



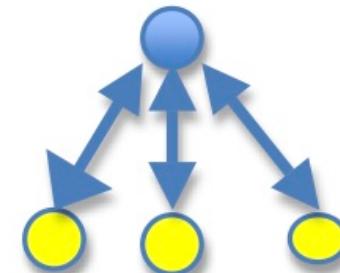
# Lifemapper Provenance Virtualization

Nadya Williams (UCSD)  
Aimee Stewart (KU)  
Quan Zhou (IU)  
Yuan Luo (IU)  
Beth Plale (IU)  
Phil Papadopoulos (UCSD)



# Outline

- **Cyber-infrastructure**
- **Domain science**
- **Provenance collection framework**



# What are we trying to do

- **Domain scientist's viewpoint**
  - Show that Lifemapper is working in a new configuration
  - Gain information from current/archived jobs
- **Data scientist's viewpoint**
  - Captured provenance: generic elements and specific Lifemapper elements
- **Cyber-infrastructure viewpoint**
  - Practical use of PRAGMA cloud
  - What is needed as a complete system – ease a burden of integrating hardware and software
  - What is missing and what can be useful

# PRAGMA experiment with Virtual Clusters and metadata capture

Lifemapper user portal

University  
of Kansas (KU)



Overflow jobs  
are sent  
to Virtual Cluster

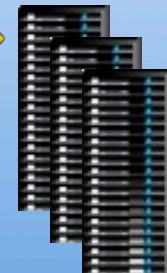
2

Submit species  
modeling and  
distribution  
experiment



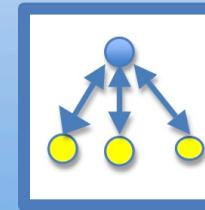
UC San Diego  
(UCSD)

ROCKS



Lifemapper virtual cluster  
on PRAGMA cloud

Indiana  
University (IU)



Provenance Data  
captured on VC,  
sent to Karma

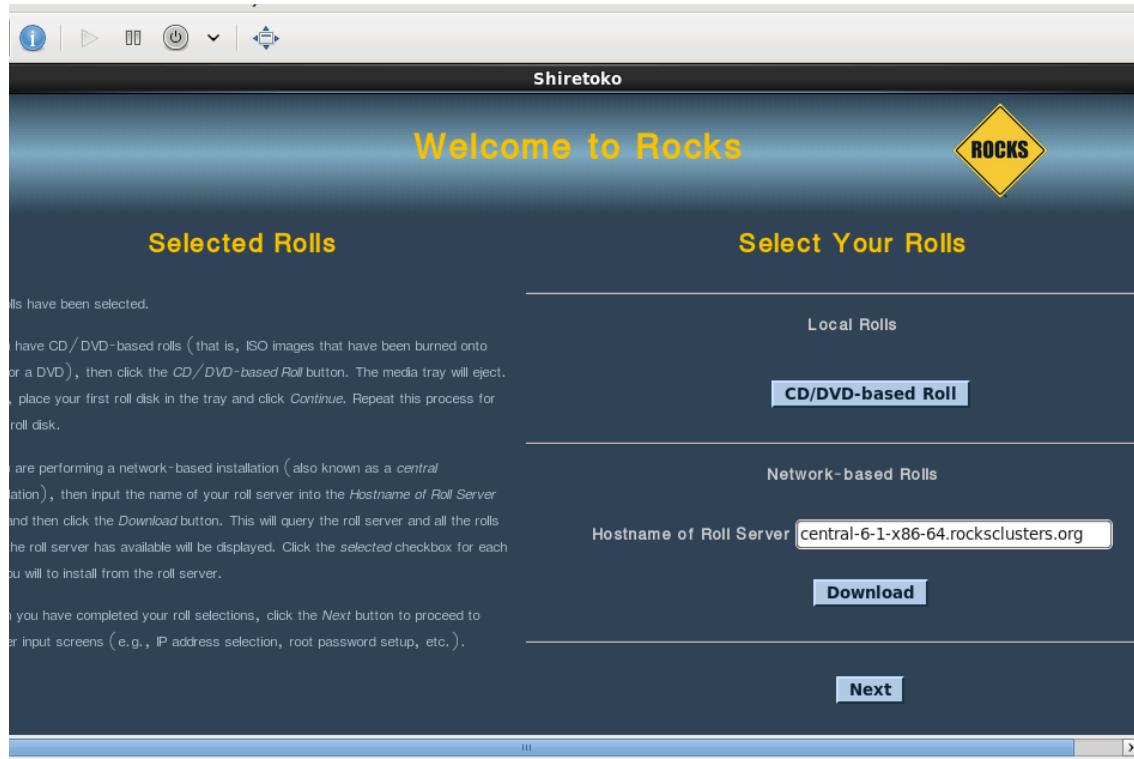
3b

View experiment  
provenance

Cytoscape

Karma provenance repository  
and analysis

# Building VC: regular rocks build



## First incarnation

```
# rocks add cluster fqdn="rocks-204.sdsc.edu" \
  ip="198.202.88.201" fe-name=rocks-204 \
  num-computes=2\
# rocks start host vm rocks-204
# virt-manager
# insert-ethers  ( to install compute nodes)
# rocks start host vm hosted-vm0-0-0
```

compute  
nodes

## Next rebuild

```
# rocks stop host vm rocks-204
# rocks set host boot rocks-204 action=install
# rocks start host vm rocks-204
# virt-manager
# rocks set host boot compute-0-0 action=install
# rocks run host compute reboot
```

# Building VC: add rolls

Selected Rolls			
Roll Name	Version	Arch	Id
ganglia	6.1	x86_64	Net
os	6.1	x86_64	Net
web-server	6.1	x86_64	Net
perl	6.1	x86_64	Net
sge	6.1	x86_64	Net
zfs-linux	0.6.0.rc12	x86_64	Net
kernel	6.1	x86_64	Net
area51	6.1	x86_64	Net
hpc	6.1	x86_64	Net
base	6.1	x86_64	Net
java	6.1	x86_64	Net

Selected	Roll Name	Version	Arch
<input type="checkbox"/>	area51	6.1	x86_64
<input type="checkbox"/>	base	6.1	x86_64
<input type="checkbox"/>	ganglia	6.1	x86_64
<input type="checkbox"/>	hpc	6.1	x86_64
<input type="checkbox"/>	java	6.1	x86_64
<input type="checkbox"/>	kernel	6.1	x86_64
<input type="checkbox"/>	kvm	6.1	x86_64
<input checked="" type="checkbox"/>	lifemapper	6.1	x86_64
<input type="checkbox"/>	os	6.1	x86_64
<input type="checkbox"/>	perl	6.1	x86_64
<input checked="" type="checkbox"/>	provenance	6.1	x86_64
<input checked="" type="checkbox"/>	python	6.1	x86_64
<input checked="" type="checkbox"/>	rocks-204.sdsc.edu-restore	2013.10.07	x86_64

# Building VC: cluster info

rocks-204 Virtual Machine

Menu Machine View Send Key

Help

Fully-Qualified Host Name:  
Enter the fully-qualified domain name  
of the cluster (optional).

Name:  
Name of the cluster (optional).

Organization:  
Name of your organization. Used when  
certifying this host (optional).

Locality:  
(optional).

State:  
(optional).

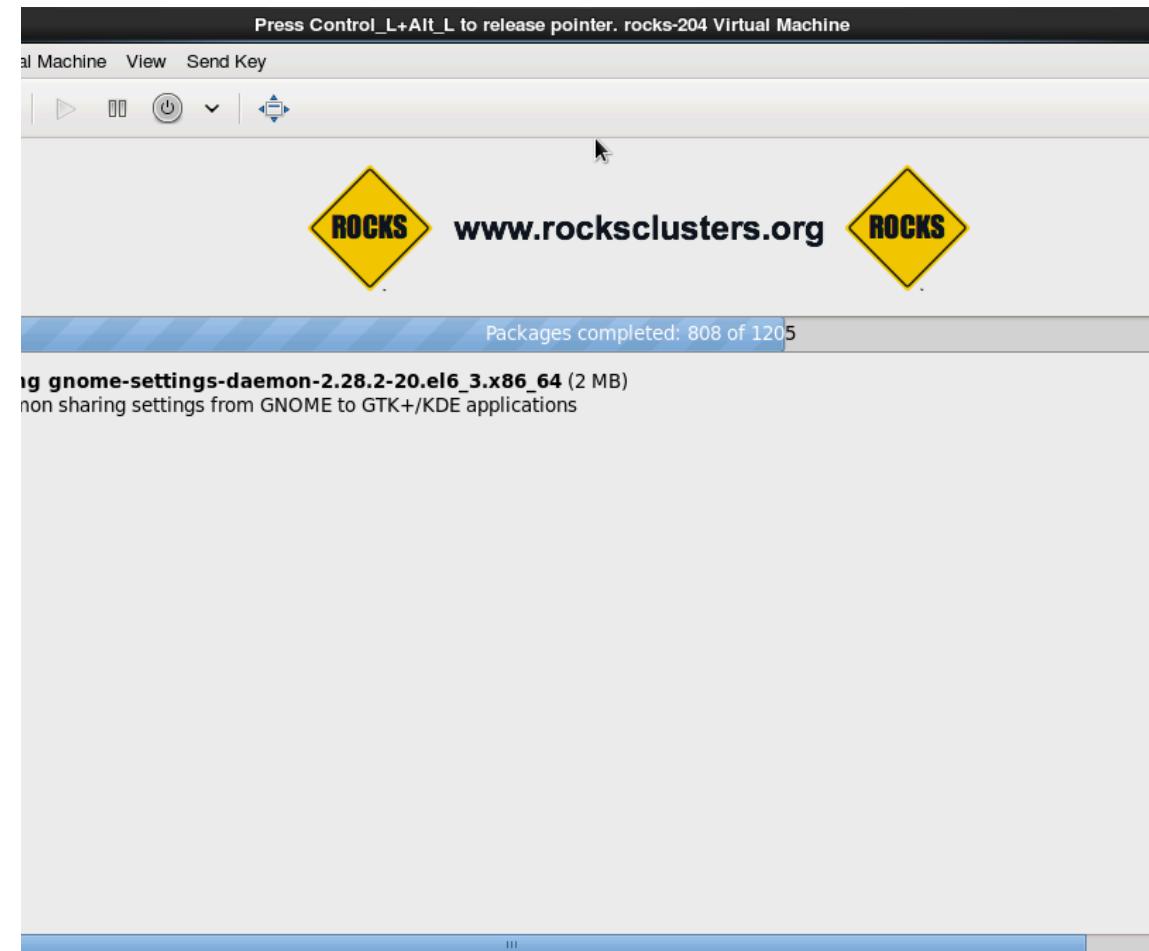
Country:  
Country (optional).

Address for the cluster admin (optional).

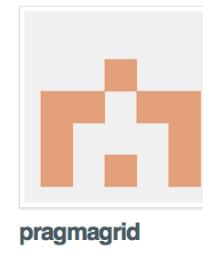
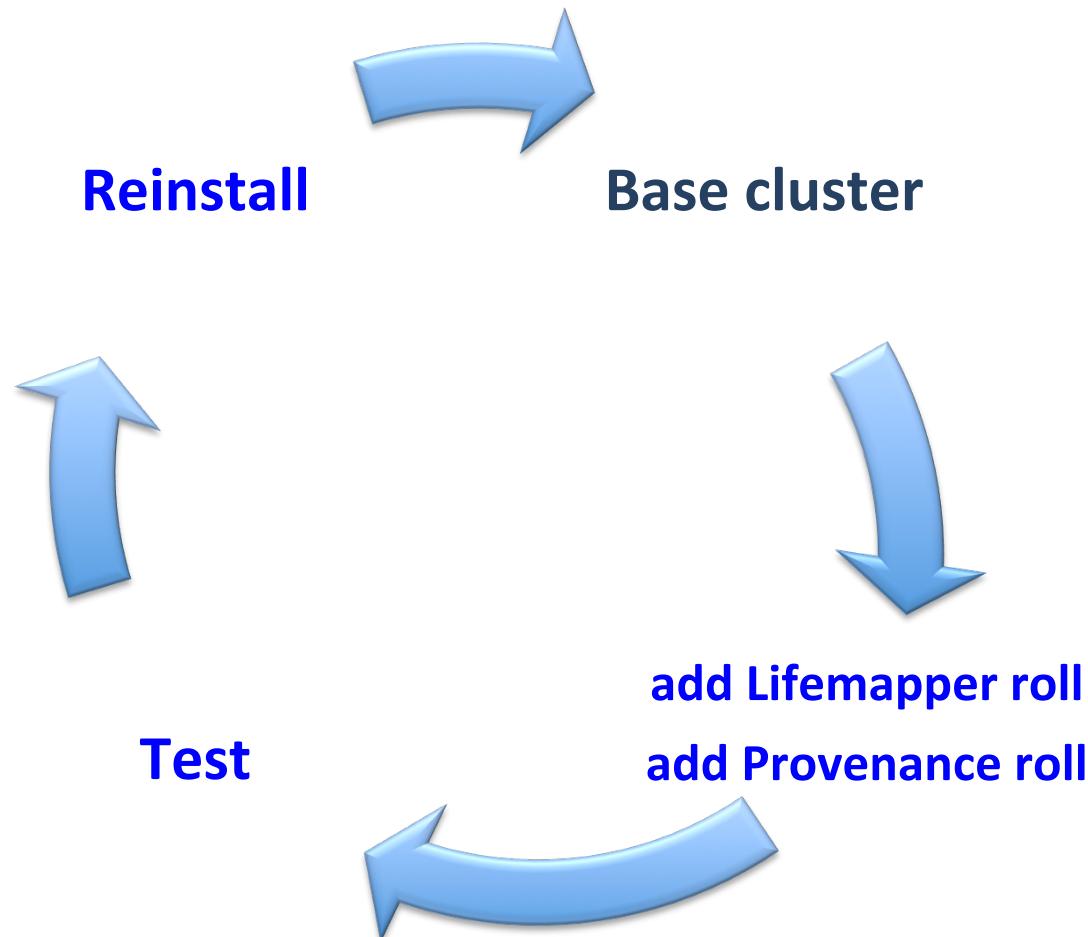
Cluster Information

Fully-Qualified Host Name	rocks-204.sdsc.edu
Cluster Name	Rocks-Cluster
Certificate Organization	SDSC
Certificate Locality	San Diego
Certificate State	California
Certificate Country	US
Contact	admin@place.org
URL	<a href="http://www.place.org/">http://www.place.org/</a>
Latitude / Longitude	N32.87 W117.22

Back Next



# VC deployment – test stage

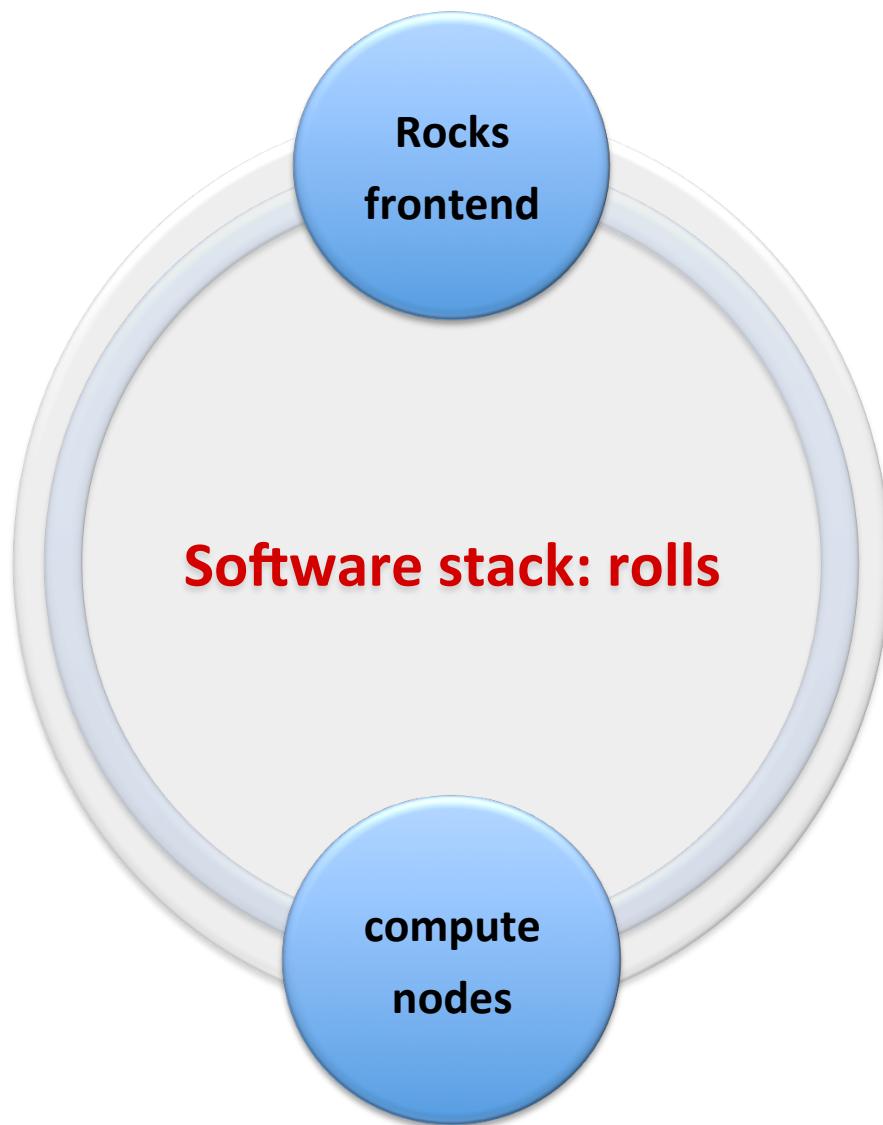


**github**  
SOCIAL CODING



<https://github.com/pragmagrid/provenance/>  
<https://github.com/pragmagrid/lifemapper/>

# Virtual cluster key parts



## Cluster

- Create and Deploy into existing environment

## Rolls

- **Lifemapper**
  - Dependencies: gdal, openmodeller, proj, tiff
  - Lifemapper compute module
  - Configuration
- **Karma provenance tools**
  - Karma server, client, adaptor, visualization plugin
  - Dependencies: erlang, rabbitmq, cytoscape
  - Configuration

## Virtual hosting environment

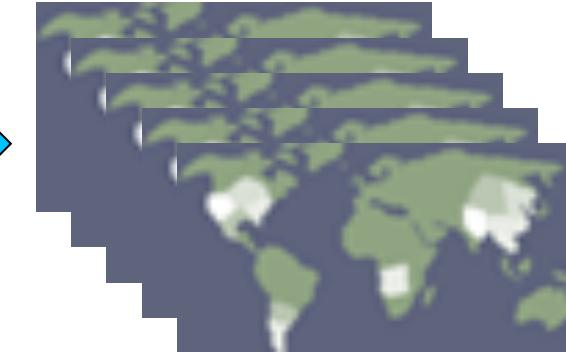
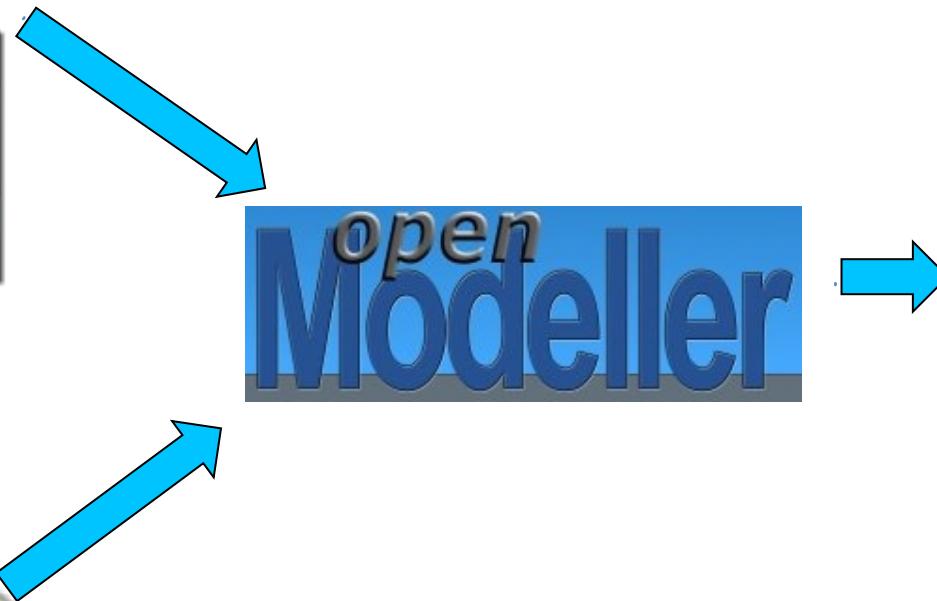
- Deploy into new environment

# Domain science

- An archive of species distribution data
- Web services for biodiversity research tools and data
  - LmSDM
  - LmRAD
- Metadata for everything
- Clients for easy access

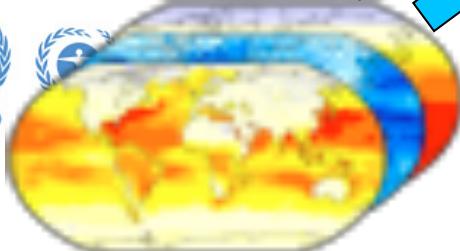


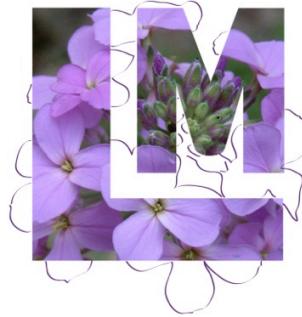
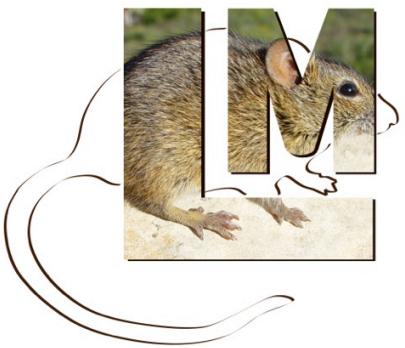
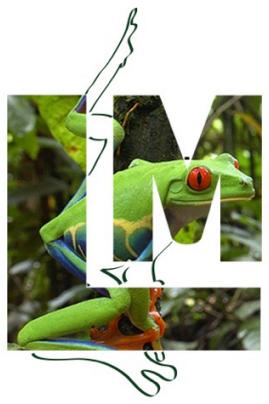
# LmSDM: Species Distribution Modeling



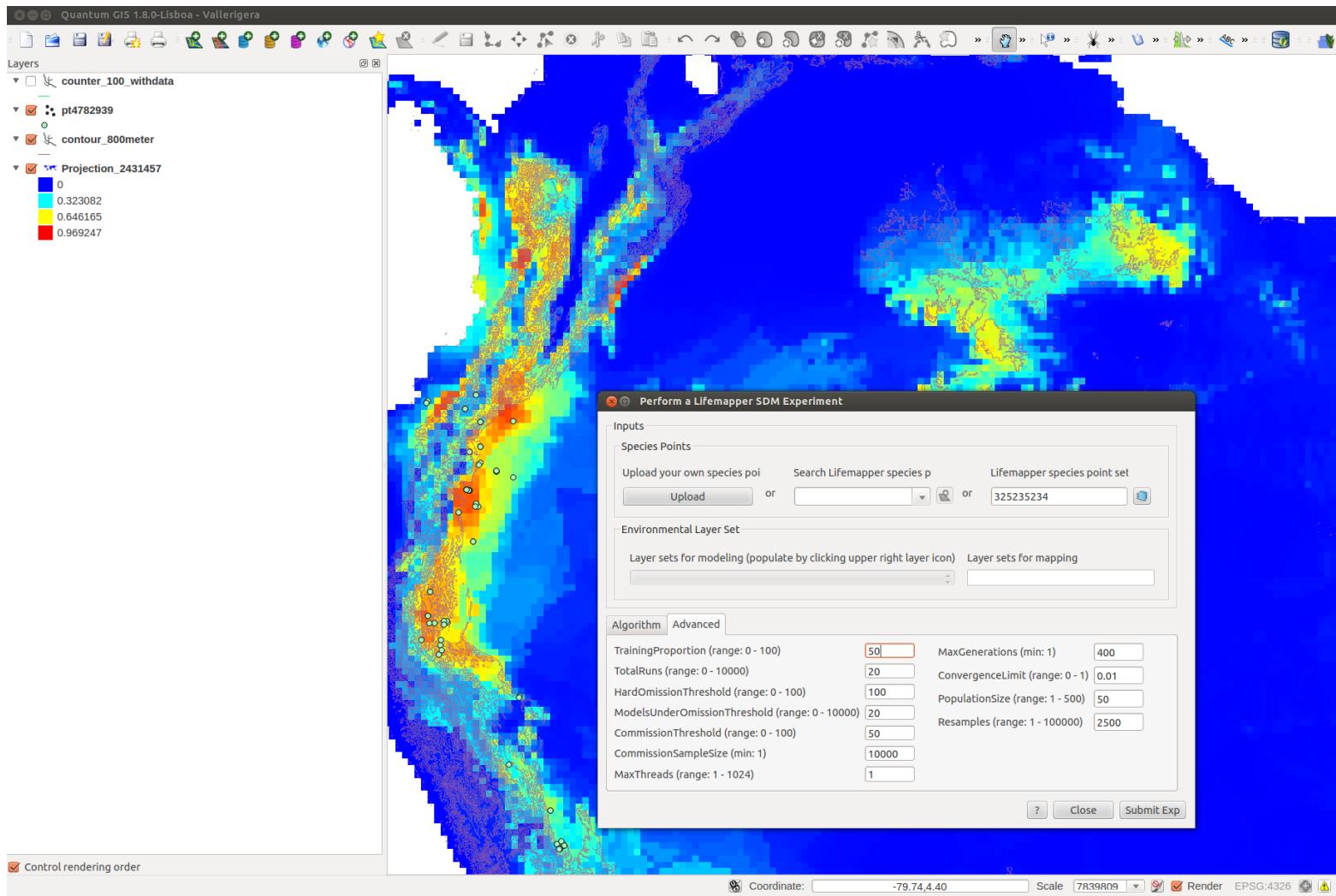
**IPCC**

INTERGOVERNMENTAL  
PANEL ON  
CLIMATE CHANGE

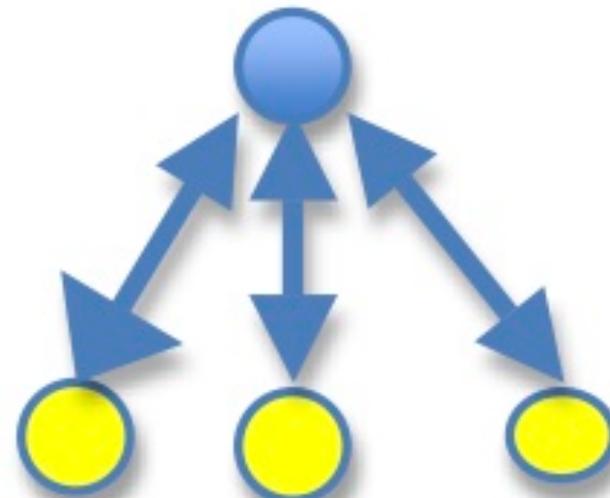




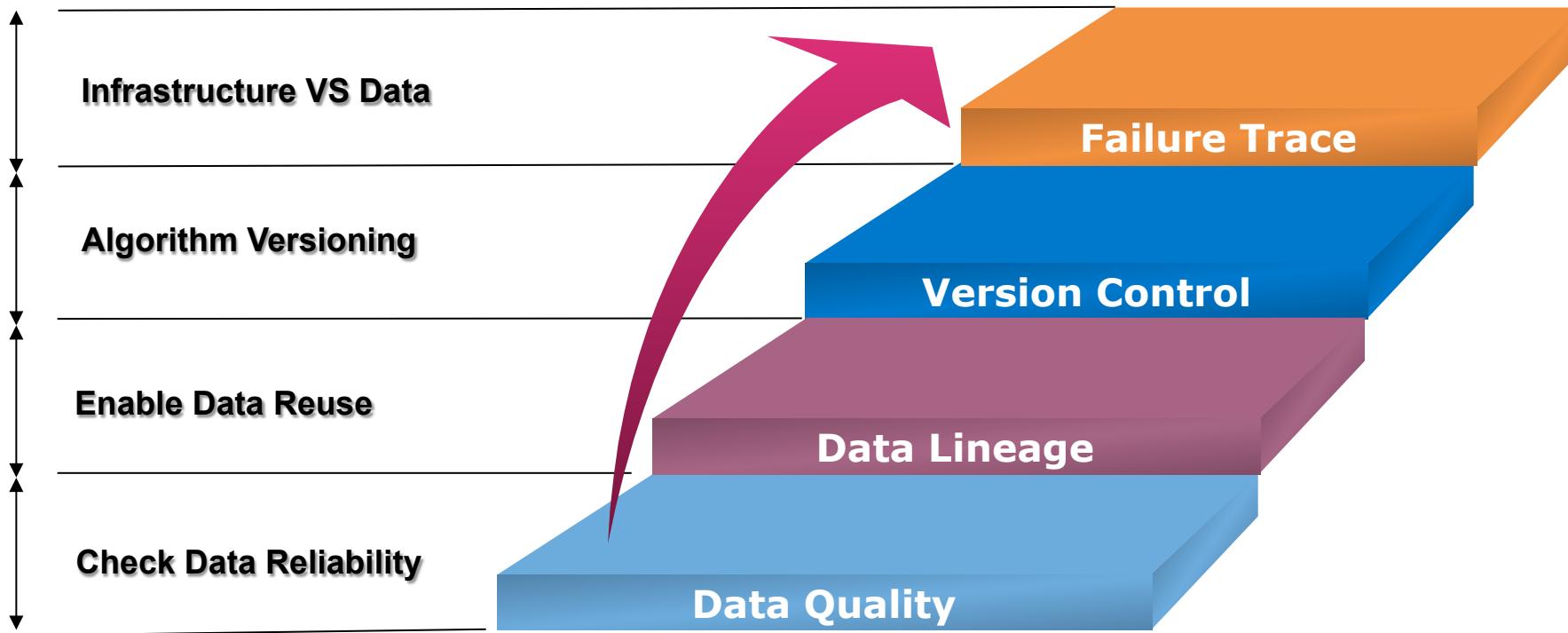
# Submit a Lifemapper SDM experiment from QGIS



# Provenance Collection Framework



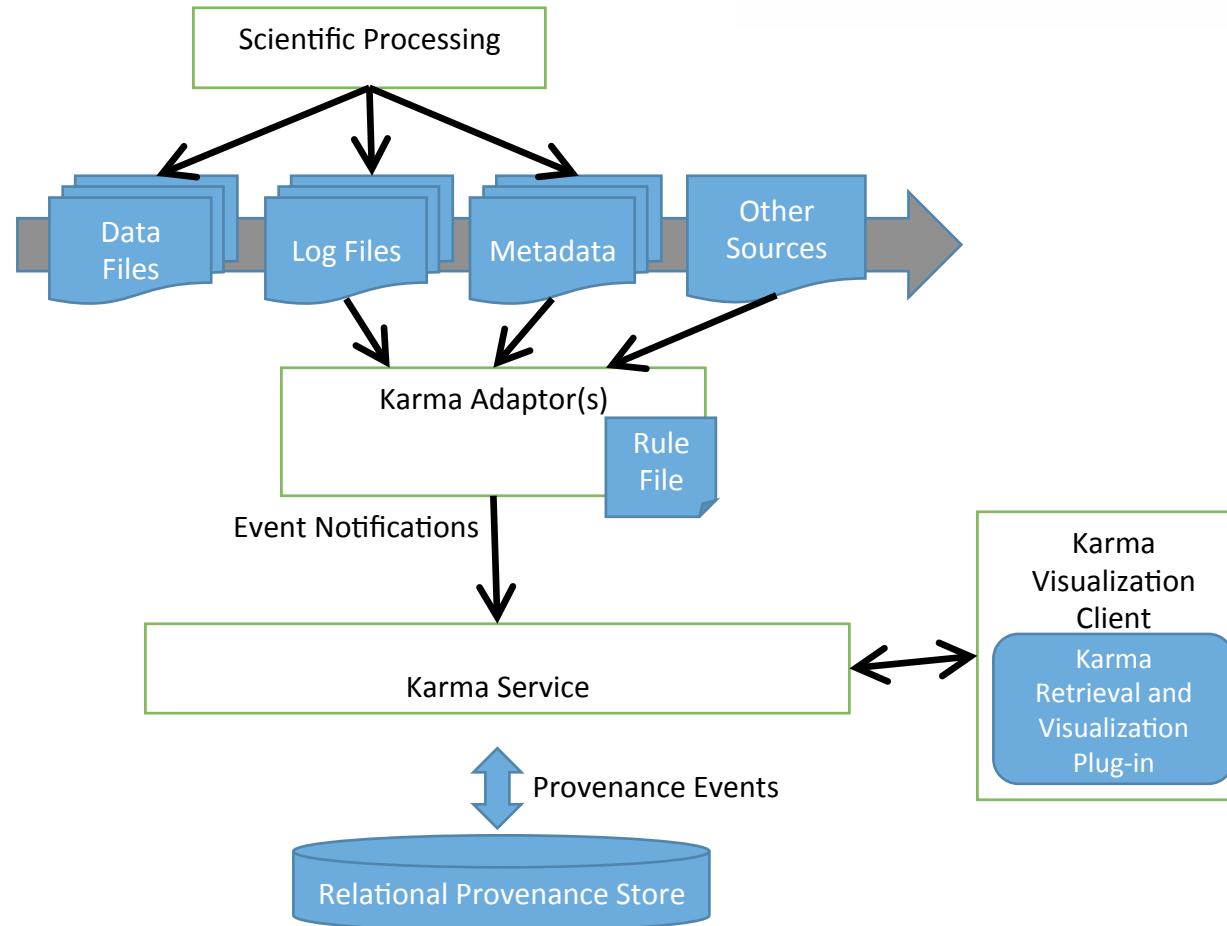
# Why provenance?



# Karma Provenance Collection Tool

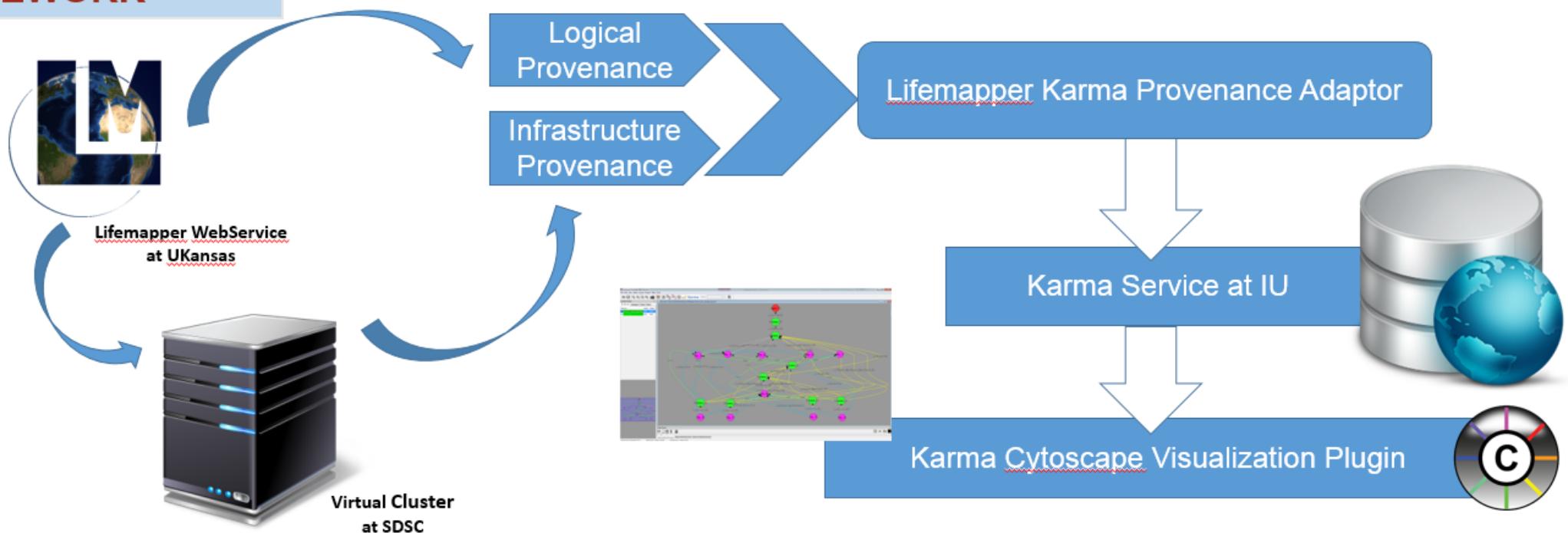


DATA TO INSIGHT CENTER  
INDIANA UNIVERSITY  
Pervasive Technology Institute

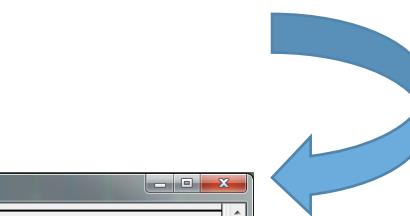
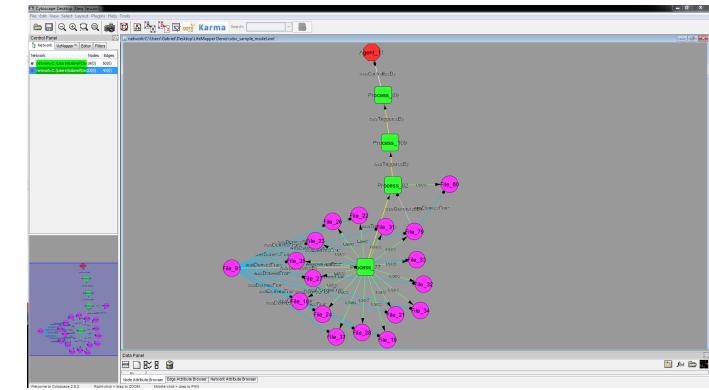
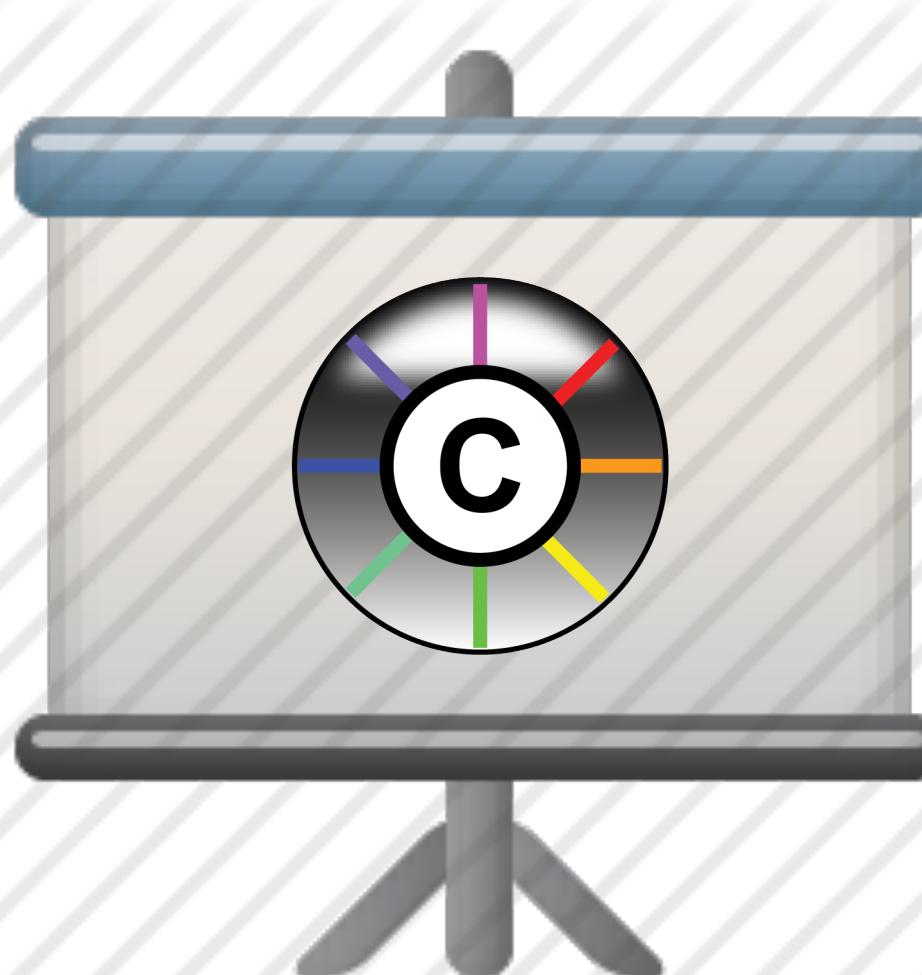
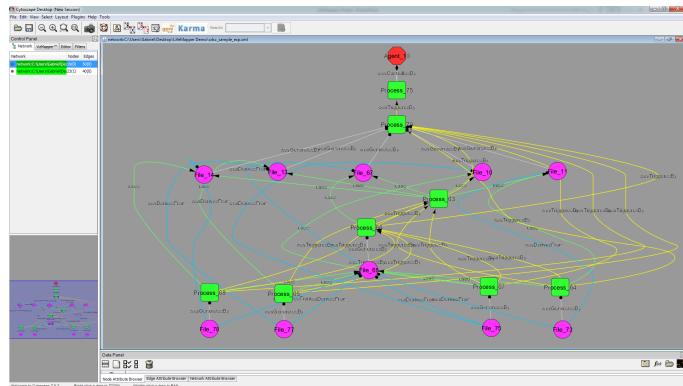


# Framework

## FRAMEWORK

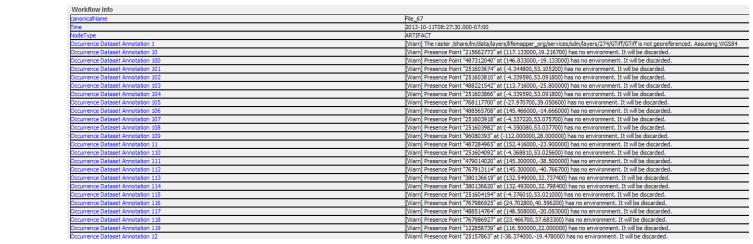


# Demo

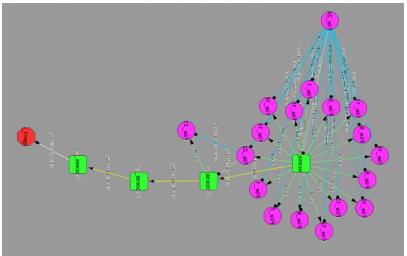


Node Attributes	
applicationID	11404052-40
process-serviceID	SGE_Job/812
process-workflowID	http://lfenner.org/services/sdm/experiments/674096/model
Time	2013-10-04T16:04:58.000-04:00
process-type	SERVICE
NodeType	PROCESS
account	sge
cpu	153.654639
end_time	1380917313
exit_status	0
failed	0
group	nadya
hostname	compute-0-1.local
io	646.795852
low	0.000000
job ID	312
jobname	runlmJob.sh
maxvmem	717987840.000000
mem	42.195375
owner	nadya

# Data Provenance

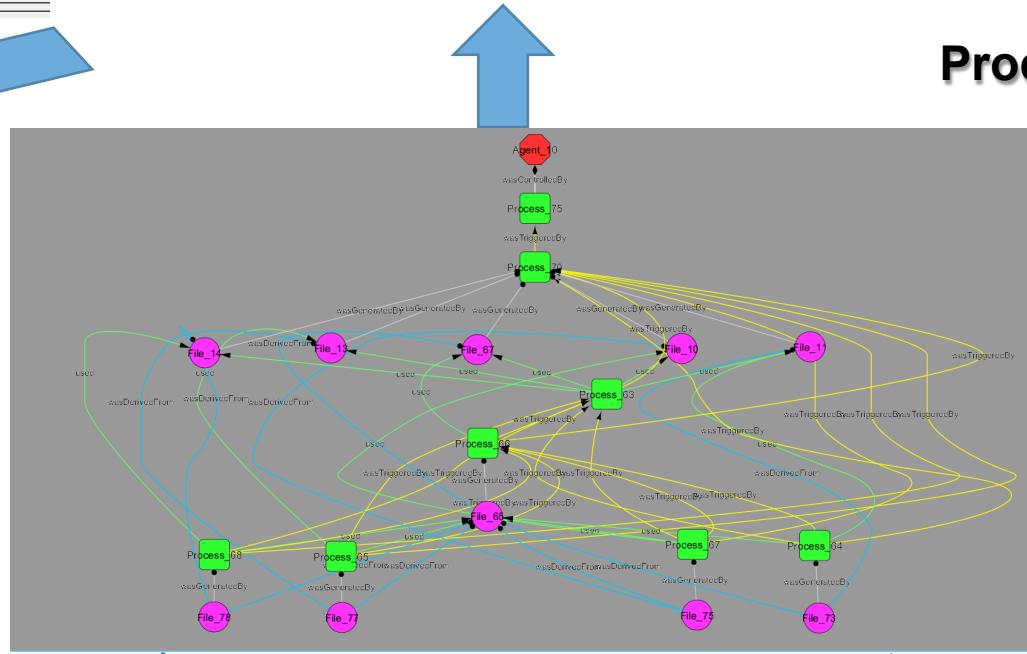
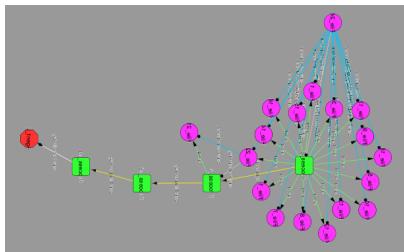
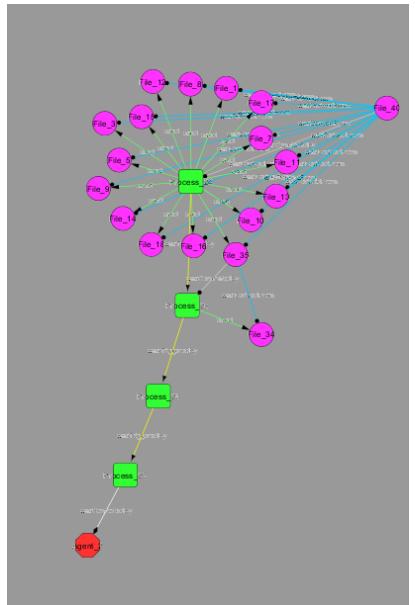


# **Component Process Provenance**

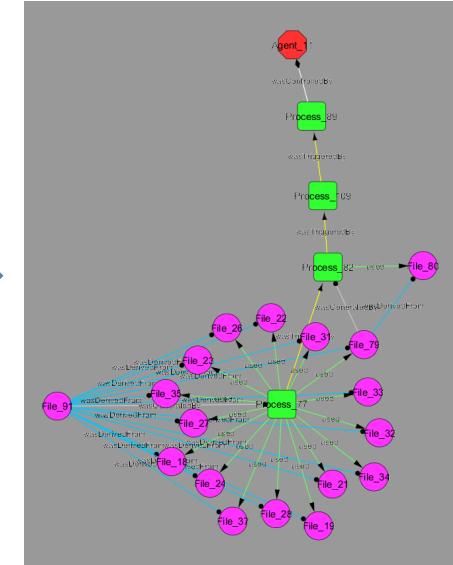
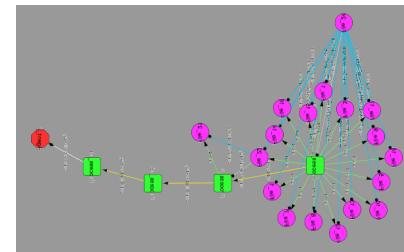


Node Attributes	
Workflow Info	
archive	Process_15
archive-on-wait	8GB /30/203
process-workflowID	<a href="http://alfresco.org/services/lbm/projections/270056">http://alfresco.org/services/lbm/projections/270056</a>
Time	2013-10-04T16:03:13.000-04:00
version	0.0.0
nodeType	PROCESS
account	lge
cmi	
ext_time	1380917088
ext_status	0
Failed	0
group	radya
hostname	192.168.0.1.local
ip	0.211939
ios	0.000000
os_id	000
osname	UbuntuJob.sh
osversion	6.02611712.000000
item	1_799279
parent	radya
priority	0
crname	mlq
subj_time	1380916561
last_time	1380916993
Registry Info	

# Process Provenance



# Experiment Provenance

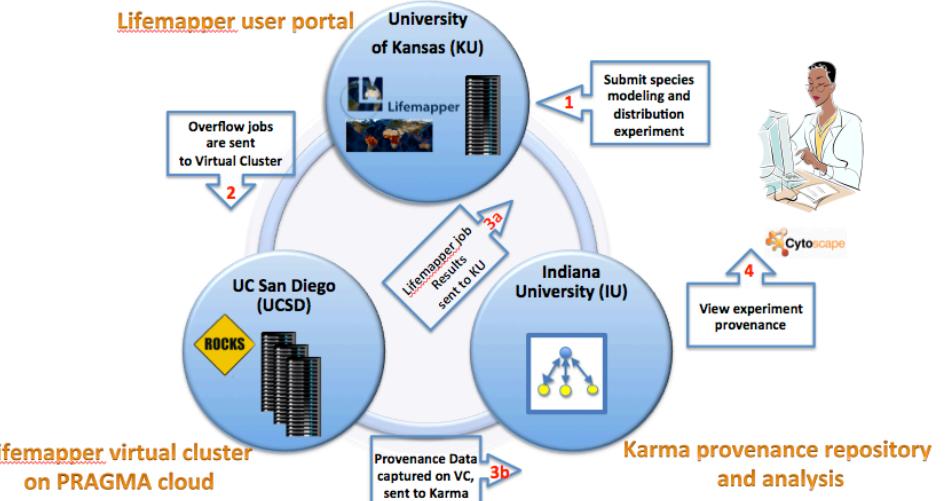


## Future Work

- Extending the Karma adaptor for Lifemapper to perform more system-based gathering of provenance
- Migrating from Open Provenance Model (OPM) to the W3C PROV data model for provenance representation. PROV allows richer expression of relationships, semantic annotations and semantic inference

# Conclusions

- **Results: it works !**
  - Practical use of PRAGMA cloud in a distributed processing environment
  - Have a framework to meet our operational imperatives that can be used as a blueprint
  - Ease of replication
- **Lessons learned**
  - What works in one environment may not apply in another
  - Multiple applications requirements, poor documentation
  - Best tools:



# Future work

## Domain science



Lifemapper

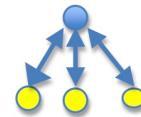
- Include UTM data, metadata catalog
- Use UFL high resolution Mt. Kinabalu imagery
- Assemble multi-species macro-ecology experiment for the area



## Cyber-infrastructure

- Enable overlay network that can span Lifemapper server, Karma server and compute clusters
- Can we handle data for specialized experiments – detached Lifemapper server usage
- Can we handle different amounts of data?
- Can we make it fault tolerant in the event of server/network outages?

## Provenance collection framework



- Extending the Karma adaptor for Lifemapper to perform more system-based gathering of provenance
- Migrating from Open Provenance Model (OPM) to the W3C PROV data model for provenance representation. PROV allows richer expression of relationships, semantic annotations and semantic inference

# Acknowledgements

This work is funded in part by National Science Foundation and NASA grants

PRAGMA

US NSF 1234953

Karma provenance tools

US NSF ACI 1234983

US NSF ACI 1148359

Lifemapper

US NSF EPSCoR 0553722

US NSF EPSCoR 0919443

US NSF EHR/DRL 0918590

US NSF BIO/DBI 0851290

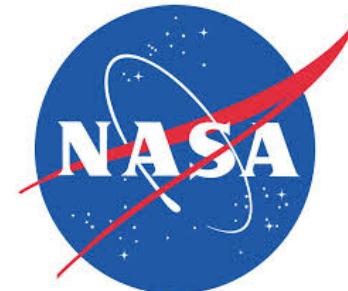
US NSF OCI/CI-TEAM 0753336

US NASA NNX12AF45A

Rocks

US NSF OCI-1032778

US NSF OCI-0721623



# Links/Contacts

## Lifemapper

<http://lifemapper.org>

<https://github.com/lifemapper/>

## Nadya Williams

[nadya@sdsc.edu](mailto:nadya@sdsc.edu)

## Karma Provenance Tools

[http://d2i.indiana.edu/provenance\\_karma](http://d2i.indiana.edu/provenance_karma)

<http://sourceforge.net/projects/karmatool/>

## Aimee Stewart

[astewart@ku.edu](mailto:astewart@ku.edu)

## Gabriel Quan Zhou

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

## Rocks

<http://www.rocksclusters.org>

## Yuan Luo

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

## Pragmagrid GitHub

<https://github.com/pragmagrid/provenance>

<https://github.com/pragmagrid/lifemapper>



**Thank You!**  
**Questions?**