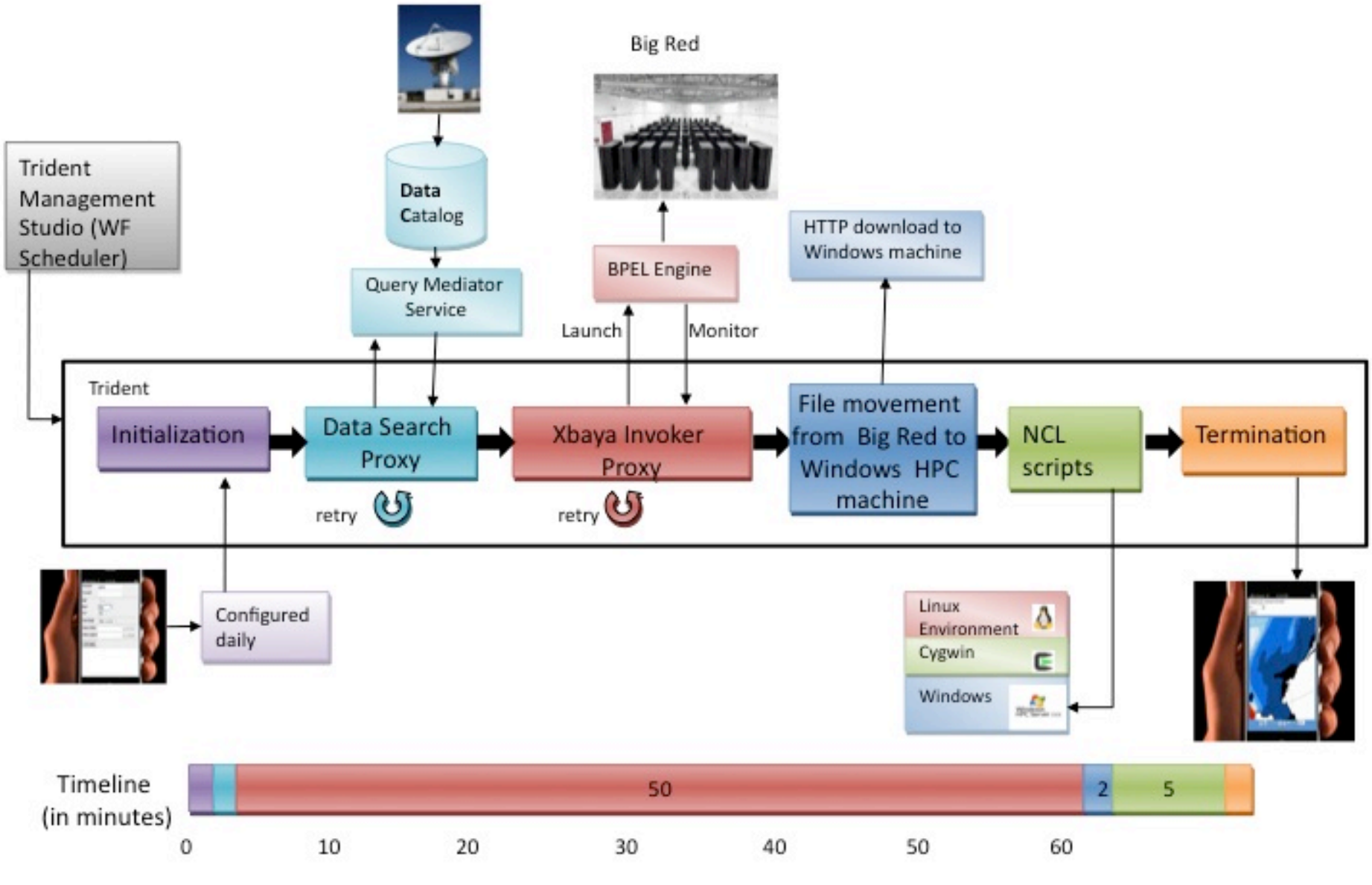
Rapid refresh RUC collection registration; index

Support data and search via a web service API or command-line client

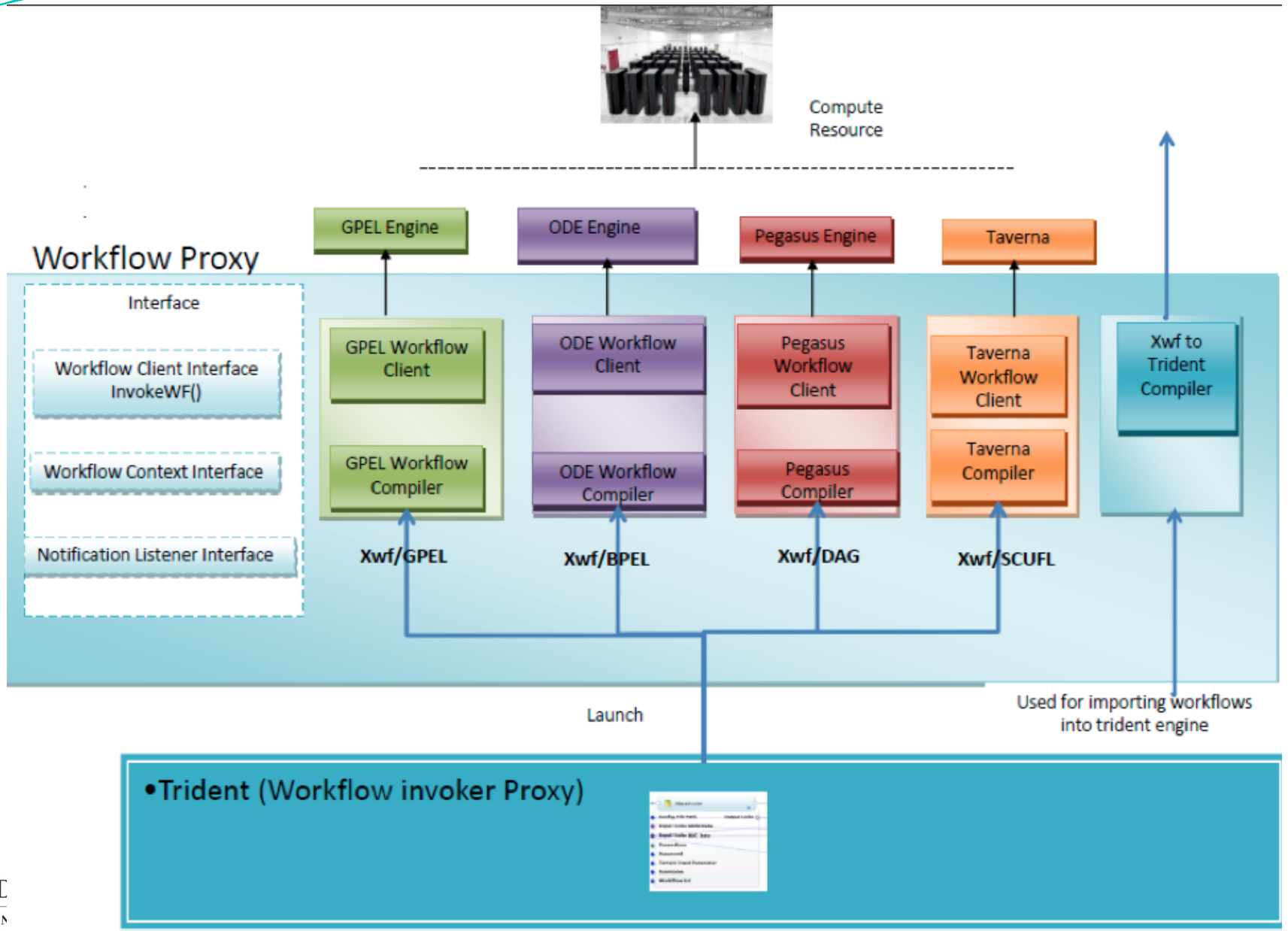
Hybrid model of vortex2



Total workflow task timeline



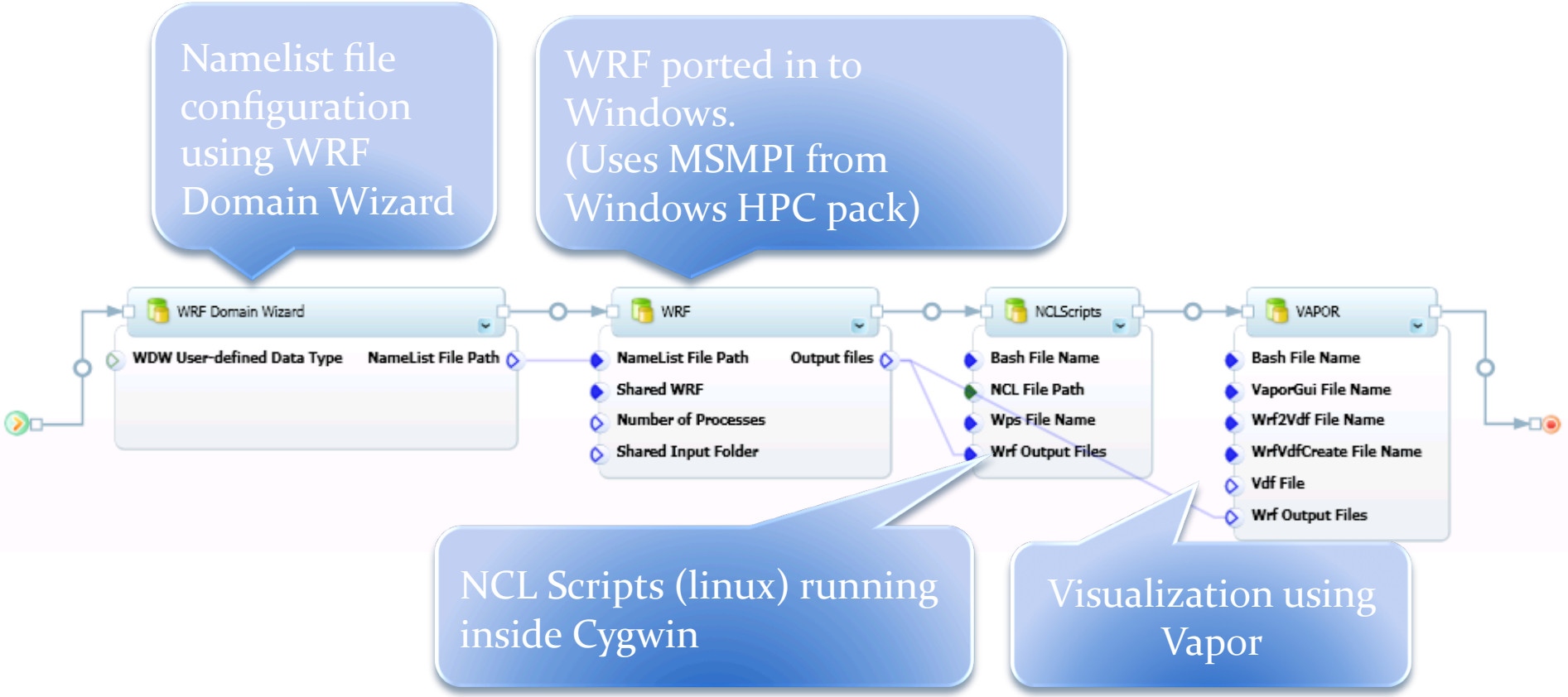
Generalized hybrid model



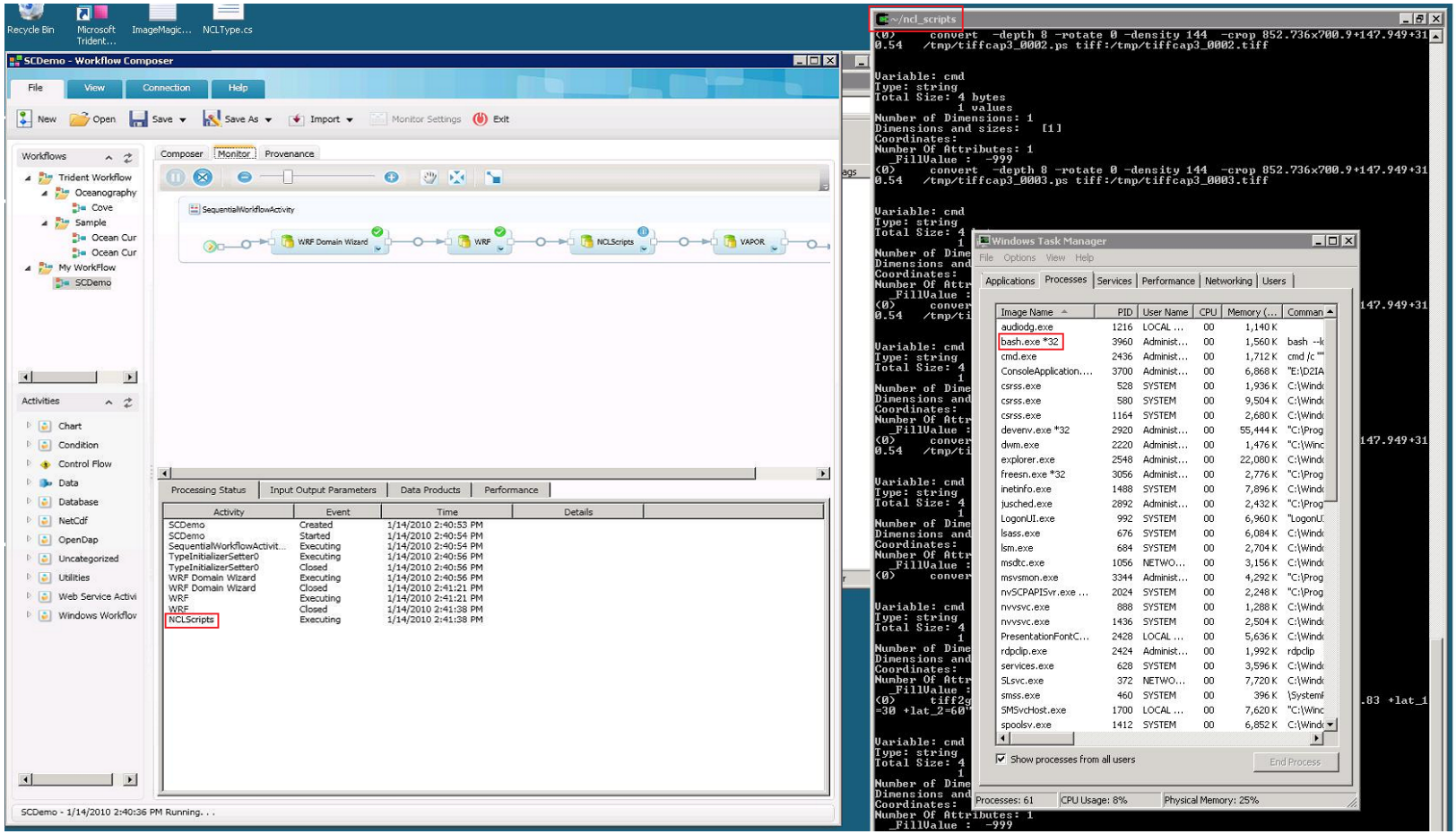
SC09 demo: Linux application on Windows

- Many scientific applications need a Linux environment to execute
- Options to run Linux applications on Windows are:
  - Port application to Windows
  - Use Linux emulator
- Cygwin (a Linux emulator) can run most Linux applications
- *LEAD-in-the-box demonstrated for first time at SC09 Trident orchestrated workflow activities running Linux applications through Cygwin*

Demonstration workflow



Linux component of workflow run on Windows box through Cygwin



NCAR NCL NCAR Command Language

Search advanced

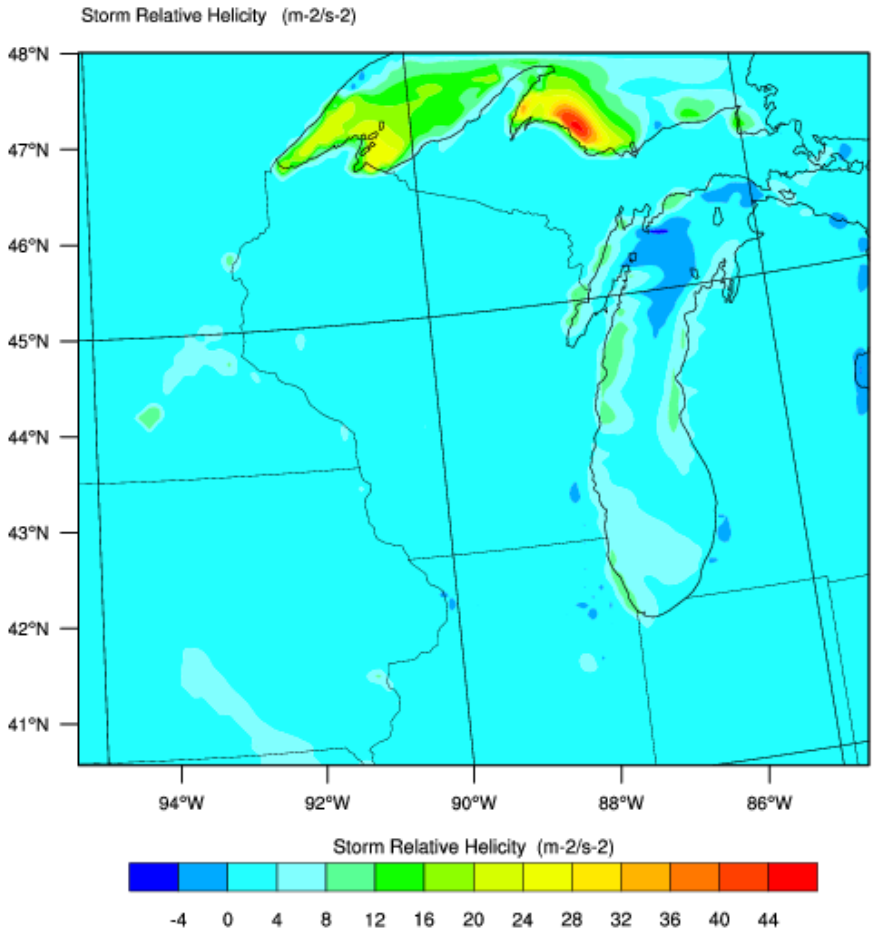
NCL Home> Application examples> Models || Data files for some examples

Example pages containing: tips | resources | functions/procedures

WRF

WRF Helicity

LEAD REAL-TIME WRF Init: 2001-06-11_12:00:00

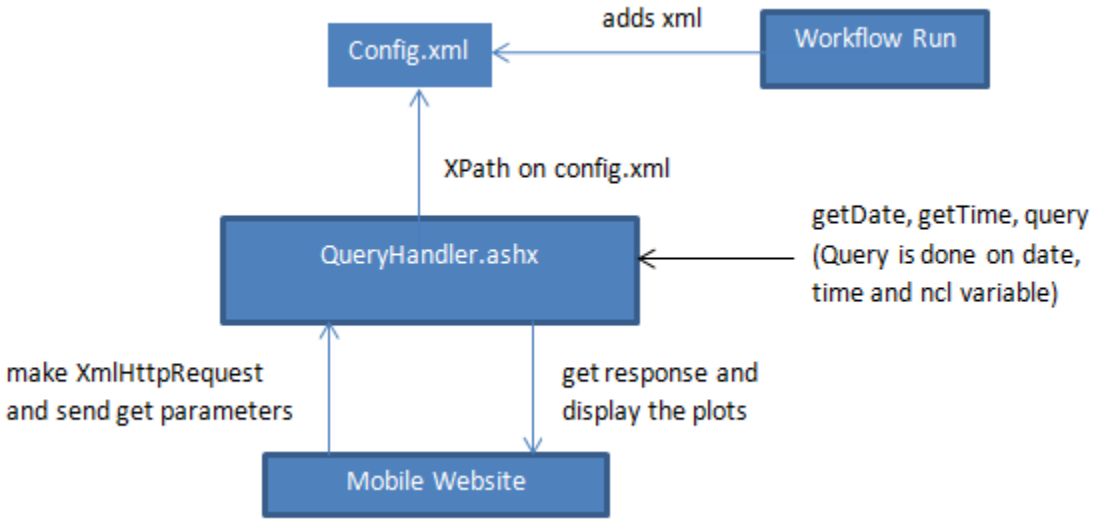


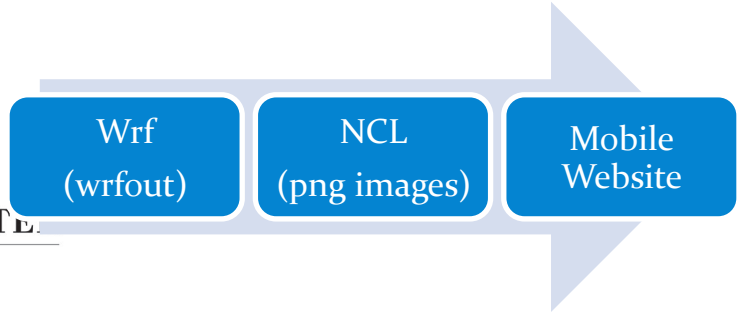
-- Helicity visualization created by Craig Mattocks.
-- Created using NCAR Command Language (NCL).
-- Helicity captures circular rotation in clouds and is particularly useful in understanding tornadoes but has other uses as well.

Mobile Website

- Mobile website
  - maintains a config.xml file
  - queries the config.xml file using QueryHandler
  - displays the plots
- Each workflow run
  - generates a log xml file
  - adds xml data to config.xml file

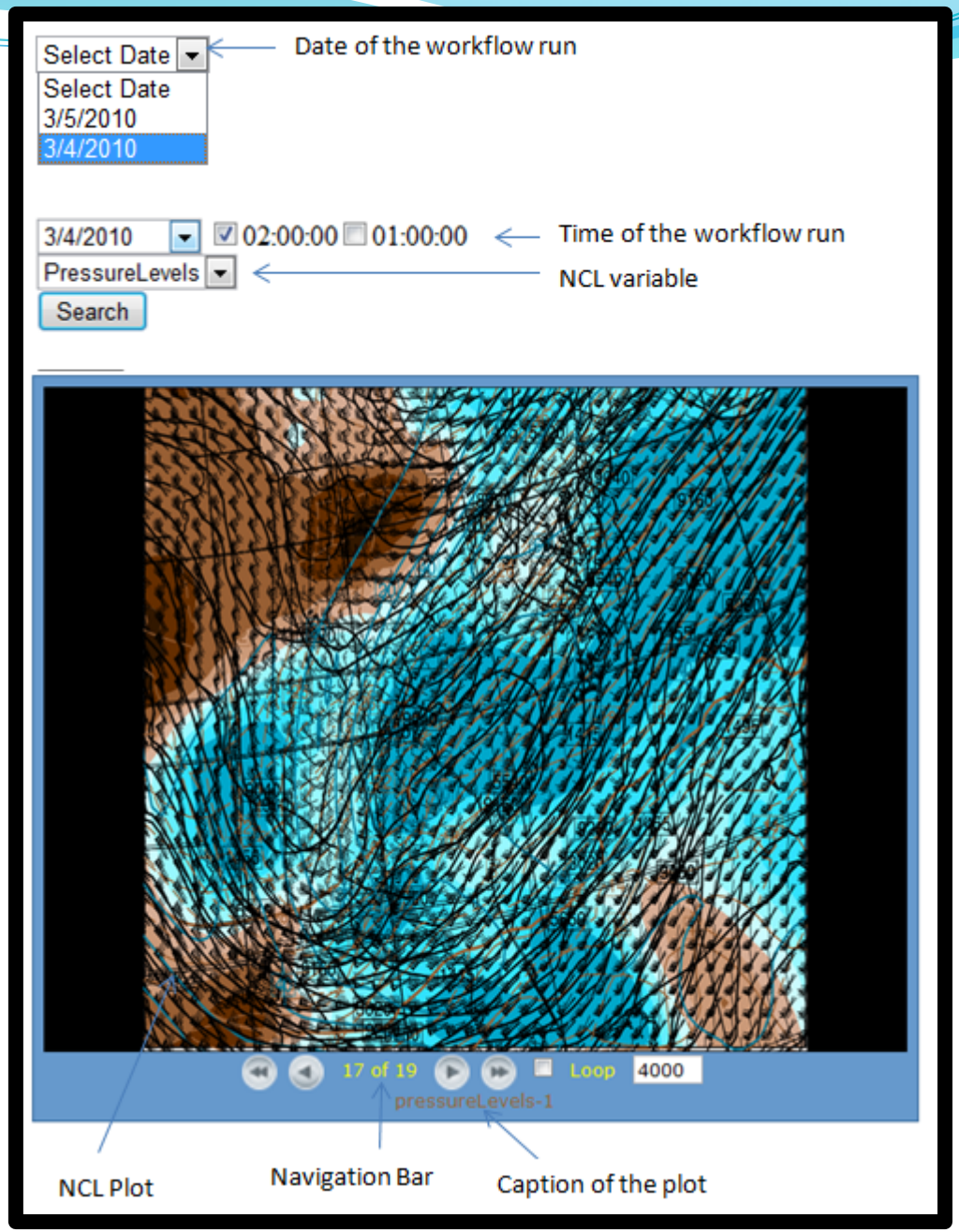
Mobile Website : query interface for smart phones





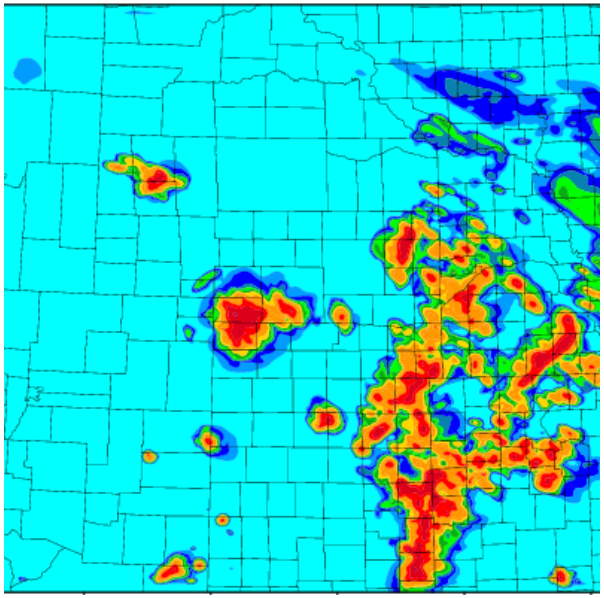
Mobile phone image retrieval

Retrieve by forecast, select product (e.g., pressure levels, radar, helicity) and
Play through forecast hours (15 hour forecast has 15 images through time).

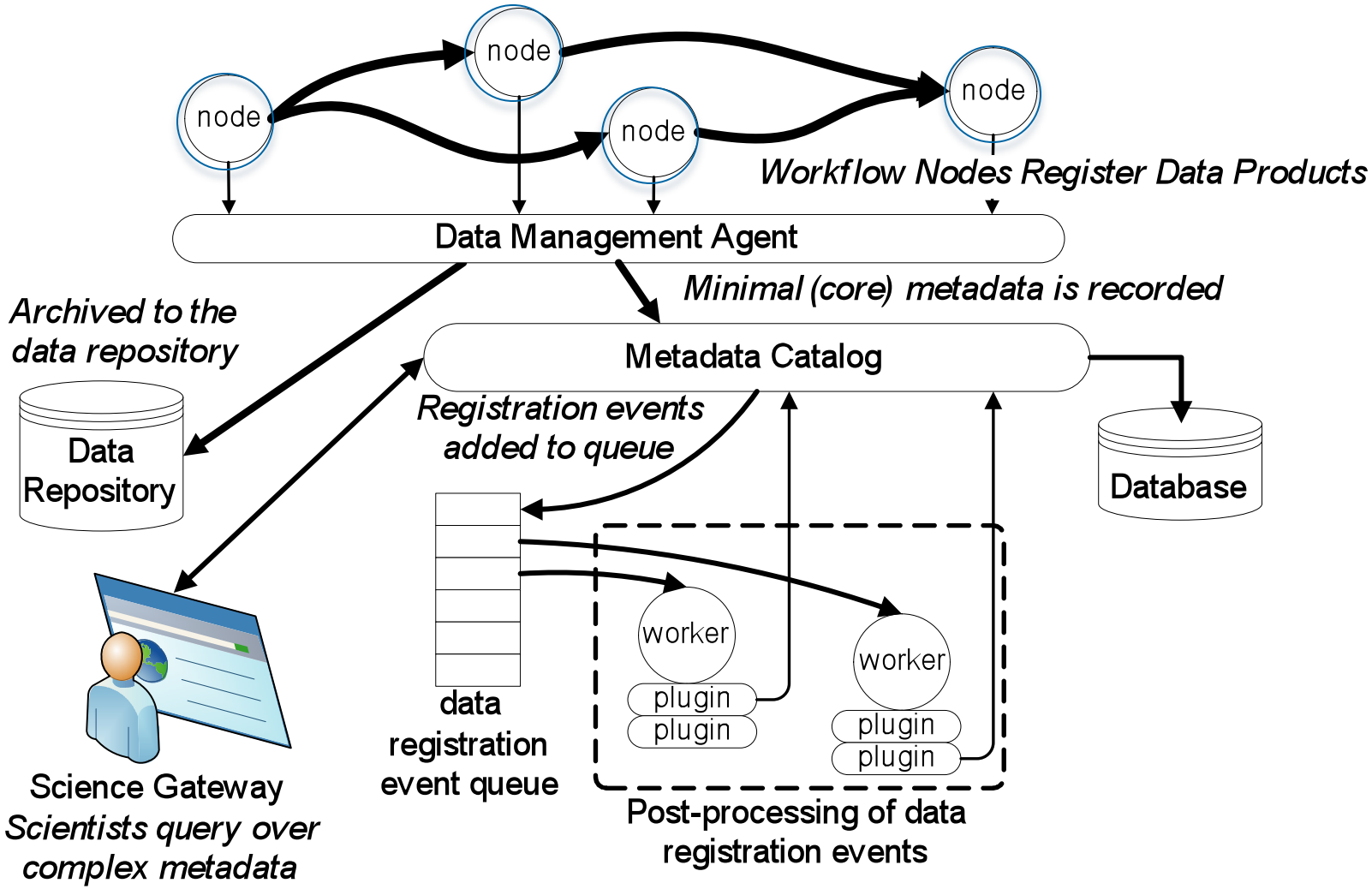


Metadata collection: essential for discovery, sharing, preservation, and reuse

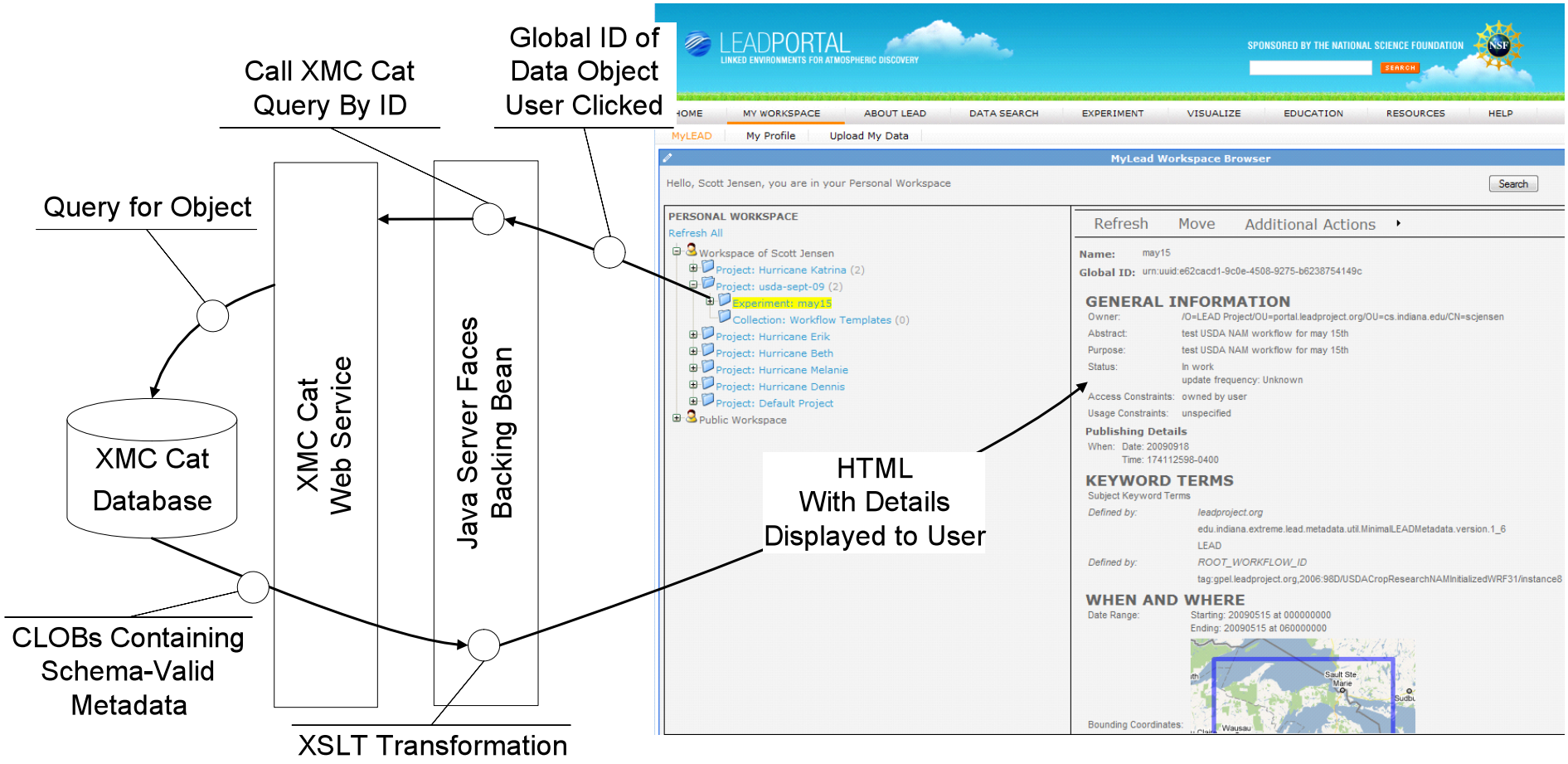
Product type: Forecast Radar Max
Experiment Timestamp: 06/07/2010 15:00:15Z
Forecast Starting Hour: 15Z
Forecast Duration: 12 hr
CenterLon: -100.5
CenterLat: 41
Loc name: Unnamed Rd, Maxwell, NE 69151, USA
SouthBound: 36.41132
EastBound: -94.75079
NorthBound: 45.58868
WestBound: -106.2492
http://www.dataandsearch.org/dsi/node/9169



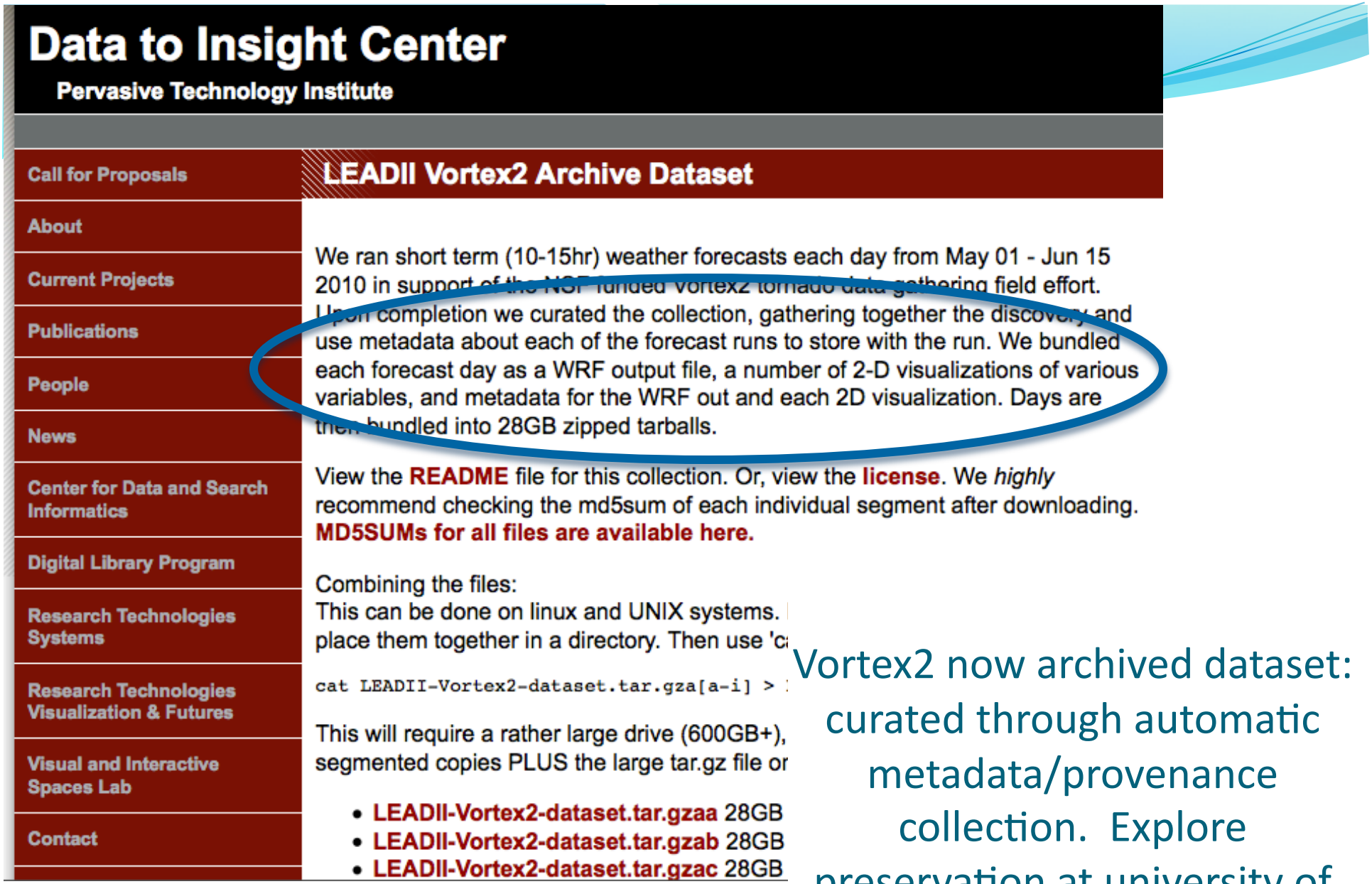
Automated Metadata Capture : XMC Cat



Metadata Queries are Interactive: invoked during browse



Queries for the full metadata must be fast – the user's mental model may not envision a query being required



Vortex2 now archived dataset: curated through automatic metadata/provenance collection. Explore preservation at university of scientific dataset

iSGTW INTERNATIONAL SCIENCE GRID THIS WEEK

Feature - The forecast before the storm

How supercomputers and hybrid workflows helped beat tornadoes to the chase

A Doppler On Wheels collects data in a tornado during VORTEX2, as PI Nolan Atkins stands nearby collecting photogrammetric data.

Image courtesy of VORTEX2.

Chasing tornadoes won't get you very far, if your goal is to understand how tornadoes form. To get results, researchers need to get their instruments on the ground before the tornado touches down.

That's the big catch 22 of VORTEX2 (Verification of the Origins of Rotation in Tornadoes Experiment), according to principal investigator Joshua Wurman. Current techniques predict tornadoes an average of only 13 minutes in advance, a fact that makes it difficult to evacuate or properly prepare for the impending disaster. To improve that lead time, or learn how to predict how destructive a tornado will be, scientists need data recorded as the tornadoes form.

"In order for us to collect good data we had to surround a supercell perhaps 40 or 50 minutes before the tornado," Wurman explained. Yet, he added, "if we knew exactly when to surround the storm, one of the big

iSGTW 8 September 2010

Feature - BiG Grid's big idea

Research - Inflated performance

Profile - People behind EGI: Tiziana Ferrari

Link of the week - News from the CCS

Image of the Week - *iSGTW* goes to Amsterdam

Announcements

Deadline extended to 12 Sept for abstracts, CGW10, Krakow, Poland

DEISA Training Courses, 14 - 16 Sept at EPCC, Edinburgh, UK

CERN Open Lab Workshop, 22-23 Sept, Geneva, Switzerland

Registration open, AGU Fall 2010 Meeting, San Francisco, USA

Jobs in distributed computing

Team members: Indiana University (lead), U North Carolina Chapel Hill, Oklahoma University