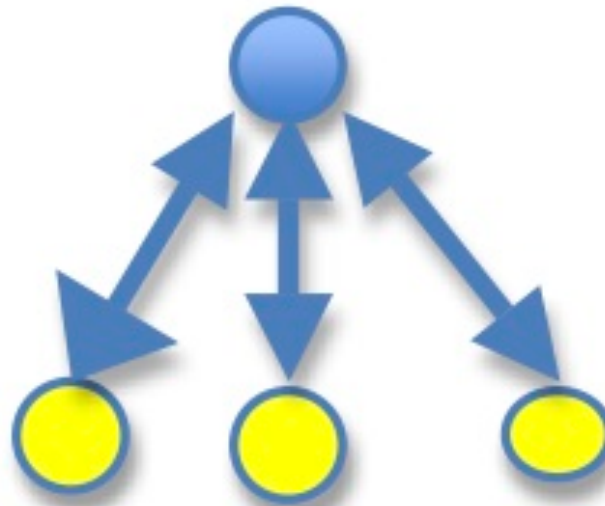


Provenance Collection Framework for Biodiversity Analysis On PRAGMA Cloud for Data Sharing

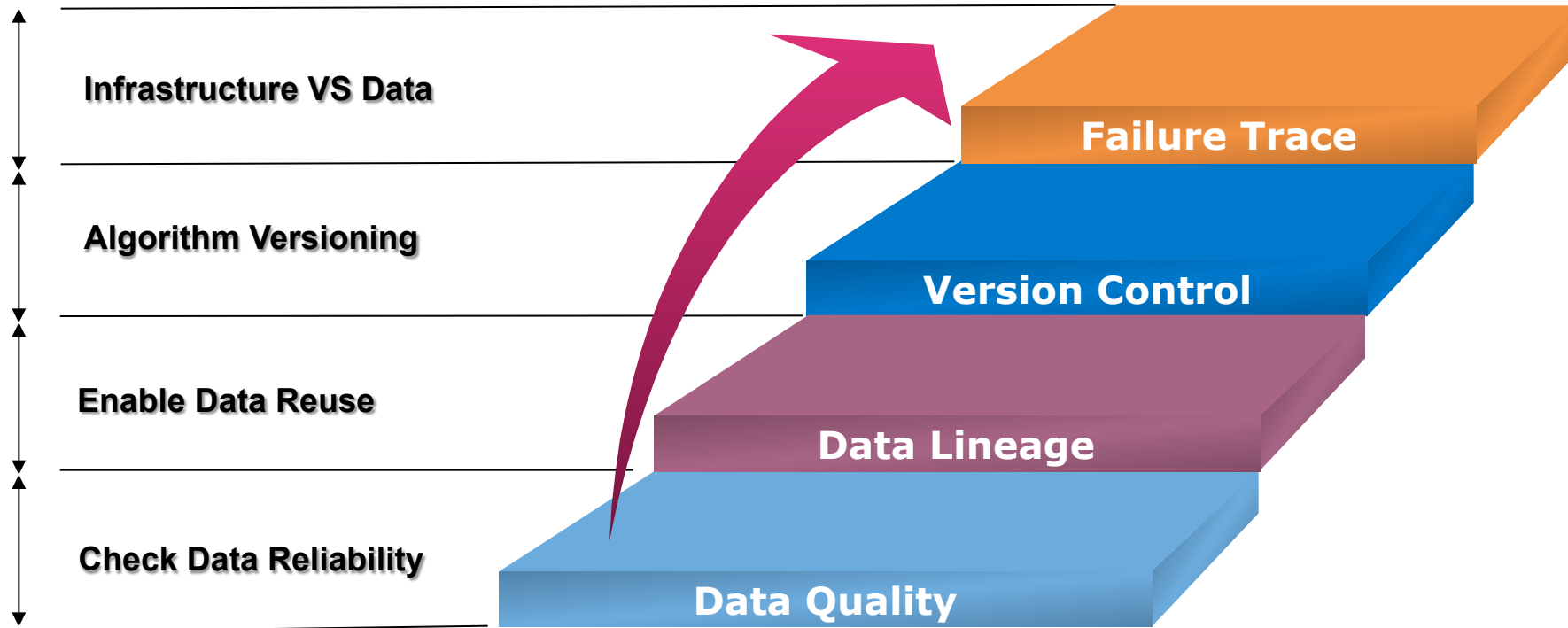
Quan Zhou and Beth Plale
Data to Insight Center, Indiana University Bloomington



DATA TO INSIGHT CENTER

INDIANA UNIVERSITY
Pervasive Technology Institute

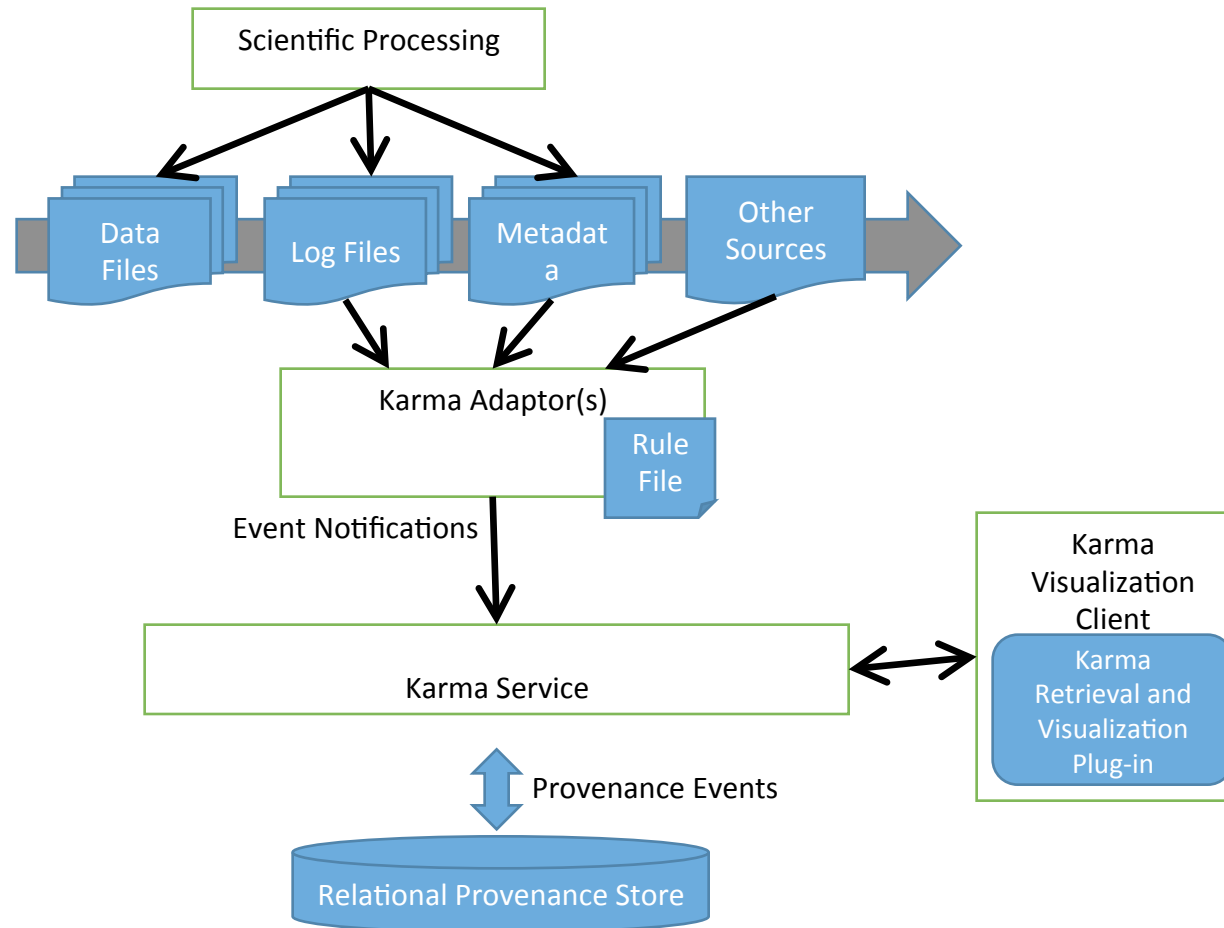
Why provenance?



DATA TO INSIGHT CENTER

INDIANA UNIVERSITY
Pervasive Technology Institute

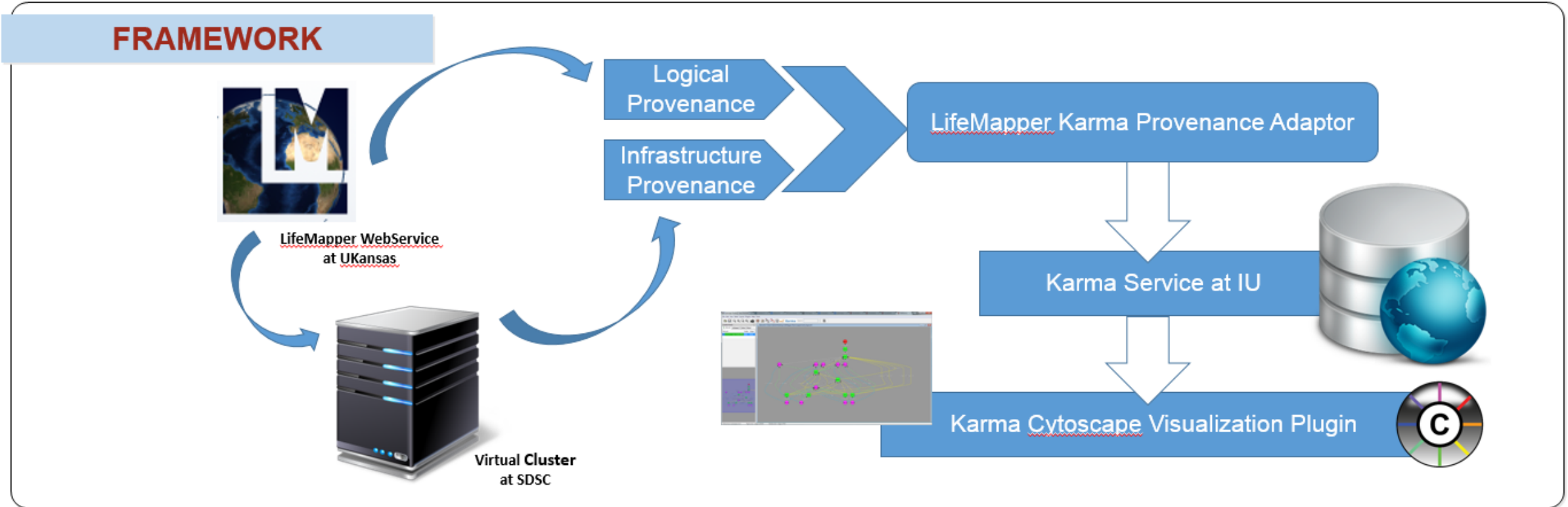
Karma Provenance Collection Tool



DATA TO INSIGHT CENTER

INDIANA UNIVERSITY
Pervasive Technology Institute

Framework

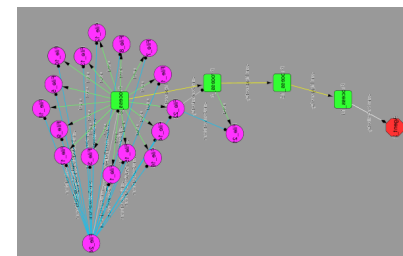


DATA TO INSIGHT CENTER

INDIANA UNIVERSITY
Pervasive Technology Institute

[illegible]

Node Attributes	
Workflow Info	
ExecutionName	process_15
process-dir-cwd	/sge_bak/2013
process-workflowID	http://Metamapper.org/services/ids/ids/projects/2700596
Time	2013-03-15 08:03:13.000+00:00
process-host	hpc-nvcc
hostType	PROCESS
AccountID	sgs
jobID	101-870746
end-time	1309917080
end-attempt	0
parentID	0
group	hadoop
hostname	hpc-nvcc-0-1.local
ip	0.211939
box	0.000000
cpu_id	302
node-name	hpc-nvcc.job.sh
main-memory	602611712.000000
owner	1.799379
owner	hadoop
priority	0
username	all
node-time	13099105961
start-time	1309916993
Registry Info	



Future Work

- ❑ Extending the Karma adaptor for LifeMapper to perform more system-based gathering of provenance by improving the provenance granularity level
- ❑ 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 inferencing



Thank You !



DATA TO INSIGHT CENTER

INDIANA UNIVERSITY
Pervasive Technology Institute