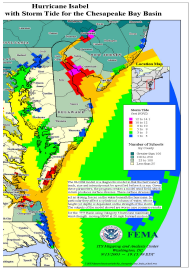


# Metadata and Preservation in Geosciences: Issues at Scale

Beth Plale

Director, Data to Insight Center of PTI

School of Informatics and Computing, Indiana University  
Bloomington



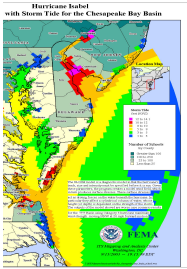
# Riches of Data Deluge

- As volume of digital scientific data increases so increases opportunities for new data-driven science
- Funding agencies beginning to mandate that research data products be made publically available
  - “The National Science Foundation is committed to the principle that the various forms of data collected with public funds belong in the public domain.”

Lab  
gains

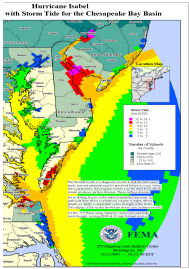
community  
gains

sharing



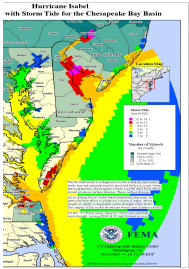
## What does it mean “in public domain”?

- It's available on my department's server, and I'm happy to share it if you ask.
- It's available on the web, but you need to know the file name to find it
- It's available on the web, but only until the machine it sits on gets too old
- *None of these good enough*



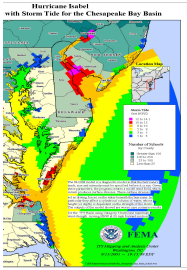
# Belonging in public domain

- A collection belonging in public domain should:
  - Be findable ..... **Discovery**
  - Be useable in another's research  
..... **Process**
  - Be as useful in 10 years as it is in 10 months  
..... **Preservation**



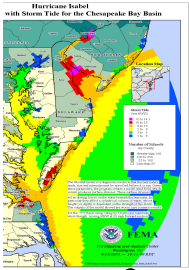
# Discovery

- Legacy solution is to share data by embedding lots of metadata into file names
  - [http://lead.unidata.ucar.edu:8080/thredds/dodsC/LEAD/radar2/KVTX/20090914/Level2\\_KVTX\\_20090914\\_1321.ar2v](http://lead.unidata.ucar.edu:8080/thredds/dodsC/LEAD/radar2/KVTX/20090914/Level2_KVTX_20090914_1321.ar2v)
  - [http://lead.unidata.ucar.edu:8080/thredds/fileServer/LEAD/model/NCEP/NAM/CONUS\\_80km/NAM\\_CONUS\\_80km\\_20090914\\_1200.grib1](http://lead.unidata.ucar.edu:8080/thredds/fileServer/LEAD/model/NCEP/NAM/CONUS_80km/NAM_CONUS_80km_20090914_1200.grib1)
- Good for those initiated into “inner-circle”
- Relying on long file names isn’t enough



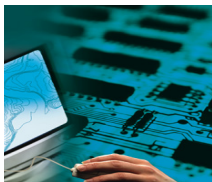
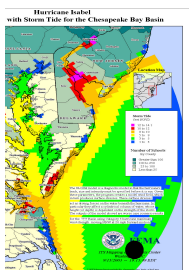
## Process: act of operating on data

- Regrid array data to match co-ordinate system of some other data set
- Sequence pattern matching
  - e.g., execution trace, DNA sequence
- Statistical analysis
  - e.g., regression analysis, confidence interval
- Inverted index for full text search



## Preservation : long term perspective

- Deals with issues of:
  - Quality of metadata about data, its accuracy, ownership. This is curation and it is expensive.
  - Media longevity – media on which digital data are stored
  - Environmental conditions – conditions in which media are stored and used. Storage on which digital files are held
  - Software and hardware requirements – tools needed to ensure longevity and usability of digital data and metadata
  - Workflow requirements – policies and practices needed to ensure longevity and usability of digital data, metadata, and tools



# Preservation

From perspective of researcher

- Researcher needs to provide more than bag of files all with long file names, which are an incomplete and inadequate form of metadata
- Researcher expects to retain ownership, receive credit
- Curation is expensive
  - Open problem to reduce cost of curation





# Discovery

Portal/web access



Database management system

**Name:** [wrfout\\_d01\\_2009-03-05\\_12:00:00](#)  
**GUID:** urn:uuid:b419247e-876d-4842-b463-e79fc50aea3b  
**Owner:** /O=LEAD Project/OU=portal.leadproject.org/  
 OU=cs.indiana.edu/CN=plale/EMAIL=plale@cs.indiana.edu  
**Create time:** ...



File system (e.g., Data Capacitor)



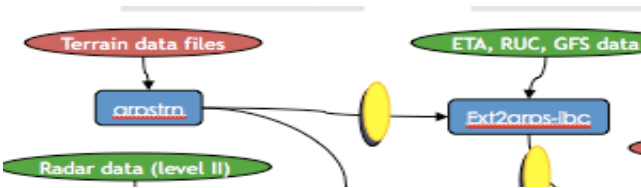
b419247e-876d-4842-b463-e79fc50aea3b

Discovery solution shown:

Metadata in database;  
data in files on disk

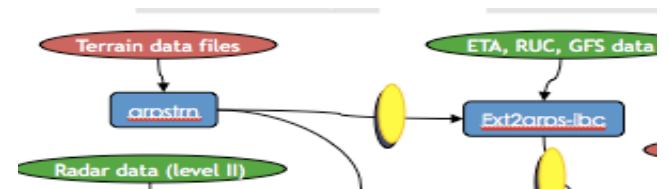
Alternate solution:

Metadata and data both in database



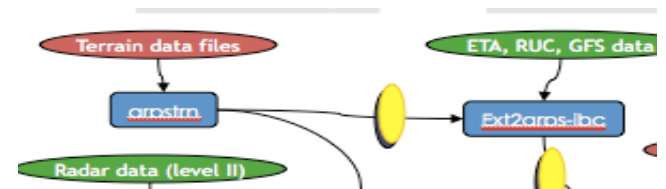
# Platforms for discovery

- Databases do Discovery very well, so using databases to store metadata makes sense.
- RDF (semantic) stores also good at storing metadata, particularly complex relationship between data objects
- Preservation platforms support discovery, but are strongest at curation and life cycle workflows



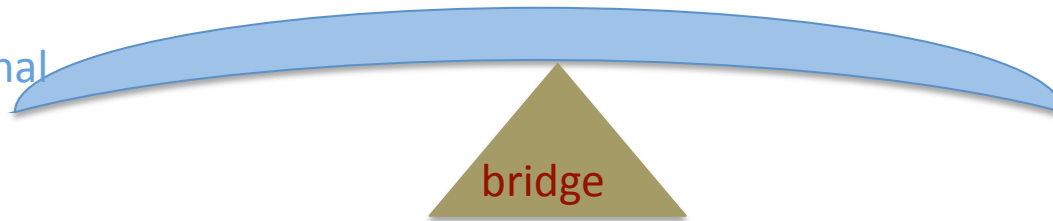
# Platforms for Process

- ... depends on Process.
  - Text applications don't use relational DBMSs
  - Regriding array data using SQL operations is nearly impossible. Earth scientists want science-specific operations as primitive operation in database.
  - Bio community uses RDF stores to store data. But column store (data warehouse servers) better at some RDF workloads.

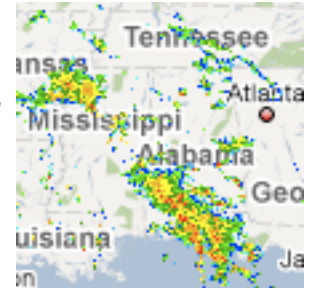


# Additional issue scientists face when sharing: community schemas.

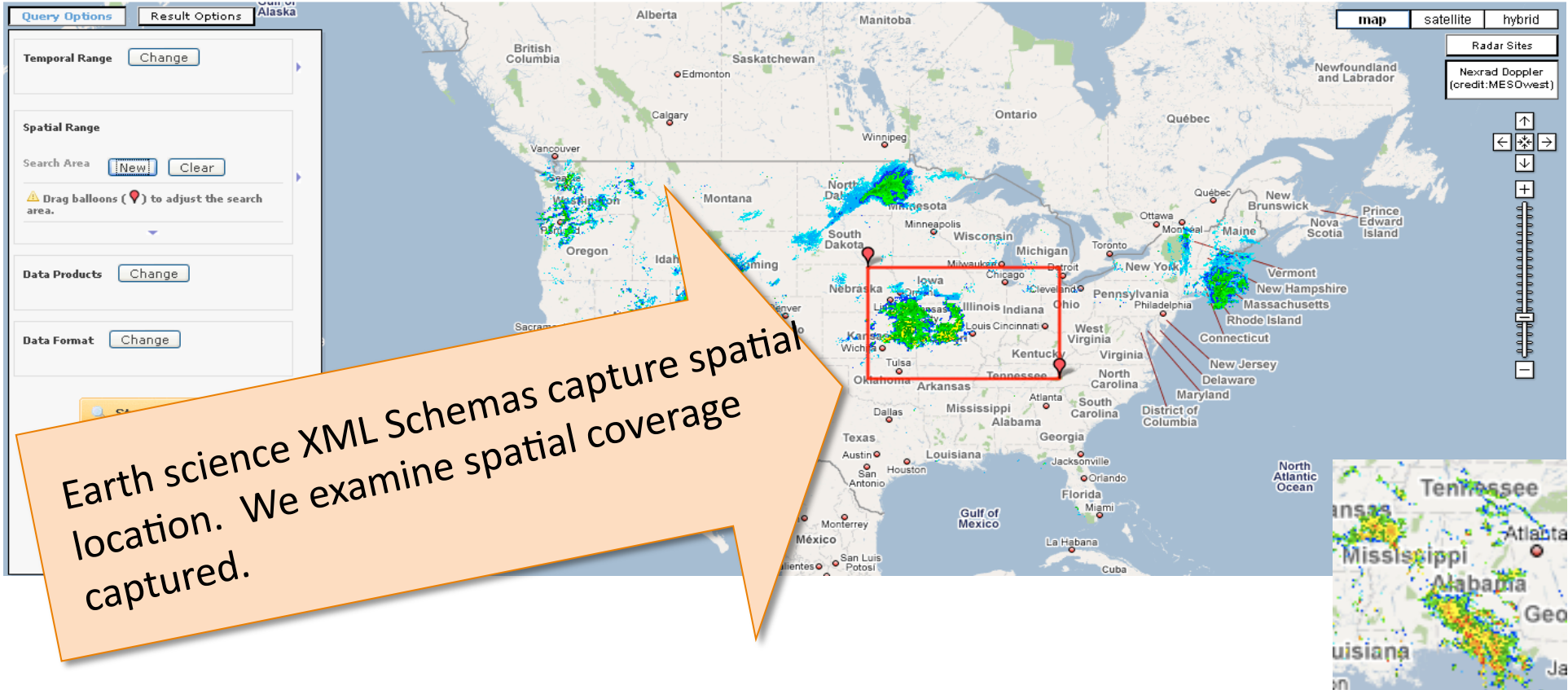
Relational model



Community schema



XMC Cat is a metadata catalog with wizard that builds a bridge between community XML Schema and general relational metadata data model

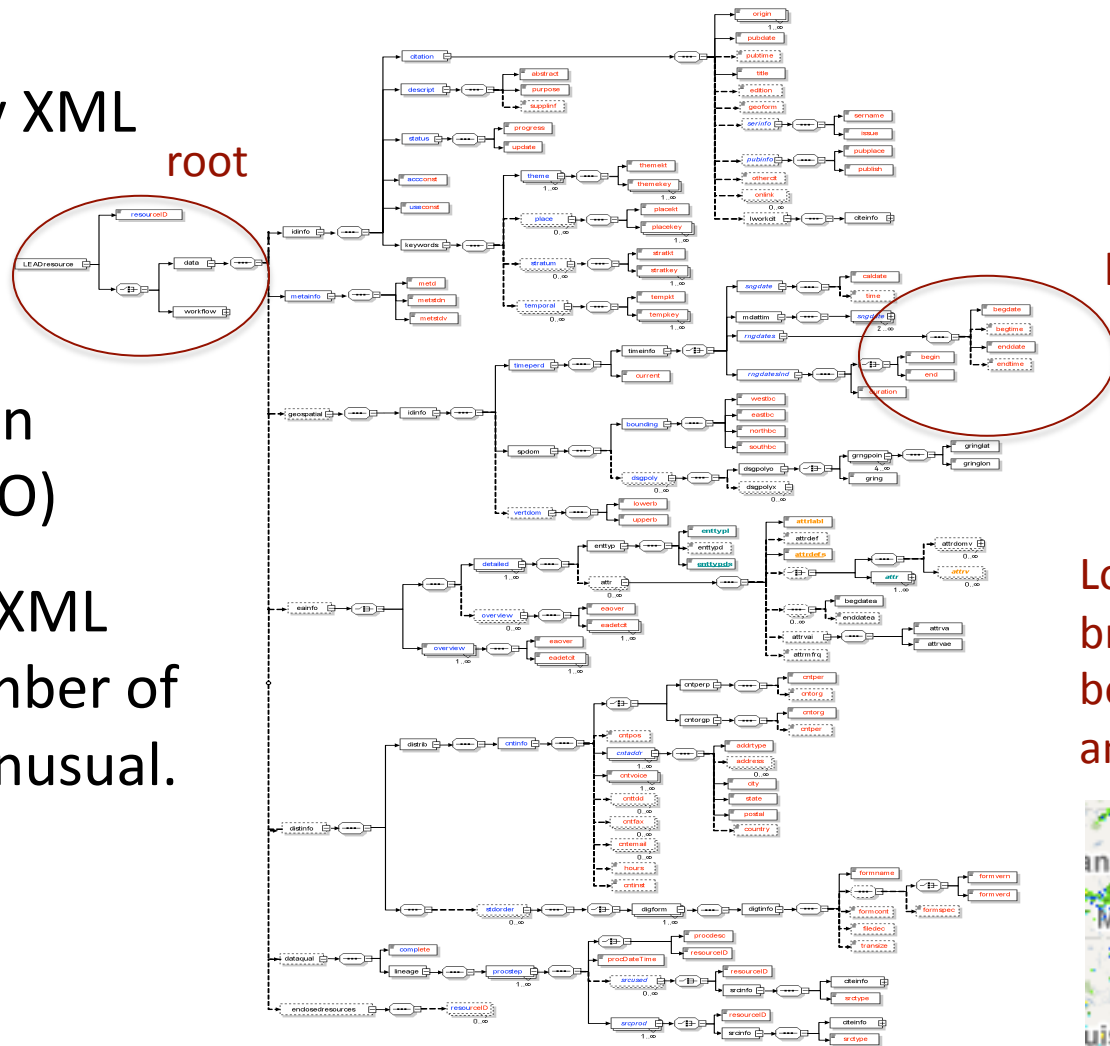


Earth science XML Schemas capture spatial location. We examine spatial coverage captured.

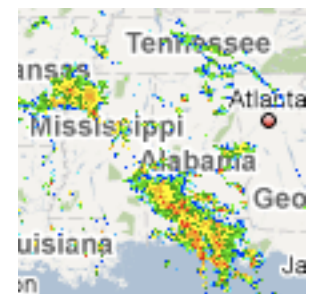
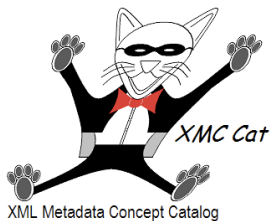
# Typical community XML Schema

- Profile of common standard (FGDC/ISO)

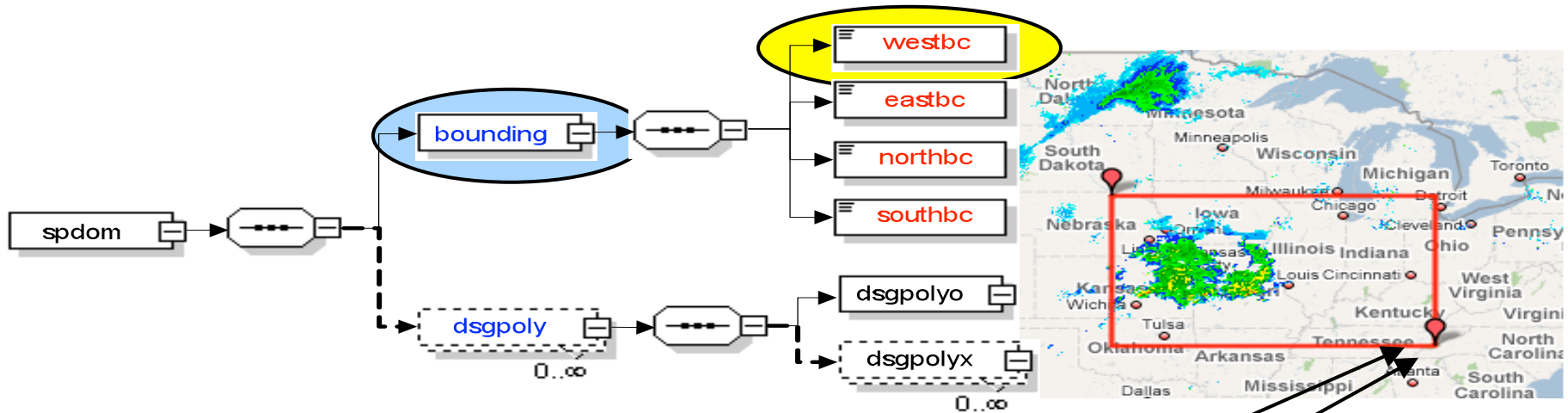
- Composed of 11 XML Schema files – number of schema files not unusual.



Lots of branches between root and leaves







```
<xs:element name="bounding" type="boundingType"/>
```

```
<xs:complexType name="boundingType">
  <xs:sequence>
    <xs:element ref="westbc"/>
    <xs:element ref="eastbc"/>
    <xs:element ref="northbc"/>
    <xs:element ref="southbc"/>
  </xs:sequence>
</xs:complexType>
```

*Need to get this XML into a relational database and query it using geospatial support provided by DBMS*



To publish your collection in community “compliant” XML, you

1. Download the XMC Cat catalog, and
2. Walk through a wizard that lets XMC Cat set itself up with your community schema.

XMC Cat Builder

Select Schema > Imports and Includes > Starting Element > Metadata Concepts > Download > Next ->

**Step 1: Select an Existing Schema:**

Select a schema you previously loaded or load a new schema using the schema import option below.

None Selected  
LEAD.xsd  
Xmccat-JSDL.xsd

**Selected Schema: None Selected**

Please select a schema from the list of schemas available to work with. If you wish to work with a new schema that you have not previously loaded, please use the schema import option.

**Or, Use a New Schema:**

/usr/XConCat/LMS/LEAD.xsd Browse...

Schema Name:  
Lead-Version2

Schema Description:  
This is version 2 of the LEAD Metadata Schema (LMS)

Load Schema

From top-level schema, Wizard can figure out the dependencies (knows that all 11 schemas are needed).

Wizard follows schema dependencies to make sure administrator gets them all identified.

The screenshot shows the XMC Cat Builder web application interface. The browser title is "XMC Cat Builder - Mozilla Firefox" and the address bar shows "http://localhost:8080/builder/SelectSchema.faces". The main content area displays "XMC Cat Builder" with a cat logo. Below the logo is a breadcrumb trail: "Select Schema > Imports and Includes > Starting Element > Metadata Concepts > Download > Next ->".

The main section is titled "Step 2: Schema Imports and Includes:" and contains the text: "Following is a list of the additional schema imports and includes specified in your schema." Below this is a list of schema dependencies:

- Import Completed:** Show Details ...  
Namespace: http://schemas.leadproject.org/2007/01/lms/fgdc  
Location: ./LeadFGDC.xsd
- Import Required:** Select File ...  
Namespace: http://schemas.leadproject.org/2007/01/lms/lea  
Location: ./LEADElements.xsd
- Include Required:** Select File ...  
Namespace: http://schemas.leadproject.org/2007/01/lms/fgc  
Location: ./Utility.xsd
- Include Required:** (highlighted in yellow)  
Namespace: http://schemas.leadproject.org/2007/01/lms/fgc  
Location: ./CitationInformation.xsd
- Include Required:** Select File ...
- Include Required:** Select File ...
- Include Required:** Select File ...

A modal dialog titled "Load Imported or Included Schema File:" is open on the right. It shows "Namespace: http://schemas.leadproject.org/2007/01/lms/fgdc" and "Schema Location: ./CitationInformation.xsd". Below this is a text input field containing "/usr/XConCat/LMS/CitationInformation.xsd" and a "Browse..." button. At the bottom of the dialog are "Load File" and "Cancel" buttons.

In the bottom right corner, there is a map of the Southeastern United States showing states like Tennessee, Mississippi, Alabama, Georgia, and Louisiana. The map displays a color-coded overlay, possibly representing a data distribution or a specific geographic area of interest.

Wizard then asks user to point out major concepts in schema and starting point in schema tree – the Root.

There's our bounding box again, now it is a major concept so will automatically have a search built.

The screenshot shows the XMC Cat Builder web application in Mozilla Firefox. The browser title is "XMC Cat Builder - Mozilla Firefox" and the address bar shows "http://localhost:8080/builder/SelectSchema.faces". The application header includes the title "XMC Cat Builder" and a cat logo. A breadcrumb trail reads: "Select Schema > Imports and Includes > Starting Element > Metadata Concepts > Download > Next ->".

The main content area is titled "Step 4: Identify Metadata Concepts:". It displays the following information:

- Schema Name: LEAD.xsd
- Starting Element: LEADresource

A schema tree is shown on the left. The root is "LEADresource", which contains several child elements. The "data" element is expanded, showing "idinfo", "keywords", "metainfo", "geospatial", and "vertdom". The "idinfo" element is further expanded, showing "Concept: citation", "Concept: describe", "Concept: status", "Element: progress", "Element: update", "Concept/Element: accconst", and "Concept/Element: useconst". The "geospatial" element is expanded, showing "idinfo", "timeperd", "spdom", and "vertdom". The "idinfo" element under "geospatial" is expanded, showing "Concept: bounding" (highlighted with a red oval), "Element: westbc", "Element: eastbc", "Element: northbc", "Element: southbc", "Concept: dsqpoly", and "vertdom".

The "Complex Type Element" configuration panel on the right is filled with the following values:

- Element name in the schema: bounding
- Defined in namespace: http://schemas.leadproject.org/200
- Contained in schema: IdentificationInformation.xsd
- Annotate as concept:
- Do not parse child elements:
- Category for this concept: Existing category: Spatial
- Name in the metadata catalog: SpatialBounds
- Source in the catalog: Default
- Existing source: http://schemas.leadproject.org/2
- Short description: Spatial Bounds
- Full description: Bounding box for spatial domain of this object.
- Schema data type: complex type
- Catalog data type: No Data Type Selected

MyLead Workspace Browser

< Return To Browser New Search Save Search Modify Search

- QUERY RESULTS**
- june 8
  - vortex test 4**
  - Vortex 3 test
  - nam wrf test 2
  - Vortex 2 test
  - mesonet OK visit
  - P436 demo of portal
  - data-driven experiment

Experiment: vortex test 4 (6)

**EXPERIMENT: VORTEX TEST 4**

Global ID: urn:uuid:dc8e66b9-ad83-4aa9-9610-a2d2bf3bb0db

**GENERAL INFORMATION**

Owner: /O=LEAD Project/OU=portal.leadproject.org/OU=cs.indiana.edu/CN=plale/EMAIL=plale@  
Abstract: test 4 - certs working now  
Purpose: test 4 - certs working now  
Status: In work  
update frequency: Unknown  
Access Constraints: owned by user  
Usage Constraints: unspecified

**Publishing Details**

When: 20090504  
215200319-0400

**KEYWORD TERMS**

Subject Keyword Terms  
Defined by: leadproject.org  
edu.indiana.extreme.lead.metadata.util.MinimalLEADMetadata.version.1\_6  
LEAD  
Defined by: ROOT\_WORKFLOW\_ID  
tag:gpel.leadproject.org,2006:86H/PublicNAMInitializedWRFForecast1/instance344

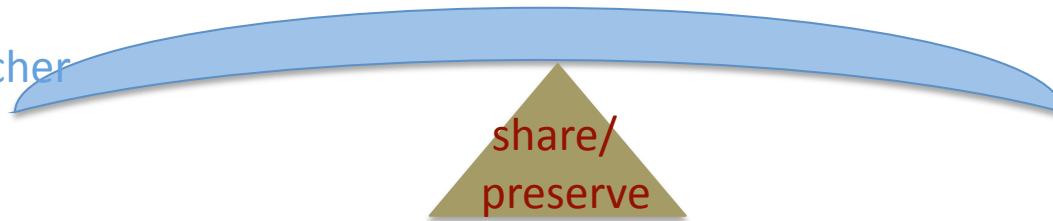
**WHEN AND WHERE**

Date Range: Starting: 20090505 at 015200321  
Ending: 20090505 at 075200321  
North  
latitude: 49.1104

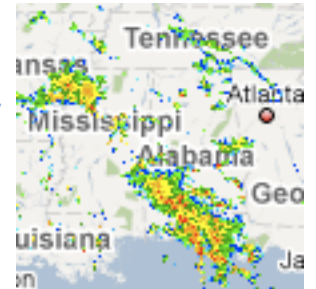
Schema configuration provisions search features: here search for failed experiments

# Additional issue scientists face when sharing over long term (preservation): metadata and annotations.

Researcher  
lab



Community

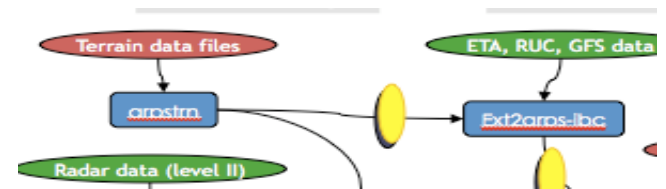


Could chuck bag of files over fence but then curation cost are very high. Open question on how cost can be reduced.

# Provenance of art works



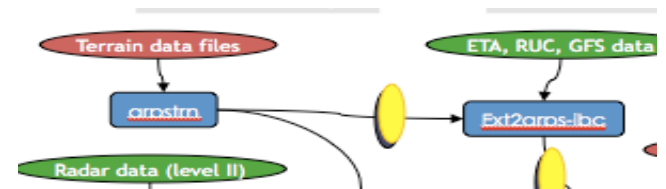
- Trace of history of work of art from moment it was made until it comes into a collection.
- Impartial and authoritative information on authenticity, ownership, theft, and other artistic, legal, and ethical issues concerning art objects.





**It can be impossible to determine complete provenance for work of art:**

- provenance records often reflect an owner or former owner's wish for anonymity.
- ephemeral nature of historical records, which are often lost or destroyed over time

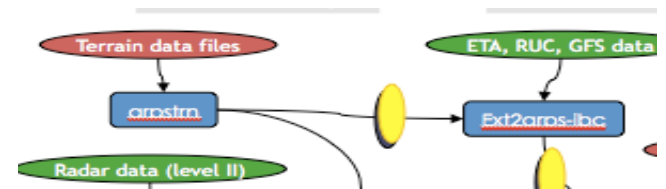


Saint John the Baptist in the Wilderness;  
circa 1535 Moretto da Brescia Italy,  
Brescia 1498–1554



Provenance:

Oskar Bondy [d. 1944], Vienna<sup>1</sup>  
[Jacob M. Heimann, Beverly  
Hills];<sup>2</sup> Philip Yordan,  
Beverly Hills (until 1951),<sup>3</sup>  
gift 1951; to Los Angeles  
County Museum of Art





**LEADPORTAL**  
LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

SPONSORED BY THE NATIONAL SCIENCE FOUNDATION

HOME MY WORKSPACE ABOUT LEAD DATA SEARCH EXPERIMENT VISUALIZE EDUCATION RESOURCES

Introduction Geographic Region Noesis

Query Options Result Options History

Temporal Range Change  
past 30 minutes

Spatial Range  
Search Area New Clear  
Drag balloons (📍) to adjust the search area.

Data Products Change  
Any

Data Format Change  
Any

Start Search

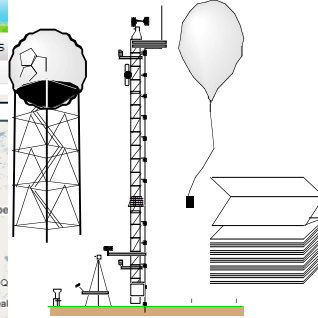
POWERED BY Google

500 km 500 mi lat, lng: 38.6855, -67.0605

COPYRIGHT © 2006 LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY. ALL RIGHTS RESERVED.

©2001 AccuWeather, Inc.

accuweather.com



## What is Provenance for Digital Scientific Data?



**LEADPORTAL**  
LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

SPONSORED BY THE NATIONAL SCIENCE FOUNDATION

HOME MY WORKSPACE ABOUT LEAD DATA SEARCH EXPERIMENT VISUALIZE EDUCATION RESOURCES HELP

MyLEAD My Profile

MyLead Workspace Browser

Hello, Yiming Sun, you are in your Personal Workspace

Collection: Workflow Instances  
 Collection: Input Data Files  
 Experiment: adas-1-jan-17-1100 (2)  
 Collection: Workflow Templates (0)  
 Project: Data Mining (14)  
 Experiment: dm-3-jan-18-1400 (11)  
 File: notiflog-cfd882d6-f917-4fe4-99aa-e8a81ed3890b.xml  
 File: KMOB\_20050829\_0028\_Clustered\_Storm\_Locations\_For\_Visualization.k  
 File: KMOB\_20050829\_0028\_Clustered\_Storm\_Locations.arff  
 File: ITSC\_SCA-V1.0-Pub.stdout  
 File: KMOB\_20050829\_0028\_Extracted\_Attributes.txt  
 File: ITSC\_Remove-V1.0-Pub.stdout  
 File: KMOB\_20050829\_0028\_Storm\_Locations\_For\_Visualization.kml  
 File: KMOB\_20050829\_0028\_Detected\_Storm\_Locations.txt  
 File: ITSC\_SDA-V1.0-Pub.stdout  
 Collection: Workflow Instances (1)  
 Collection: Instance of Data Mining: Detect & Cluster Storms (dm-3-jan-18-1400) (0)  
 Collection: Input Data Files  
 Experiment: dm-2-jan-18-1400  
 Experiment: dm-1-jan-18-1400  
 Experiment: dm-3-jan-18-1115  
 Experiment: dm-2-jan-18-1115  
 Experiment: dm-1-jan-18-1115  
 Experiment: dm-1-jan-17-1515  
 Experiment: dm-7-jan-17-1100  
 Experiment: dm-6-jan-17-1100  
 Experiment: dm-5-jan-17-1100  
 Experiment: dm-4-jan-17-1100  
 Experiment: dm-3-jan-17-1100

**INFORMATION FOR CURRENTLY SELECTED ITEM**  
 Refresh · Delete · View · Download · Publish · Annotate Experiment · Toggle Provenance · Move

**Name:** dm-3-jan-18-1400  
**Type:** experiment  
**GUID:** urn:uuid:cfd882d6-f917-4fe4-99aa-e8a81ed3890b  
**Description:** dm  
**Owner:** /O=LEAD Project/OU=portal.leadproject.org/OU=cs.indiana.edu/CN=yiming  
**Create time:** Fri Jan 18 00:00:00 EST 2008 14037420-0500  
**Purpose:** dm  
**Progress:** In work  
**Update:** Unknown  
**Access Constraint:** owned by user  
**Use Constraint:** unspecified  
**Theme:** leadproject.org  
 key = LEAD  
 key = edu.indiana.extreme.lead.metadata.util.MinimalLEADMetadata.version.1\_6  
**Theme:** ROOT\_WORKFLOW\_ID  
 key = tag:gpel.leadproject.org,2006:7CH/DataMiningDetectClusterStorms/instance290  
**East Coordinate:** 0.0  
**West Coordinate:** 0.0  
**North Coordinate:** 0.0  
**South Coordinate:** 0.0  
**topicID:** topicID = cfd882d6-f917-4fe4-99aa-e8a81ed3890b  
**WorkflowInitialized:** serviceID = tag:gpel.leadproject.org,2006:7CH/DataMiningDetectClusterStorms/instance290  
 timeStamp = 2008-01-18T14:03:44  
**WorkflowTerminated:** serviceID = tag:gpel.leadproject.org,2006:7CH/DataMiningDetectClusterStorms/instance290  
 timeStamp = 2008-01-18T14:33:21

Science Gateway:  
 User workspace captures metadata and activity through time

LEADPORTAL  
LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

SPONSORED BY THE NATIONAL SCIENCE FOUNDATION

HOME MY WORKSPACE ABOUT LEAD DATA SEARCH **EXPERIMENT** VISUALIZE EDUCATION RESOURCES HELP

Introduction Experiment Builder Application Tools

**Experiment Wizard**

User: yiming Project: Mesoscale Weather Forecast  
 Name: WRF  
 Description: Using WRF model to perform forecast over North America regions  
 Workflow: Public: NAM Initialized WRF Forecast

**Model Domain Configuration**

**Region Type Selection**

- Regional 1000Km X 1000Km X 51 Domain with 20 Km Grid Spacing
- Regional 1000Km X 1000Km X 51 Domain with 5 Km Grid Spacing
- CONUS (5520Km X 3520Km X 51) Domain with 20 Km Grid Spacing

**Forecast Start Time**  
 Dates and times in Greenwich Mean Time (GMT)

- Now (in other words, run a forecast using the most recent data available)
- Please specify:

Start Date: 2008/11/06 Current Time: 2008/11/06 03:29Z  
 Start Hour: 0Z

Forecast Duration: 6 hours

Using your mouse, drag and drop the center of the model domain grid to position it as desired on the map

**Forecast Domain**  
 center latitude: 35.1000  
 center longitude: -97.4000

Drag the balloon (📍) to move the region.  
 Settings

map satellite hybrid  
 Radar Sites  
 Nexrad Doppler (credit: MESOwest)

Science Gateway:  
 Wizard guides scientist through workflow configuration

LEADPORTAL  
LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

SPONSORED BY THE NATIONAL SCIENCE FOUNDATION

HOME MY WORKSPACE ABOUT LEAD DATA SEARCH EXPERIMENT VISUALIZE EDUCATION RESOURCES HELP

Introduction Experiment Builder Application Tools

Experiment Builder Portlet

**Experiment Wizard**

User: yiming Project: Mesoscale Weather Forecast  
 Name: WRF  
 Description: Using WRF model to perform forecast over North America regions  
 Workflow: Public: NAM Initialized WRF Forecast

**View/Modify WRF Namelist Parameters**

Physics Options		
Parameter	Value	Description
mp_physics	2	<b>Microphysics Option</b> = 0, no microphysics = 1, Kessler scheme = 2, Lin et al. scheme = 3, WSM 3-class simple ice scheme = 4, WSM 5-class scheme = 5, Ferrier (new Eta) microphysics = 6, WSM 6-class graupel scheme = 8, Thompson 7-class scheme = 98, NCEP 3-class simple ice scheme = 99, NCEP 5-class scheme
ra_lw_physics	1	<b>Longwave Radiation Option</b> = 0, no longwave radiation = 1, rrtm scheme = 3, cam scheme = 99, GFDL (Eta) longwave
ra_sw_physics	1	<b>Shortwave Radiation Option</b> = 0, no shortwave radiation = 1, Dudhia scheme = 2, Goddard short wave = 3, cam scheme = 99, GFDL (Eta) longwave
radt	10.00	<b>Minutes Between Radiation Physics Calls</b>
sf_sfclay_physics	1	<b>Surface-layer Option</b> = 0, no surface-layer = 1, Monin-Obukhov scheme = 2, Monin-Obukhov (Janjic Eta) scheme = 3, NCEP Global Forecast System scheme

Science Gateway: Configuration includes physics parameters, making workflows better suited to high level experts.

# Typical e-Science Experiment

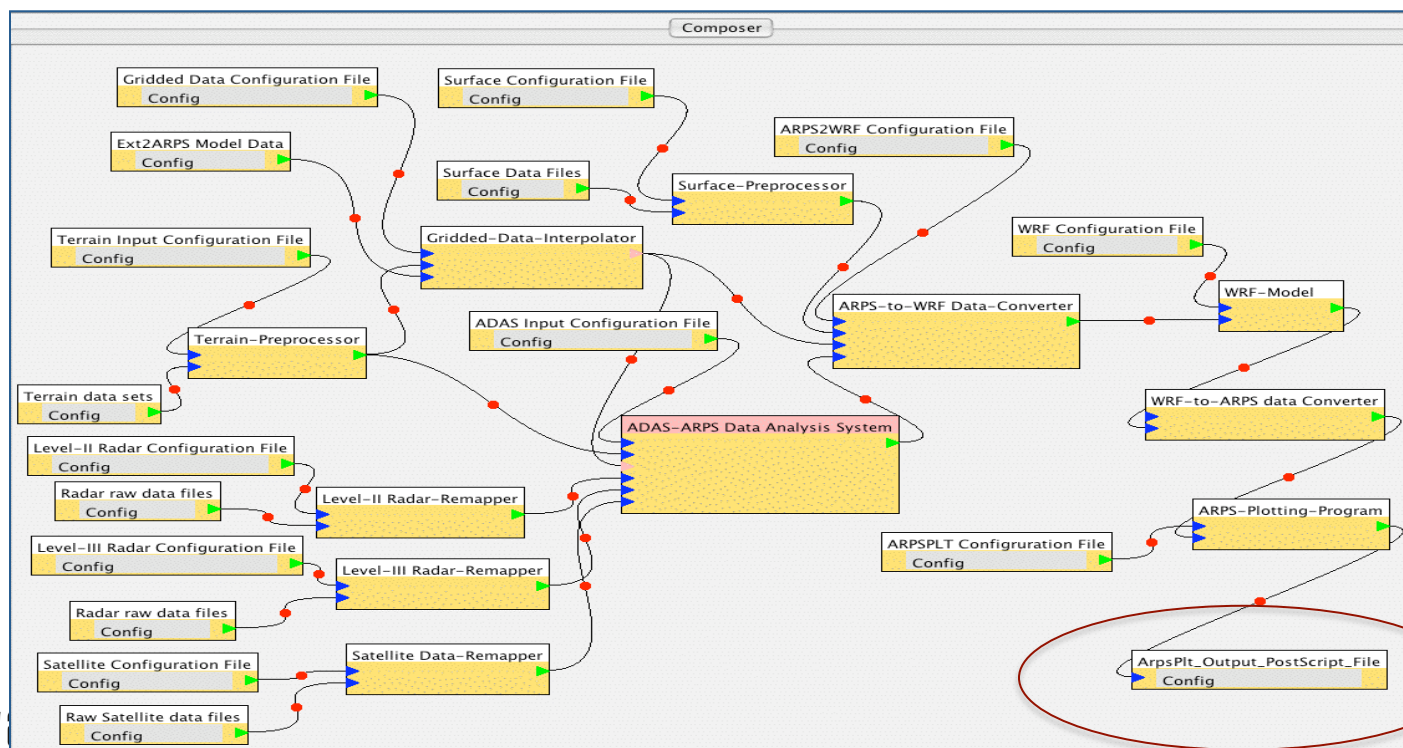
*Weather forecast using WRF in LEAD : role of provenance?*

Pre-Processing

Assimilation

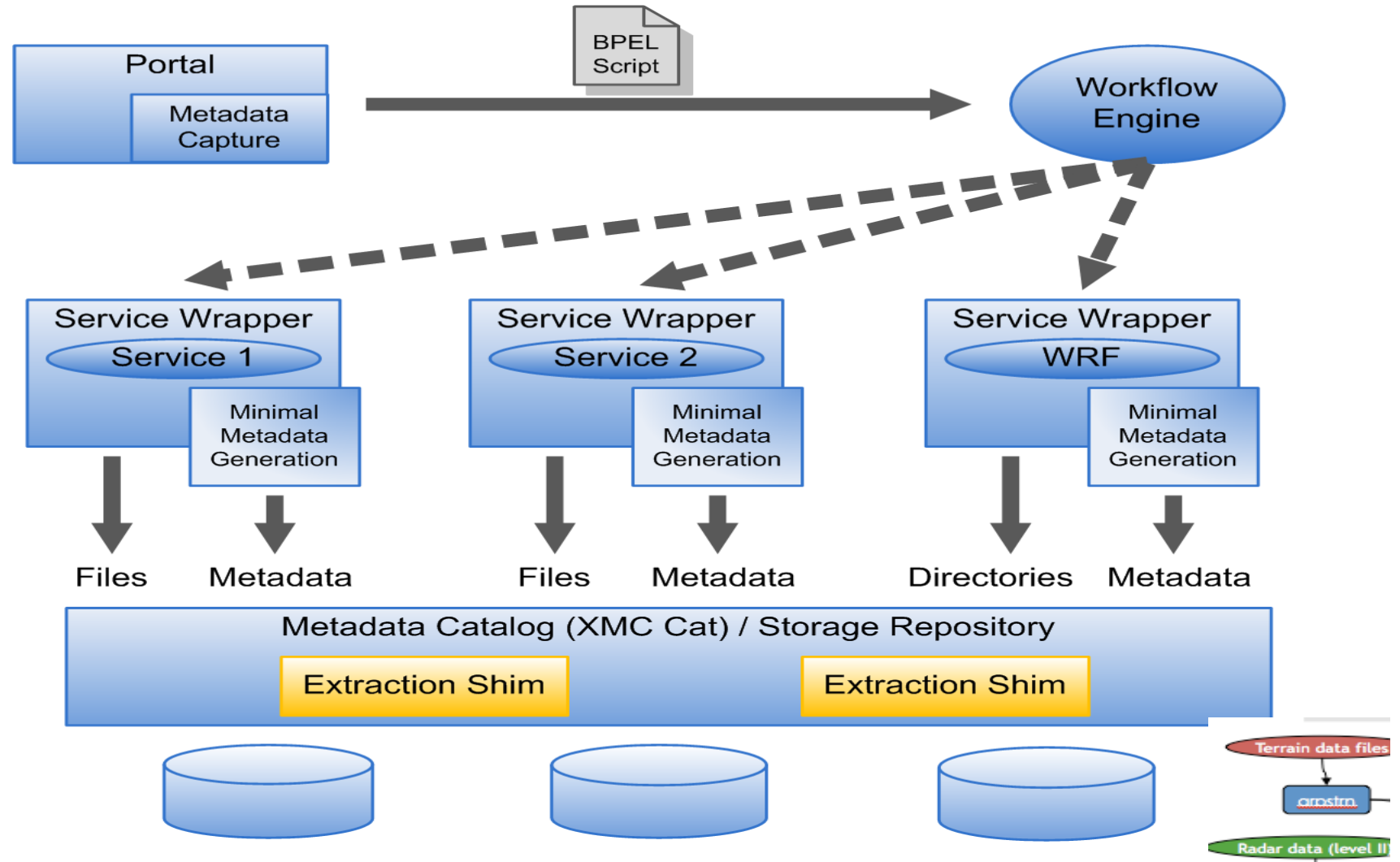
Forecast

Visualization



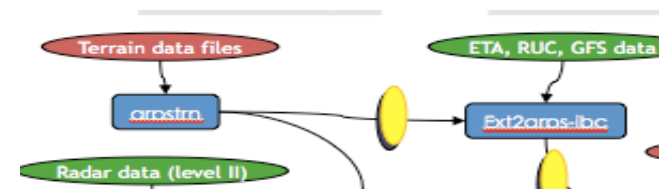
Give me provenance of this product

# Automated capture

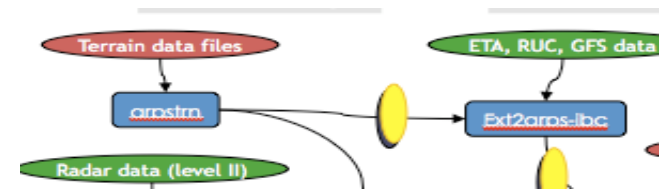
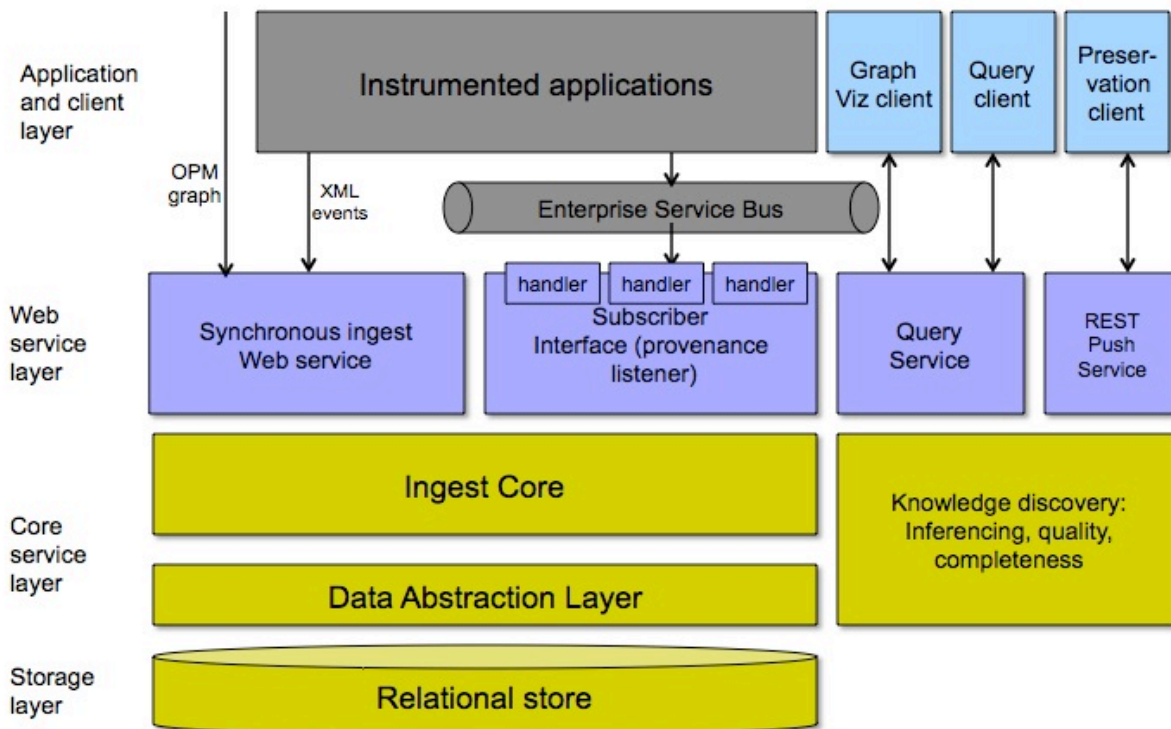


# Forms of provenance capture

<b>User Annotation</b>	<b>Scavenging</b>	<b>Full Provenance Instrumentation</b>
Low application burden	Low application burden	High application burden
High human burden	Low human burden	Low human burden (high developer setup cost)
Error rates and omissions can be high leading to incomplete information	Incomplete information	Complete information

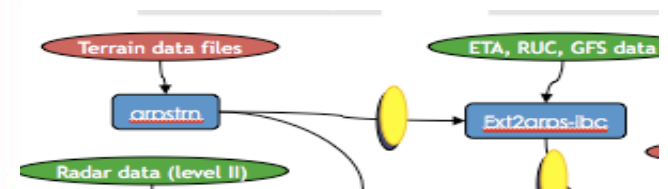
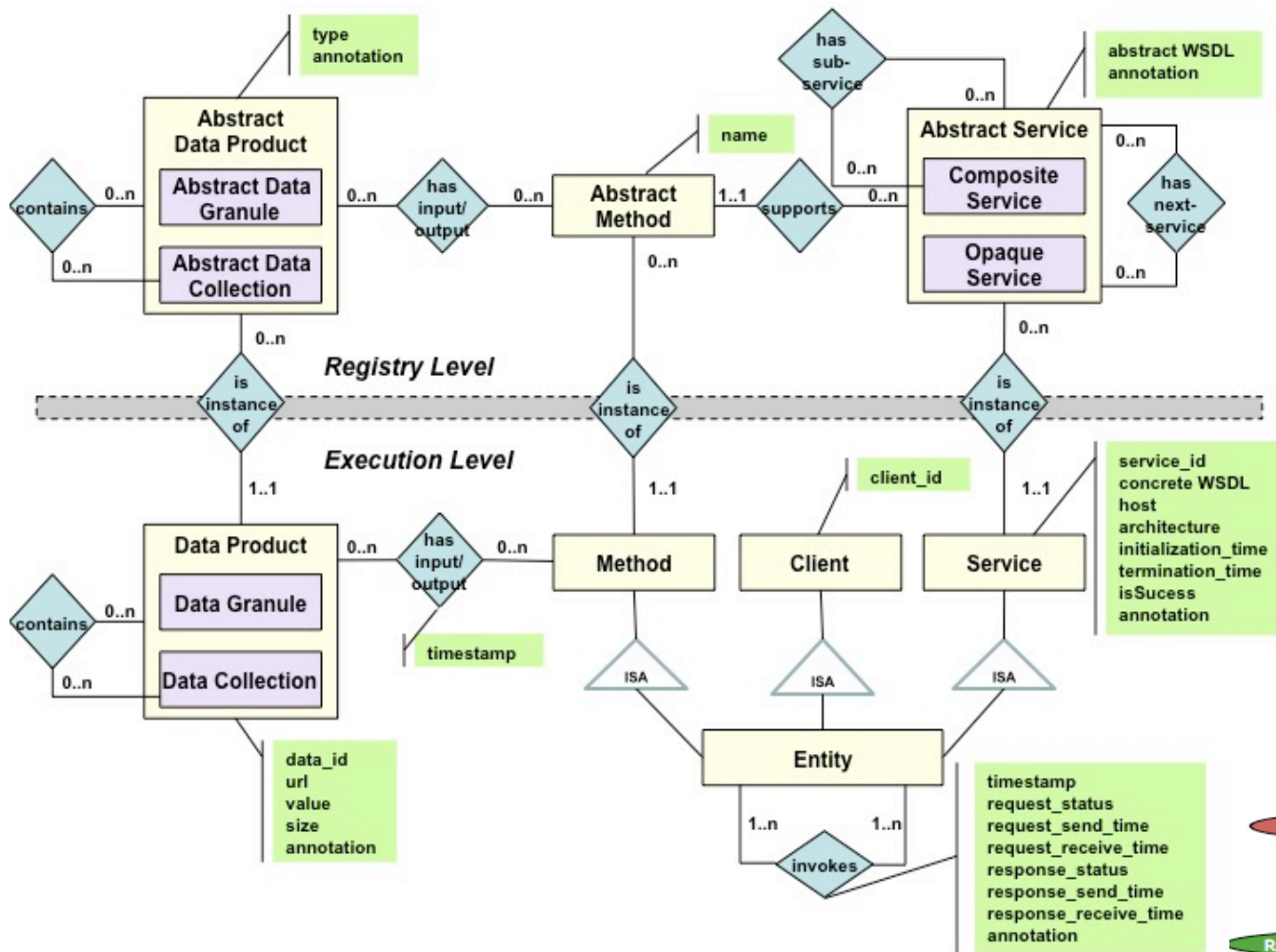


# Architecture for Provenance Capture, Storage, and Analysis and Discovery

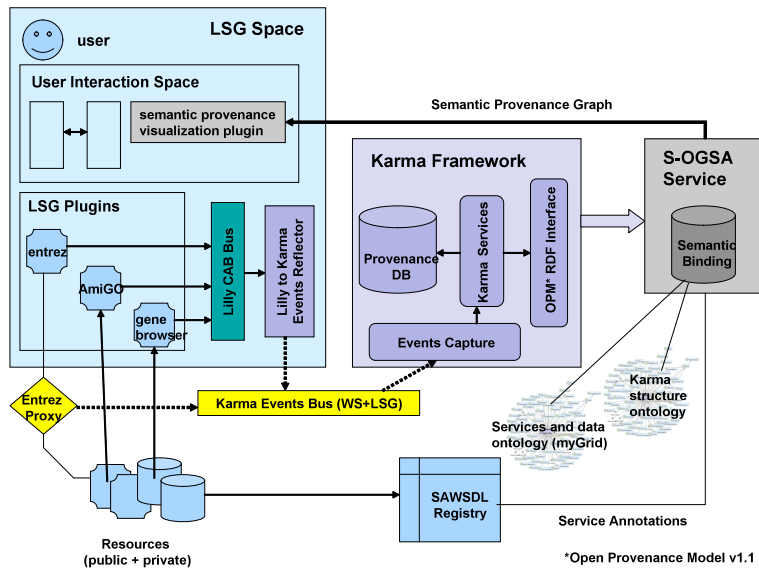




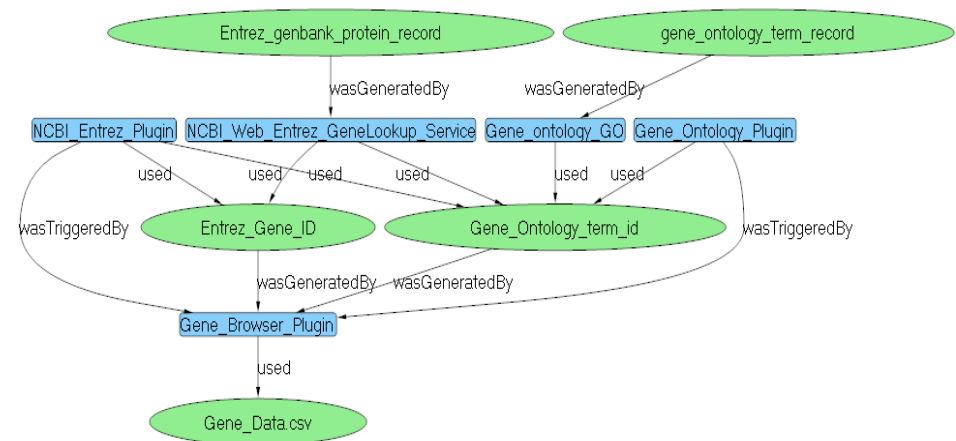
# Data model for provenance representation



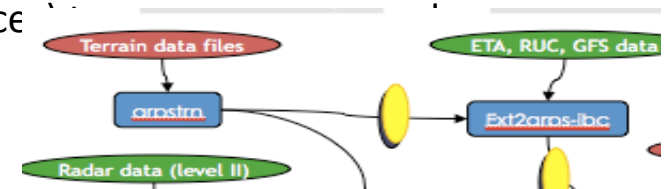
# Application to the Life Science Grid

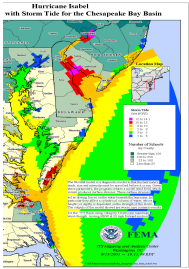


- Life Science Grid (LSG): open source cyberinfrastructure for drug discovery



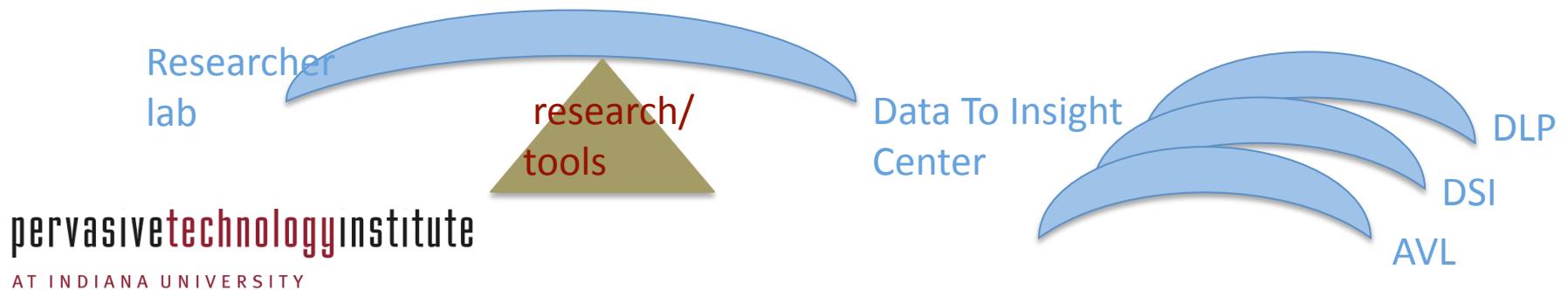
- Karma: captures where and how an artifact originates, and processes previously applied to an artifact.
- S-OGSA adds semantic annotations (services and user preference)
- Collaborators: Eli Lilly and University of Manchester.

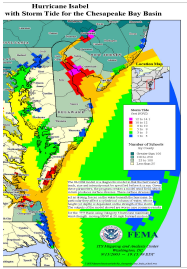




# Data To Insight Center Scientific Mission

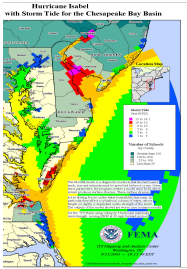
The mission of Data to Insight Center is to bring *first class research* in data-driven technology to *interdisciplinary problems* and to *provide value* to the citizens and business of the State of Indiana and beyond.





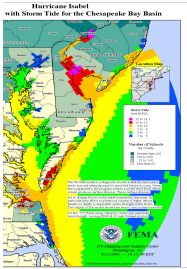
## Center Goals

1. Establish open structures for involvement in Center by IU faculty and staff; and beyond.
2. Aid in the creation and growth of interdisciplinary research through identification of and channeling effort channeling into **research thrusts**
3. Perform basic and applied research within or across thrusts with focus on large scale projects
4. Harden selected data-focused technology solutions that have broader use potential (translational research)
5. Prototype new data-focused technology with opportunity for broader use potential
6. Train students, postdocs



# Purpose of Center Thrusts

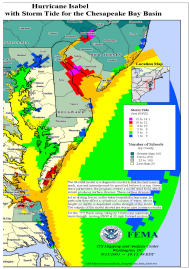
- The Data To Insight Center adopted an organizational model of using thrusts as a way of focusing effort.
  - A *thrust* is an area of research focus that concentrates efforts and provides clarity in pursuit of opportunities.
- Three thrusts identified for 2009-2011.
- Thrusts are reviewed every 2 years to assess their continued viability to stimulate.
- The Center may also engage in additional thrusts that are shared across one or more of the PTI centers.



# 2009-2011 Thrusts

## *Scientific data preservation*

- Long term preservation of scientific data collections
- *Sustainability, climate and the environment*
  - Sustainability science deals with interactions between natural and social systems and how those interactions conserve the planet's life support system. Climate and the environment are broadly defined as encompassing atmospheric, oceanographic, hydrologic, ecological, and earth physical systems.
  - Data are geo-located social, observational, and derived data. Includes but not limited to visualization; data provenance and preservation; leverages campus strength in human-oriented sustainability
- *Data at scale*
  - IU's significant storage and compute resources create opportunities for understanding how to deal with large-scale data, that is, data at scale.
  - Includes data mining over large scale data sets, visualizations and intelligent user interfaces that connect users and their goals (scale in user diversity); challenges in integration and interoperability across databases. Data from devices distributed across a population. Incentives and barriers to data sharing.



# Researcher Engagement in Center

## Seminar Series:

- Jointly (DSI and DLP) host several speakers in '09-'10 year.
- Distribute talks at Lindley Hall, Library, Innovation Center; extend welcome to IUPUI and host venue there.
- Fellows Program
  - Proposal driven; contribute to center's thrusts and goals
  - 2010
- Consortia affiliation

Thank you.

Beth Plale

[plale@indiana.edu](mailto:plale@indiana.edu)

XMC Cat

Scott Jensen [scjensen@indiana.edu](mailto:scjensen@indiana.edu)

Karma

Bin Cao [bincao@indiana.edu](mailto:bincao@indiana.edu) and Girish Subramanian

<http://www.dataandsearch.org/dsi/research>

