Experiences with Measurement Data Collection on R&E Networks

Christopher Small

Indiana University

The Global Research NOC

- Formed in 1998 to provide operations for the Abilene Network
- Groups
  - Service Desk: 24x7x365 Call Center & Monitoring Center
  - Network Engineering: 16 engineers providing troubleshooting & planning
  - Systems Engineering & Tool Development: 11 engineers
    - Developing & supporting GRNOC toolset and systems
    - Operating research platforms like the Internet2 Observatory and NLRview
    - Measurement collection and visualization
    - GENI Meta Operation Center (GMOC)

The Global Research NOC

CEN

C·I·C 1958 – 2008 OmniPoP

gMOC GENI META-OPERATIONS CENTER

HOPI HYBRID OPTICAL PACKET INFRASTRUCTURE

I-LIGHT Indiana's Optical Network

Indiana GigaPOP indiana.gigapop.net

INDIANA UNIVERSITY

INTERNET2®

IP grid

manLan Manhattan Landing Exchange Point

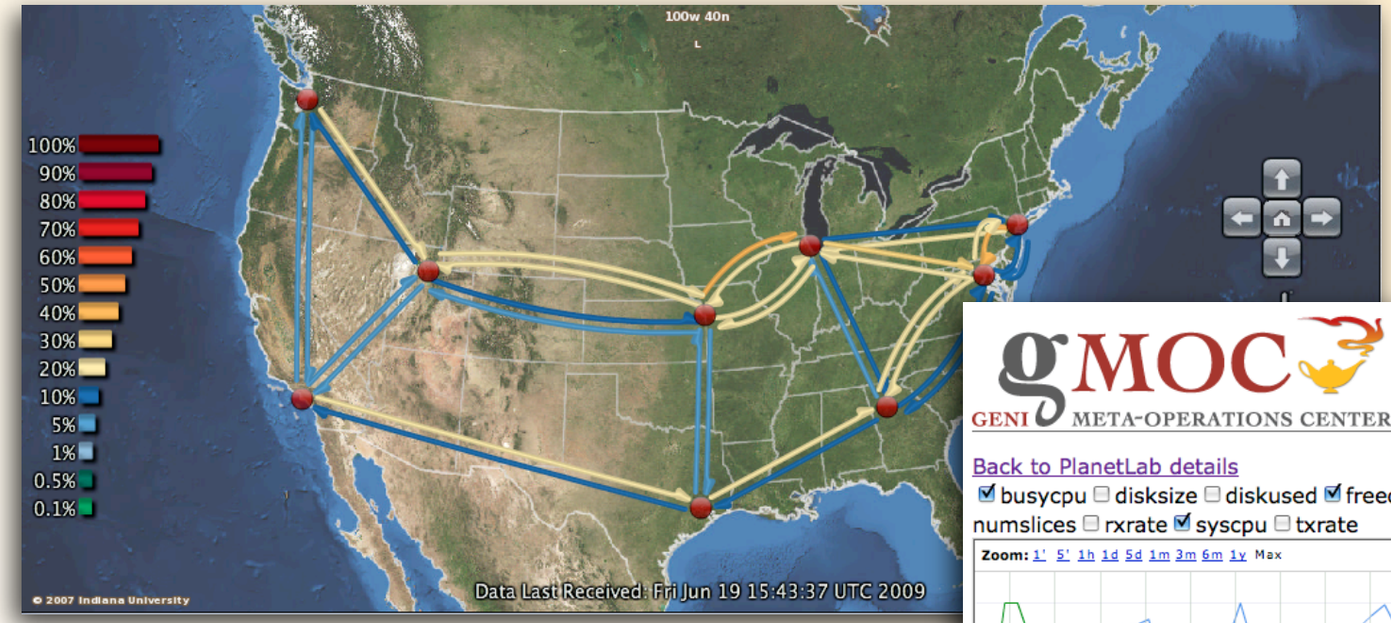
NLR light the future

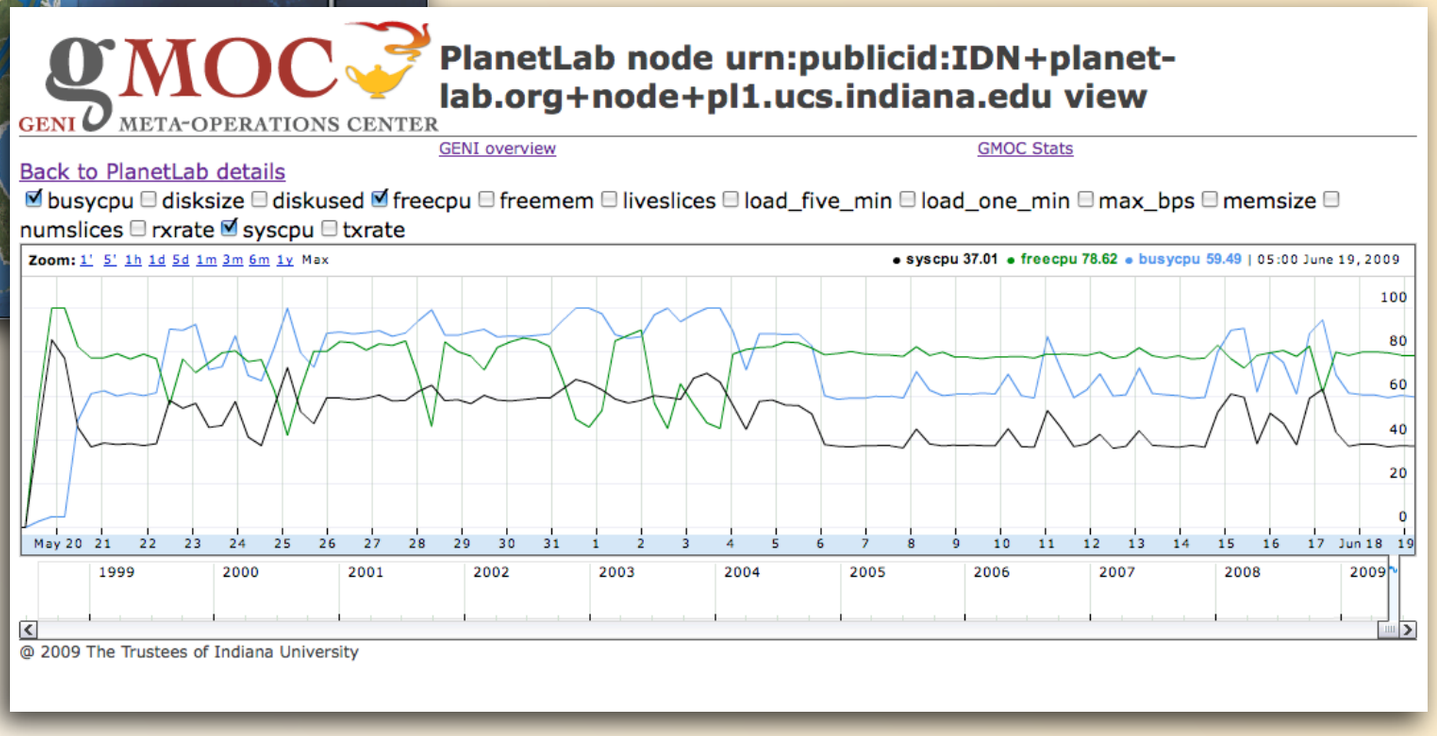
TransPAC2®

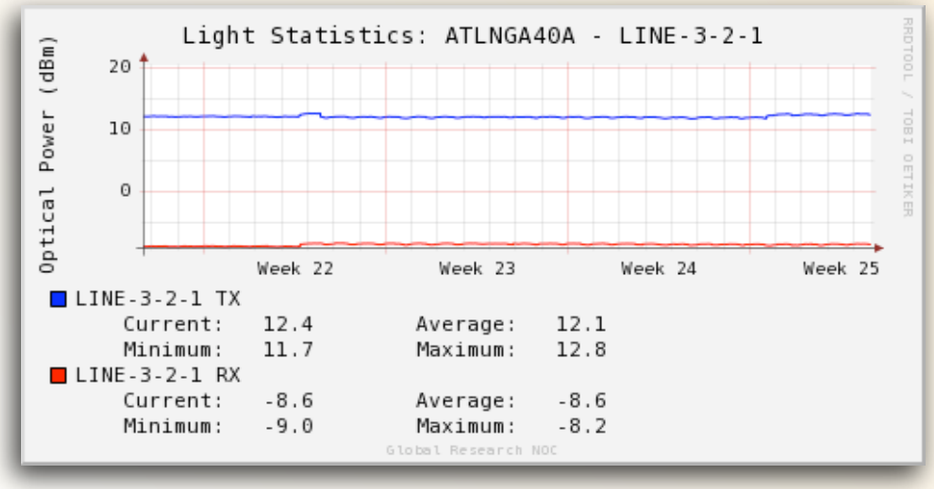
IU GlobalNOC and Measurement

- GENI Meta Operations Center (GMOC)
  - Operations and Management will need inputs to understand the state of the network
- Experience with research infrastructure
  - >200 I2 Observatory, NLRView, CIC, I-Light, Gigapop, IU campus, etc.. measurement points
  - Co-located research equipment
- Substrate operator
  - Currently collect data that can be added to measurement repositories
- Visualization systems

Measurement Data and Visualization







Typical Measurement Points

- 1-10 servers per pop -- 1 and 2U rackmount, central data collection
- 10 gig capable
- Time source (GPS or CDMA)
- Software
  - Netflow/sflow, packet capture
  - Active measurements (BWCTL, OWAMP, NPAD, NDT)
  - Routing data (ISIS and BGP)
  - SNMP stats (Utilization, Environmental, etc..)
  - “Visible Network”, Router/Switch/NE Proxy, LLAMA

Researcher questions (#1)

- What is the state of the network?
  - Researcher often doesn't know what they are looking for but see an anomaly in their data
  - Topology (Virtual and Physical), Outages, Utilization, etc..
  - Can you tell me what happened on the night of Mat 10th 2007?
  - Time period can be exact or very loose
  - R&E networks are fairly static in topology yet hard to describe changes to researchers

Researcher questions (#2)

- Where is the data?
  - Data is often hard to find
  - Human readable vs. Machine readable
  - Raw data vs. "Cooked" data
    - Information is often available but hard to find/construct
    - Descriptions of data transformation (anonymization, sampling, filtering)
- Data Reuse - Does this fit in the GENI model?

Researcher questions (#3)

- Why can't I get the data?
  - Privacy
    - Policy needs to be well defined
    - Drives what data collected and architecture
  - Cost
    - If OTDR/DAG card needed at every pop it will get expensive
    - Data Retention
  - Disruptive - Breakable networks
  - Technology - probably the easiest

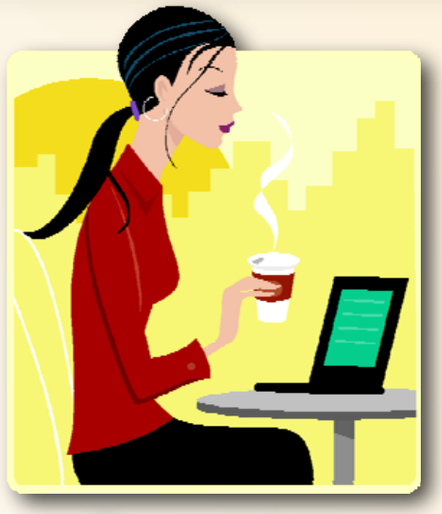
GENI Relevance

- Slightly different model than R&E networks
  - Separation of experiments
  - Is shared data useful in GENI?
    - Intermediate results
    - Repeatable of experiment
    - Manage your own experiment and data
    - “An Experimentation Workbench for Replayable Networking Research” [Eide, Stoller and Lepreau NSDI 07]

Data Management

- What possible things can be done to streamline research?
  - Metadata attached to measurement data
    - Operational Data - state of slice (and all slivers)
    - Data transformations
    - Slice Provisioning as a source of data
    - Experiment description language
    - Policy info??
    - Formats??? (RSpec, OGF schemas, Workflow Languages)

GENI Slice Creation

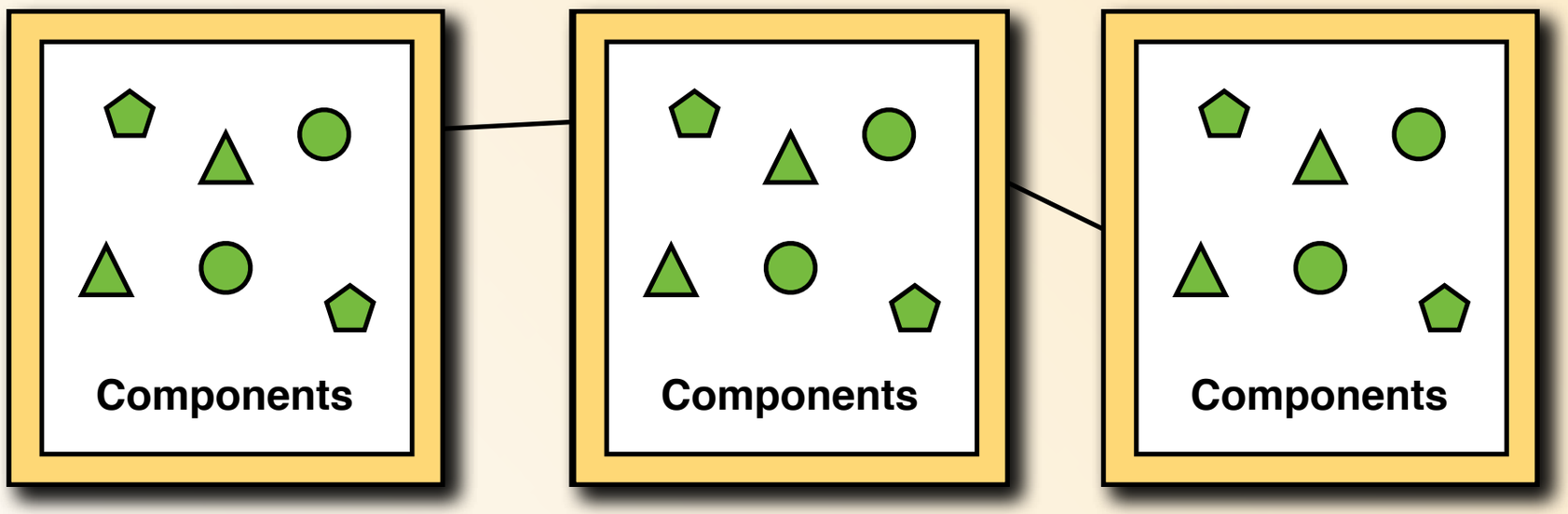


Researcher

GENI Experimental Workflow Tools

GIMS (GENI Instrumentation and Measurement Systems)

GMOC Operation DataSet

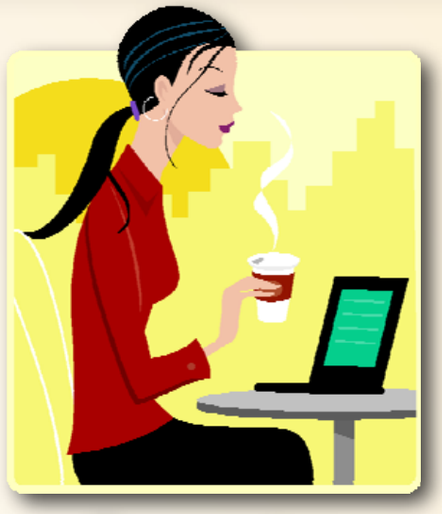


Aggregate A
Compute Nodes

Aggregate B
Backbone Net

Aggregate C
Mobile Net

GENI Slice Creation

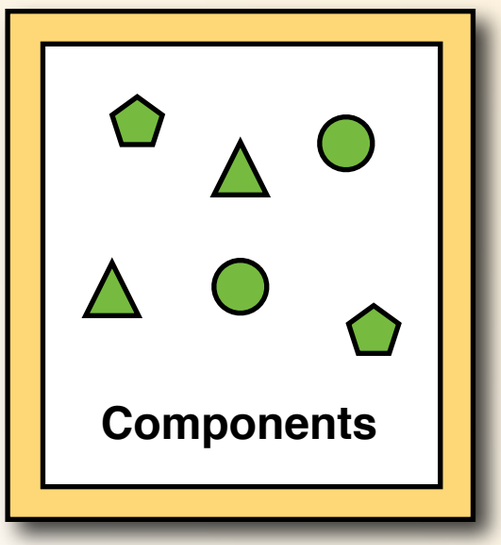


Researcher

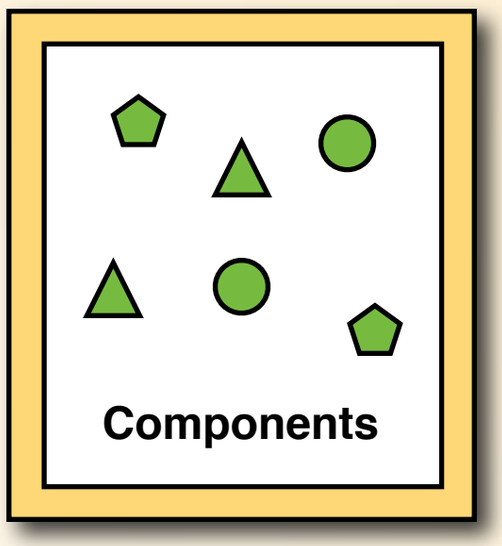
GENI Experimental Workflow Tools

GIMS (GENI Instrumentation and Measurement Systems)

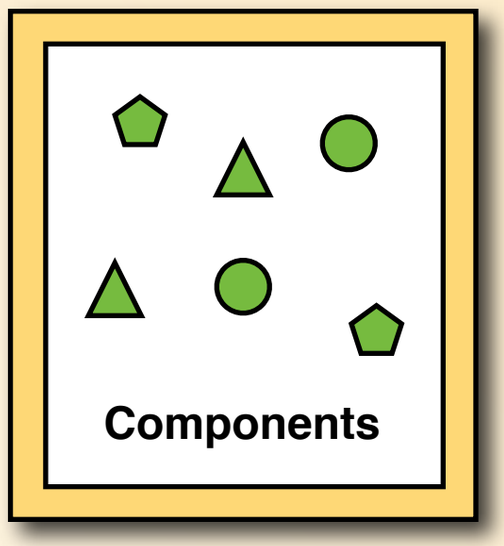
GMOC Operation DataSet



Aggregate A
Compute Nodes

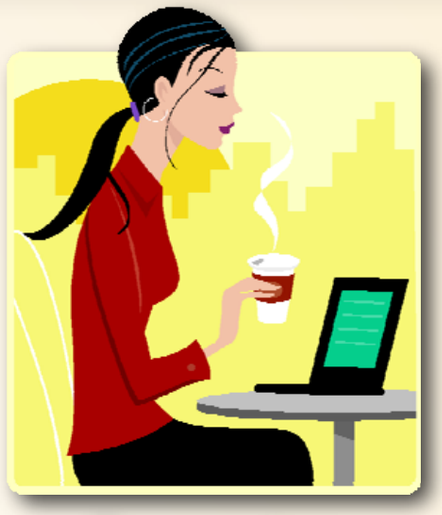


Aggregate B
Backbone Net



Aggregate C
Mobile Net

GENI Slice Creation



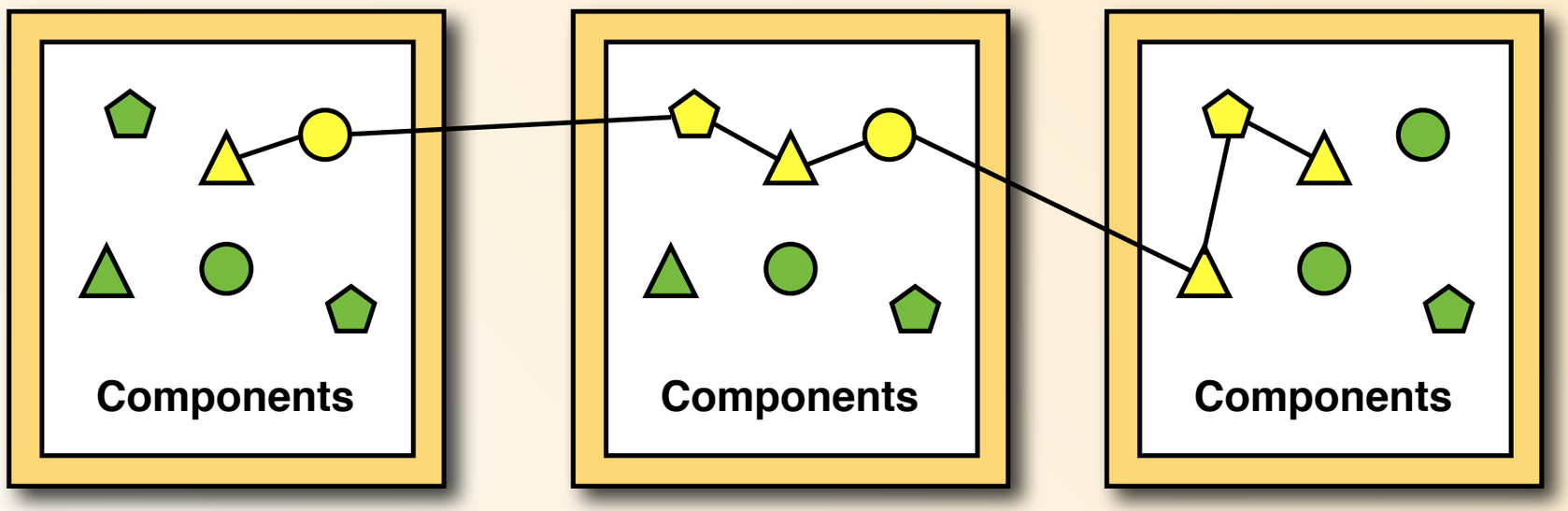
Researcher

GENI Experimental Workflow Tools

GIMS (GENI Instrumentation and Measurement Systems)

GENI:
"Slice Allocated"

GMOC Operation DataSet

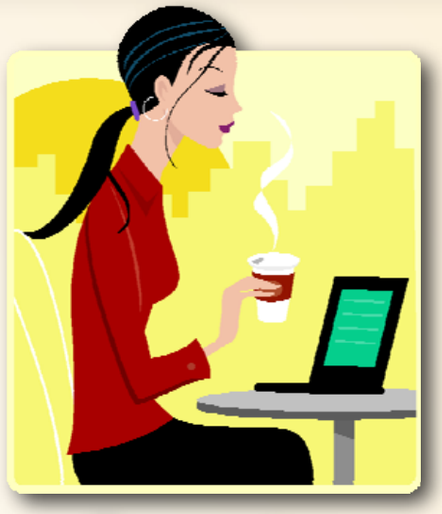


Aggregate A
Compute Nodes

Aggregate B
Backbone Net

Aggregate C
Mobile Net

GENI Slice Creation



Researcher

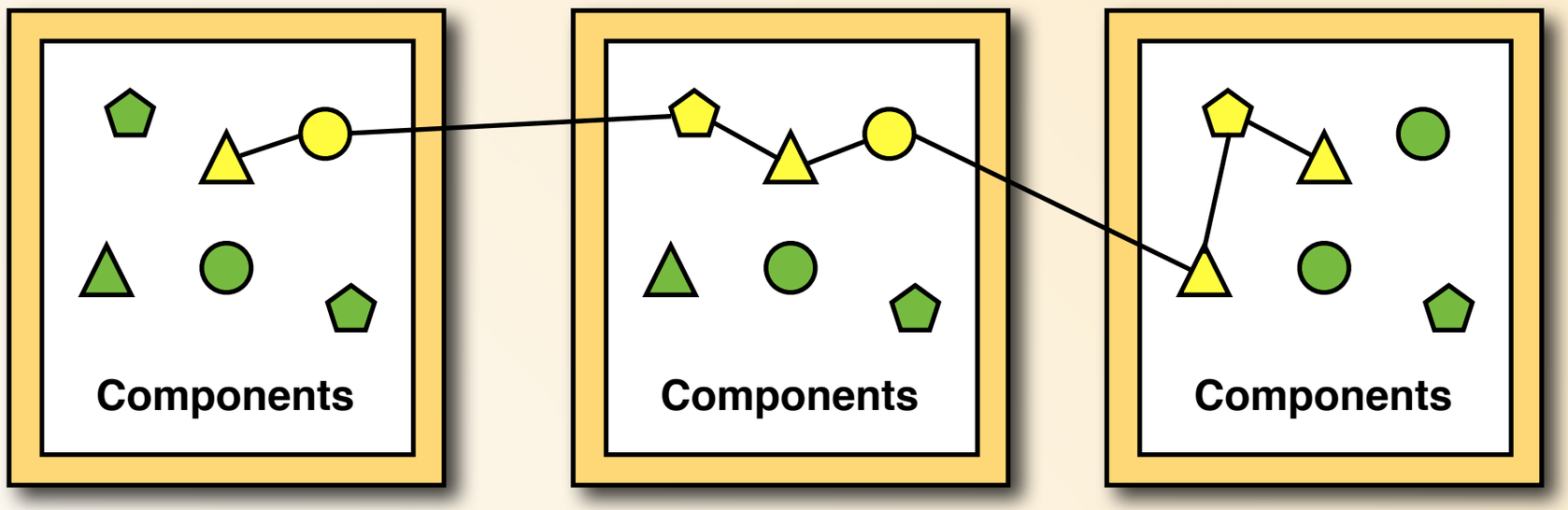
Researcher:
"Start Experiment"

GENI Experimental Workflow Tools

GIMS (GENI Instrumentation and Measurement Systems)

GMOC Operation DataSet

Experiment Configuration and Software pushed to slice

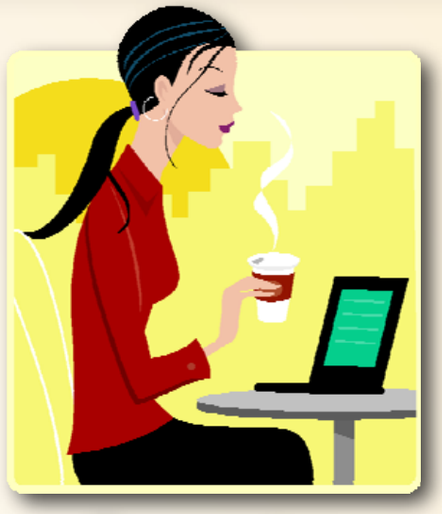


Aggregate A
Compute Nodes

Aggregate B
Backbone Net

Aggregate C
Mobile Net

GENI Slice Creation



Researcher

Researcher: "Start Experiment"

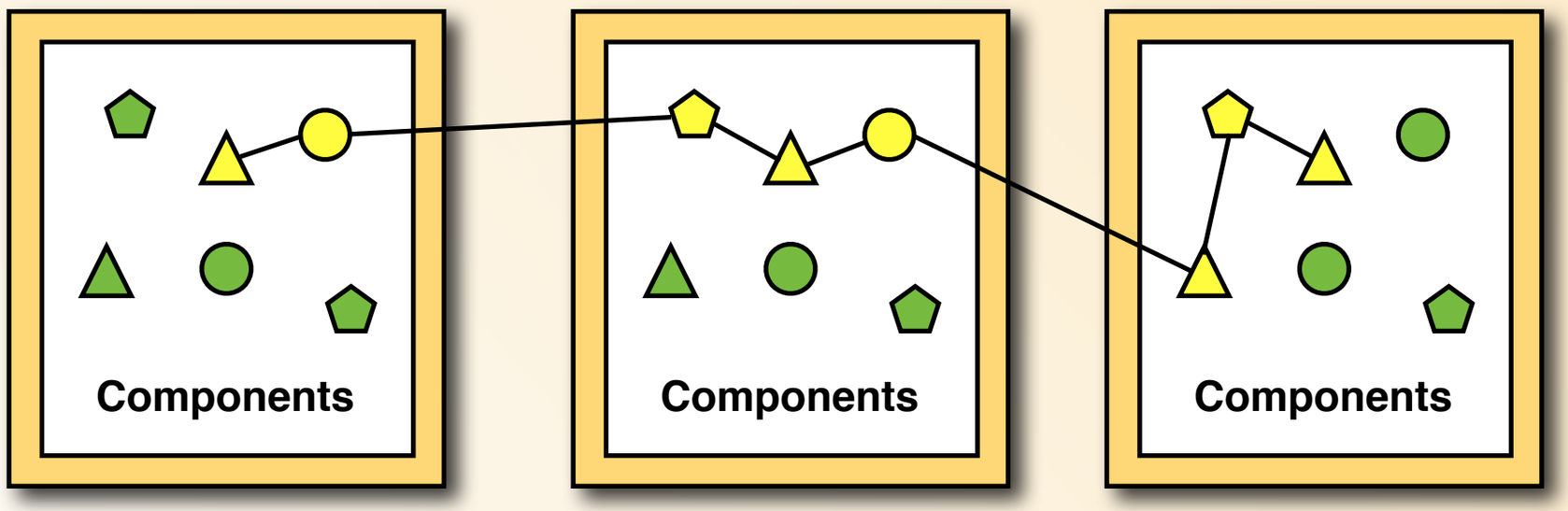
Can this workflow be captured and tied to the data collected in the experiment?

GENI Experimental Workflow Tools

GIMS (GENI Instrumentation and Measurement Systems)

GMOC Operation DataSet

Experiment Configuration and Software pushed to slice



Aggregate A
Compute Nodes

Aggregate B
Backbone Net

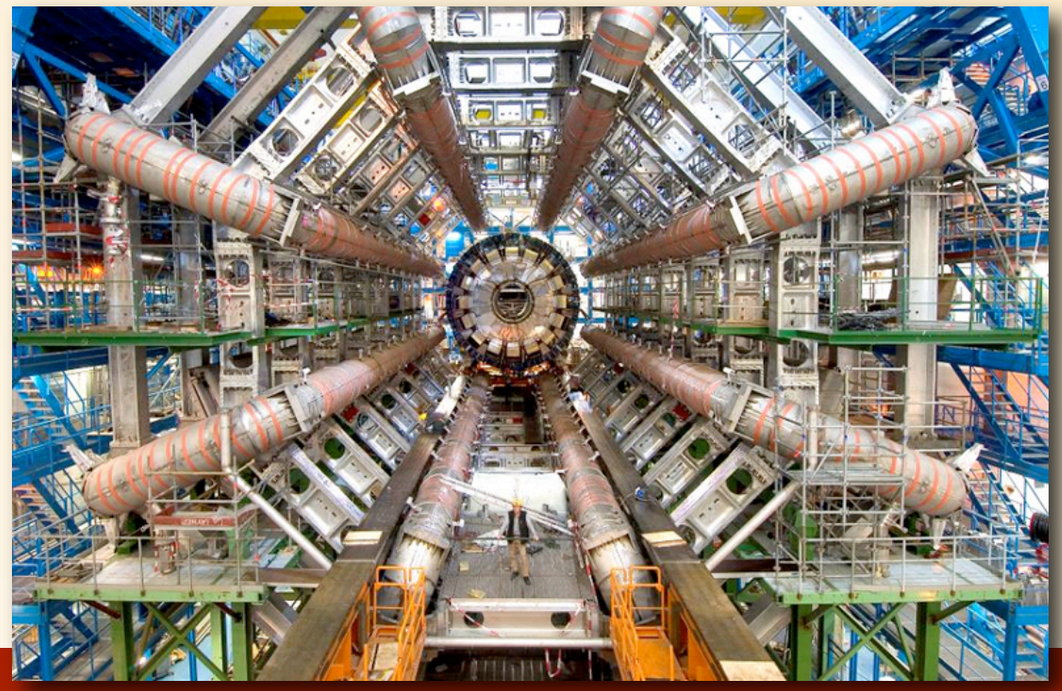
Aggregate C
Mobile Net

Data Management - Finding Data

- Data Discovery mechanisms
  - Portals - One or many
  - Web Services
  - Handle service
  - Find data “in the slice”
  - Storage - Where? Does it exist after the slice is released? Can it be used as an incentive to make data available to others.

Possible relevant work

- Grid
  - Data provenance - Karma, Workflow languages (BPEL, YAML)
    - Use Workflow Language to describe an experiment and infrastructure state
  - OGF schemas - PerfSonar
- Other Sciences (HEP, Astronomy, Meteriology)
- Emulab/Flexlab -- replayable experiments
- DATCAT
- CRAWDAD



Links

- IU GlobalNOC - http://globalnoc.iu.edu/
- GMOC - http://gmoc.grnoc.iu.edu/
- NSDI07 Replayable Research - http://www.cs.utah.edu/flux/papers/workbench-nsdi07-base.html
- Karma (Data Provenance) - http://www.dataandsearch.org/provenance/?q=taxonomy/term/3
- DatCat - http://www.datcat.org/
- Workflow Languages - http://www.gridworkflow.org/snips/gridworkflow/space/Workflow+Description+Languages