Globus Toolkit & Grids for Synchrotrons

Advanced Photon Source Experiment Requirements + others => Grids

Gregor von Laszewski

Rochester Institute of Technology

laszewski@gmail.com

Service Oriented Cyberinfrastructure Lab

- Previous Tenured at ANL: 2-3 people
  - Possibly the 4th Grid person
  - What comes next?
  - I did research not production …
- Small focused group
  - 1 Director
  - 1 Assoc. Director (soon)
  - 2 PhD (soon 3, come apply ;-))
  - 5 Masters Thesis
  - 15 students in Graduate Projects
  - 5 undergraduate student
- SOA, Clouds, Grid, Commodity Web 2.0 Grids, GPU (speedup of 70 is disruptive ;-))

R·I·T
Rochester Institute of Technology

Outline

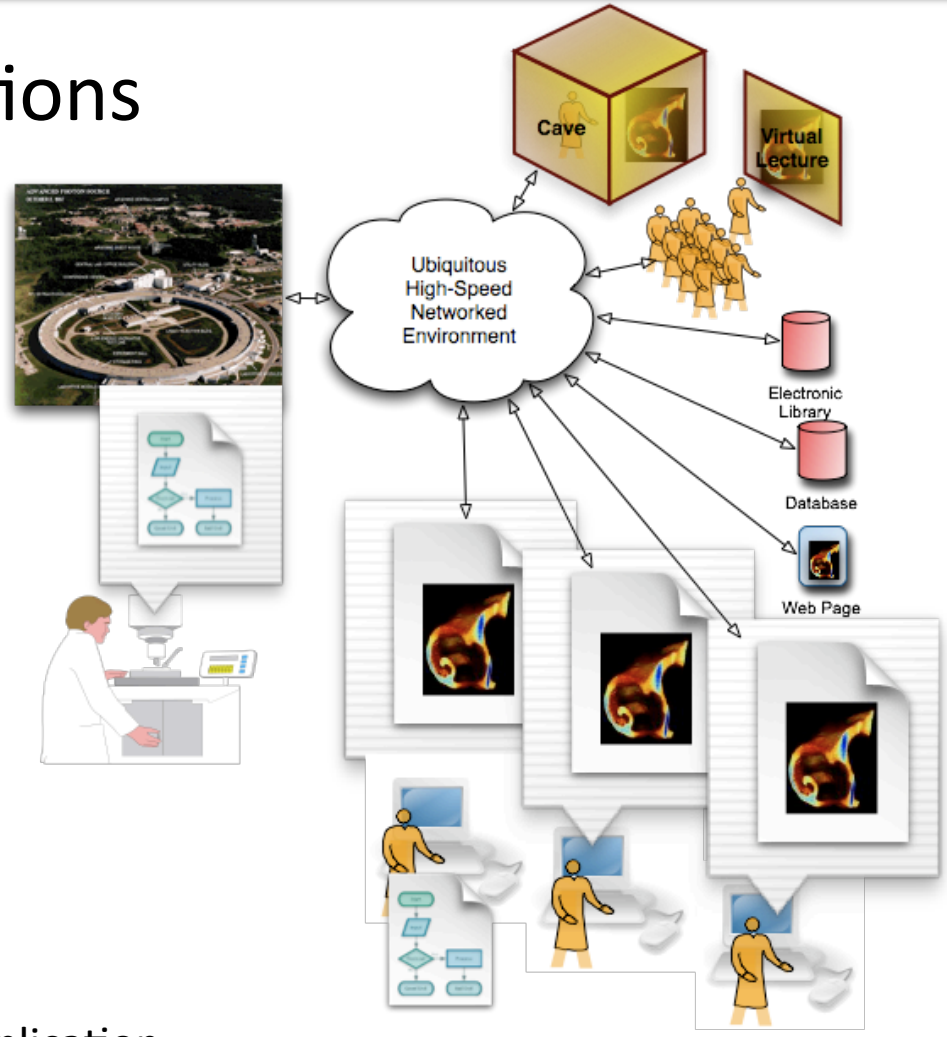
- Globus My View
- GT 4
- CoG Kits
- Cyberaide

- Administration Bottlenecks

Grid: My View

1996/7

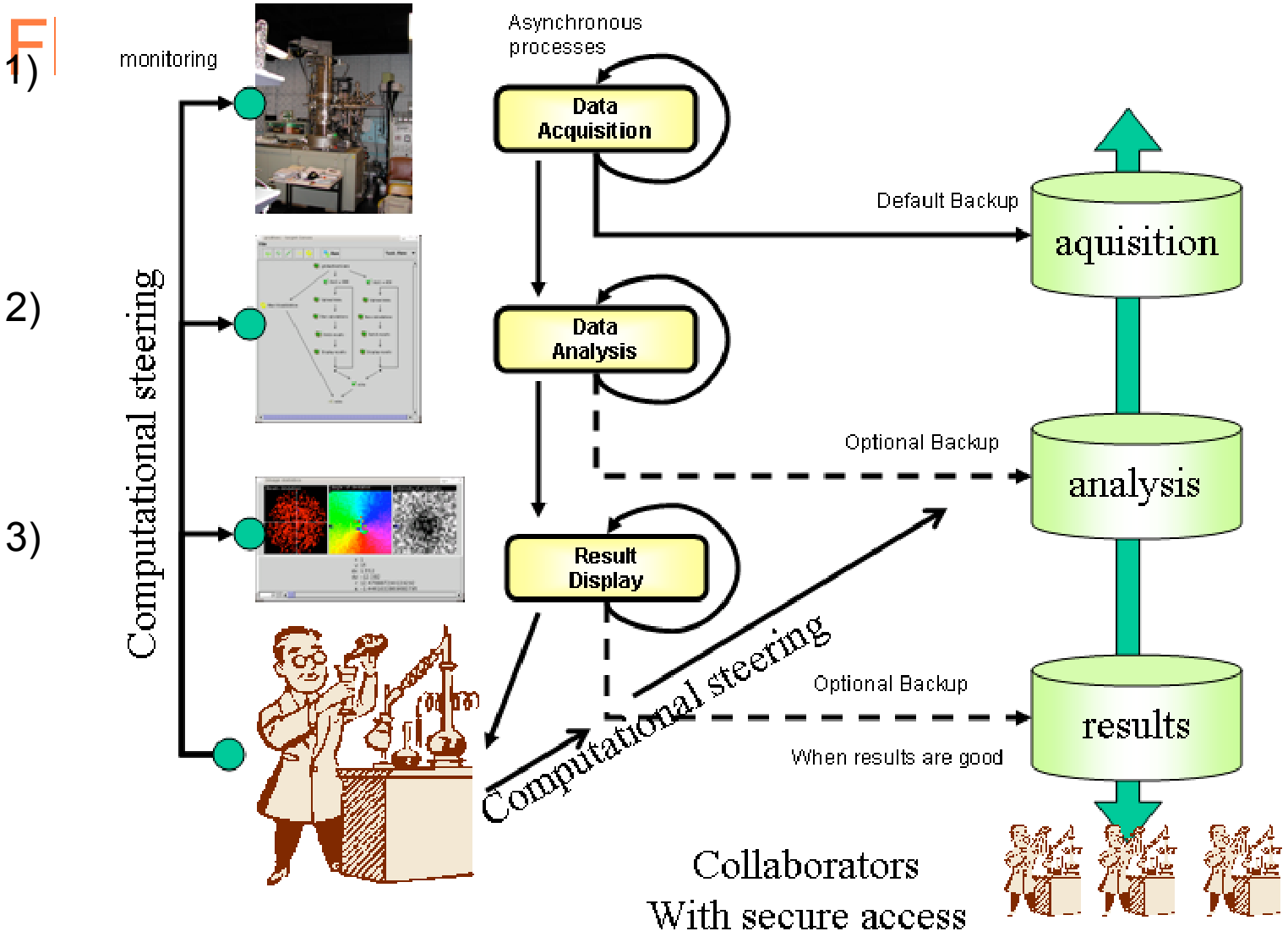
- Do real time calculations
- Queues are full
- Computers are remote
- People should not travel
- Role based Security model with ACL



⇒Most requirements came form this application.
Used in DOE slides for years.

R·I·T
Rochester Institute of Technology

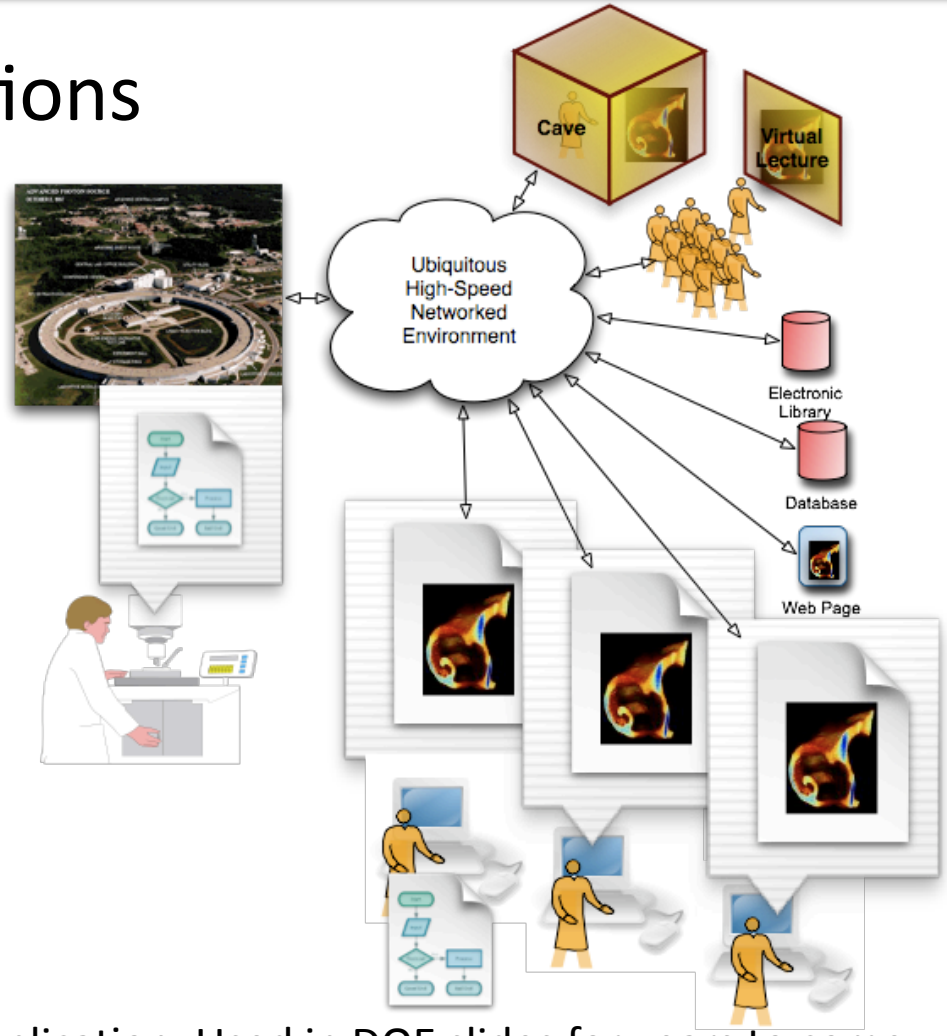
Nanomaterials/Electron Microscope



Realtime Data Analysis

1996/7

- Do real time calculations
- Queues are full
- Computers are remote
- People should not travel
- Role based Security model with ACL



=> Most requirements came form this application. Used in DOE slides for years to come.

Grids for Massive Data Analysis

Did not exist in 1996, 4 years later?

- Goal use (Grids and) remote Supercomputers for massive Crystallographic and Microtomographic Structural Problems and address
  - **Infrastructure:** develop new capabilities for the reconstruction of high-resolution tomographic data (including time resolved data) obtained from advanced X-ray sources such as the Advanced Photon Source (APS)
  - **Science:** develop new capabilities for the direct phase solution of atomic resolution, large-molecule crystallographic data obtained from APS detectors
  - **Modality:** demonstrate the tight coupling of Teraflop/sec-class supercomputers with APS detectors, to enable quasi-real-time analysis and visualization, and hence **interactive steering of experiments**.

Requirements

- Many users
- Do my calculation quickly
- Give me my data
- Security, what are you talking about?
- Where is the GUI?
- Where is the script/ commandline?

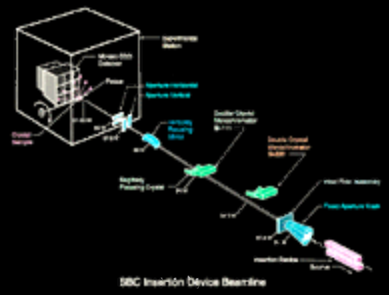
- Which Language
  - C/C++
  - FORTAN
  - Python (new)

- I do not want to care about
  - Metacomputing
  - Grids
  - Clouds

1997

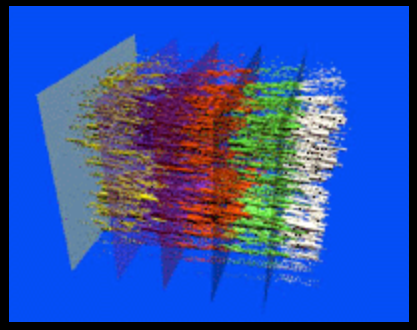
Argonne National Laboratory
Globus & Structural Biology

APS

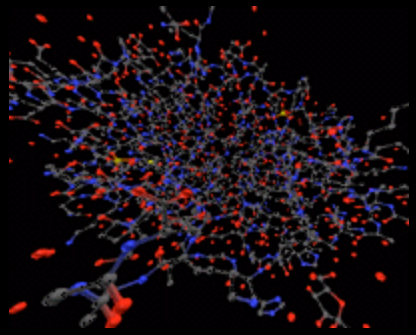


Experiment Setup

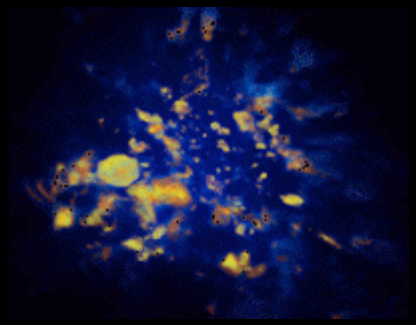
Globus Environment



Shoebox View



Molecule View



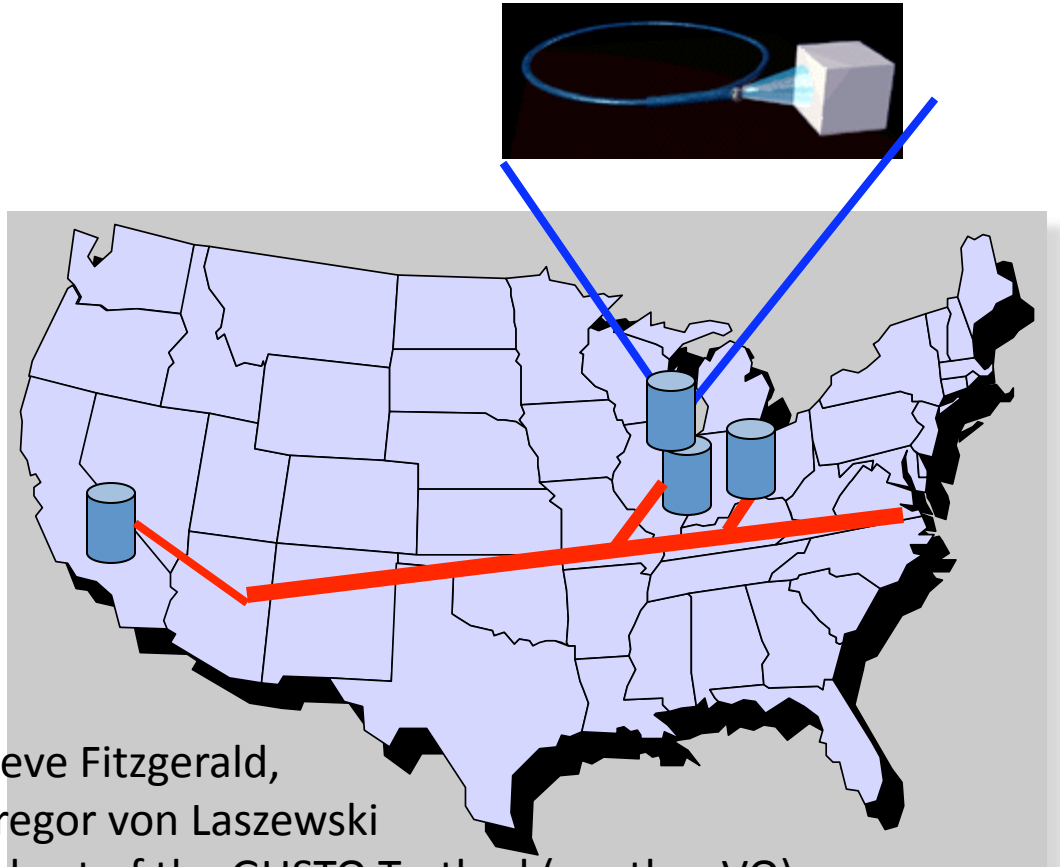
Density View

The first "VO": SC'97 APS - testbed

- (16) ANL, IL
- (8) ISI, CA
- (64) U Indiana
- (128) NCSA, IL

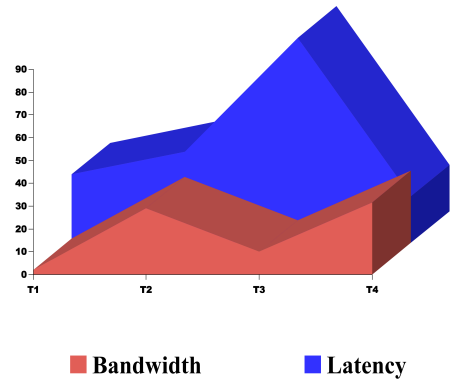
- Tomographic analysis

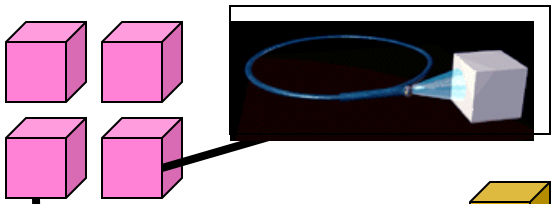
- One Grid (Foster) vs many Grids (Laszewski)

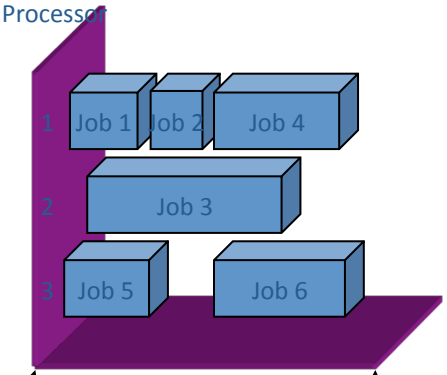


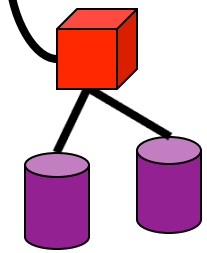
Steve Fitzgerald,
Gregor von Laszewski
Subset of the GUSTO Testbed (another VO)

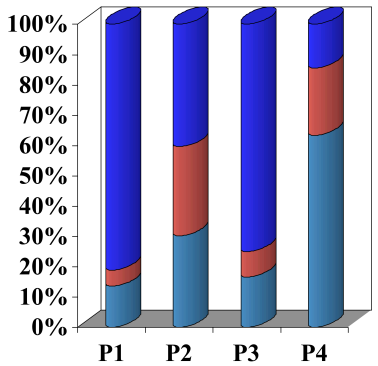
State of the Network



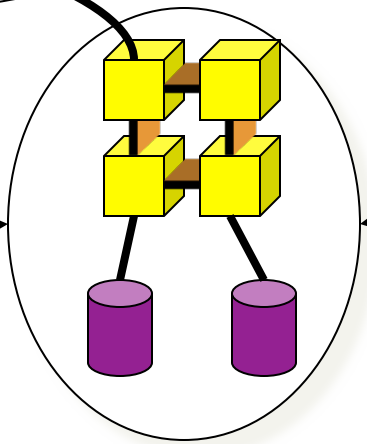








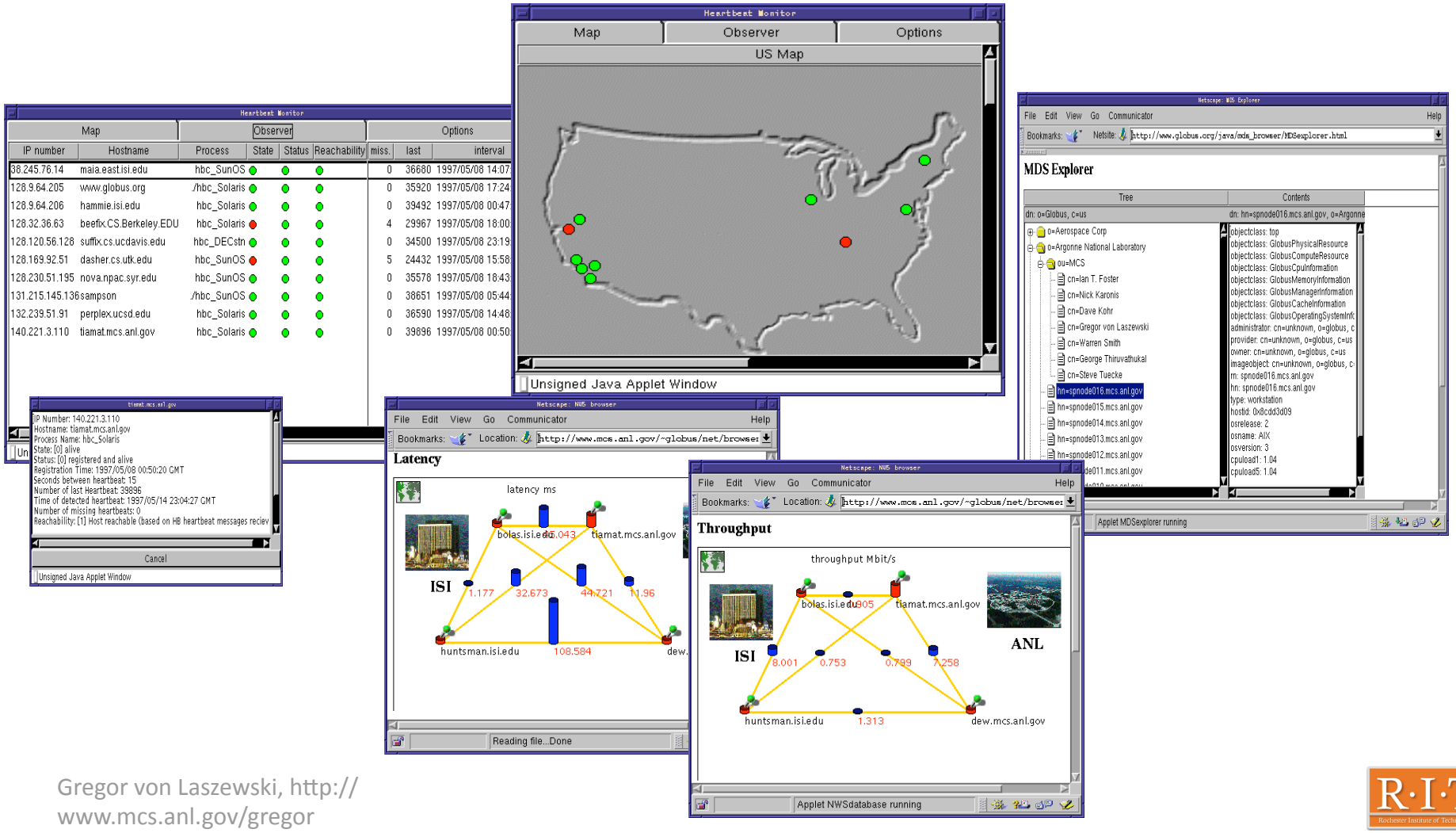
Free
System
User



Gregor von Laszewski, http://www.mcs.anl.gov/gregor

R·I·T
Rochester Institute of Technology

1997: von Laszewski
Desktop Access To Remote Resources

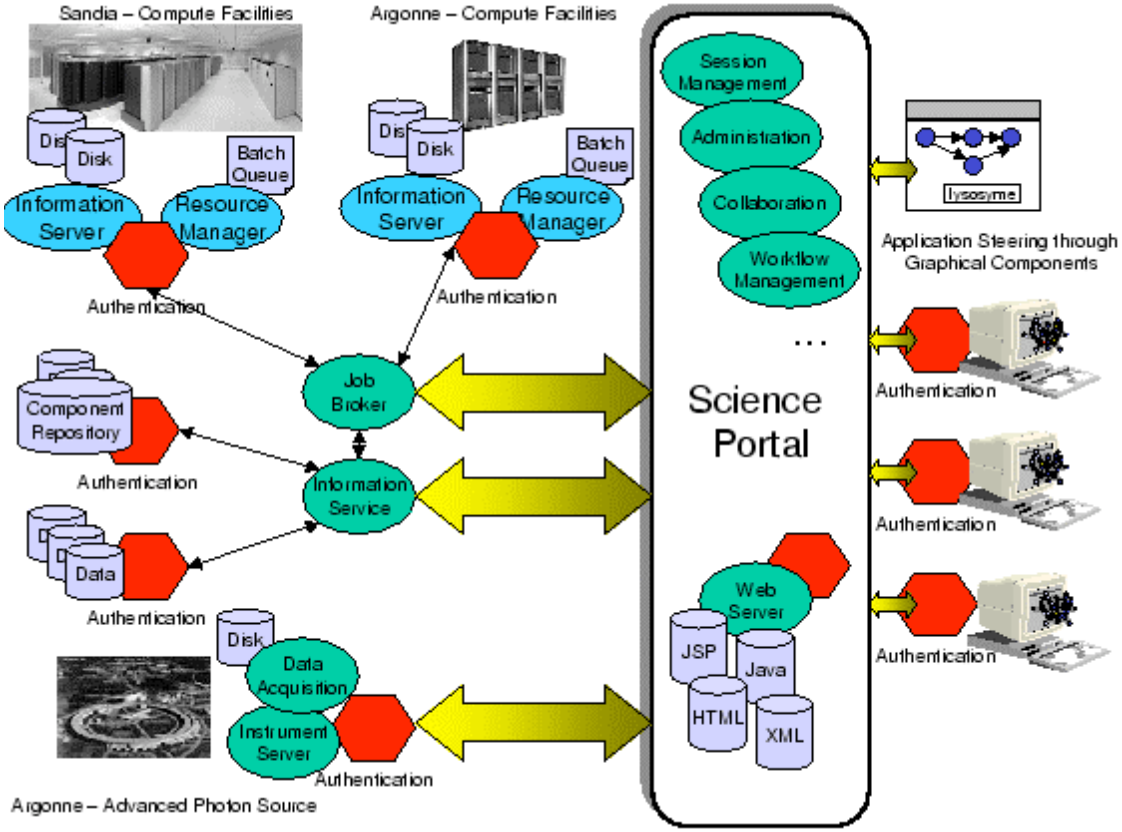


Gregor von Laszewski, http://www.mcs.anl.gov/gregor

R·I·T
Rochester Institute of Technology

The Portal is Important

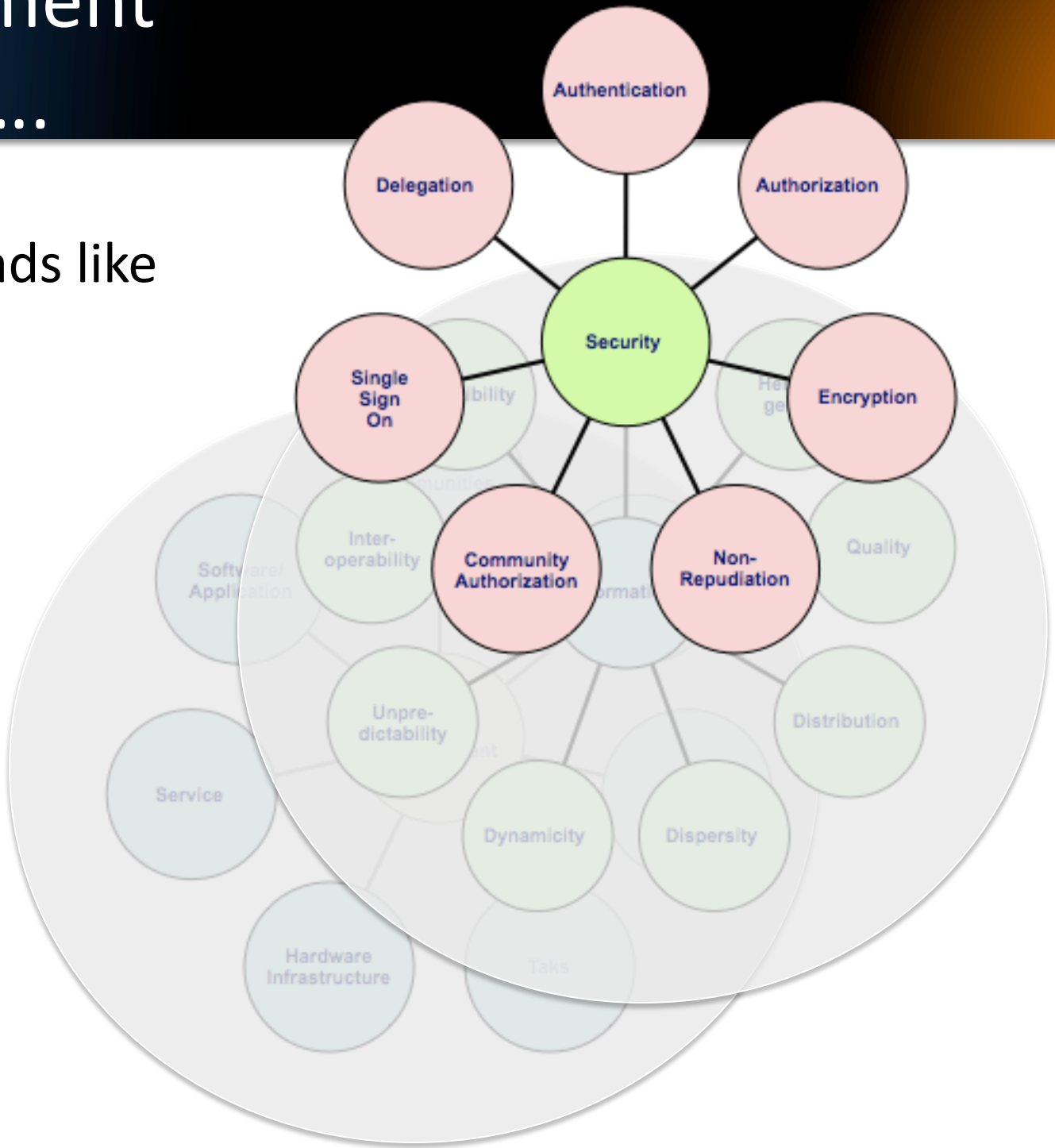
- ACM article



Management aspects ….

1997

- It spreads like a weed



Evolution of Globus

Cloud

CoG Kit workflow

Cyberaide

If you used the CoG Kit for on demand batch Processing and file transfer, little changes needed. CoG is above middleware.

Meta Computer

1995/6

Generation 0:

Message Passing

(Nexus, I-Way)

1998

Generation 1:

On Demand Batch Computing, Runtime Library, Protocols, API, message passing (Gusto Testbed/ Metacomputing)

GvL: everything must be a service

1999

Generation 2:

Addressing scalability

removal of message passing, focus on

Elementary functionality

Bottom up design

gridFTP

GridForum

GvL: top down design

2003

Generation 3:

Everything is a service

Global Grid Forum

Stateful Services

Java

Not the right way.

2006

Generation 4:

WSRF

Stateful Services

Scalability Improvements

TeraGrid

Community Driven

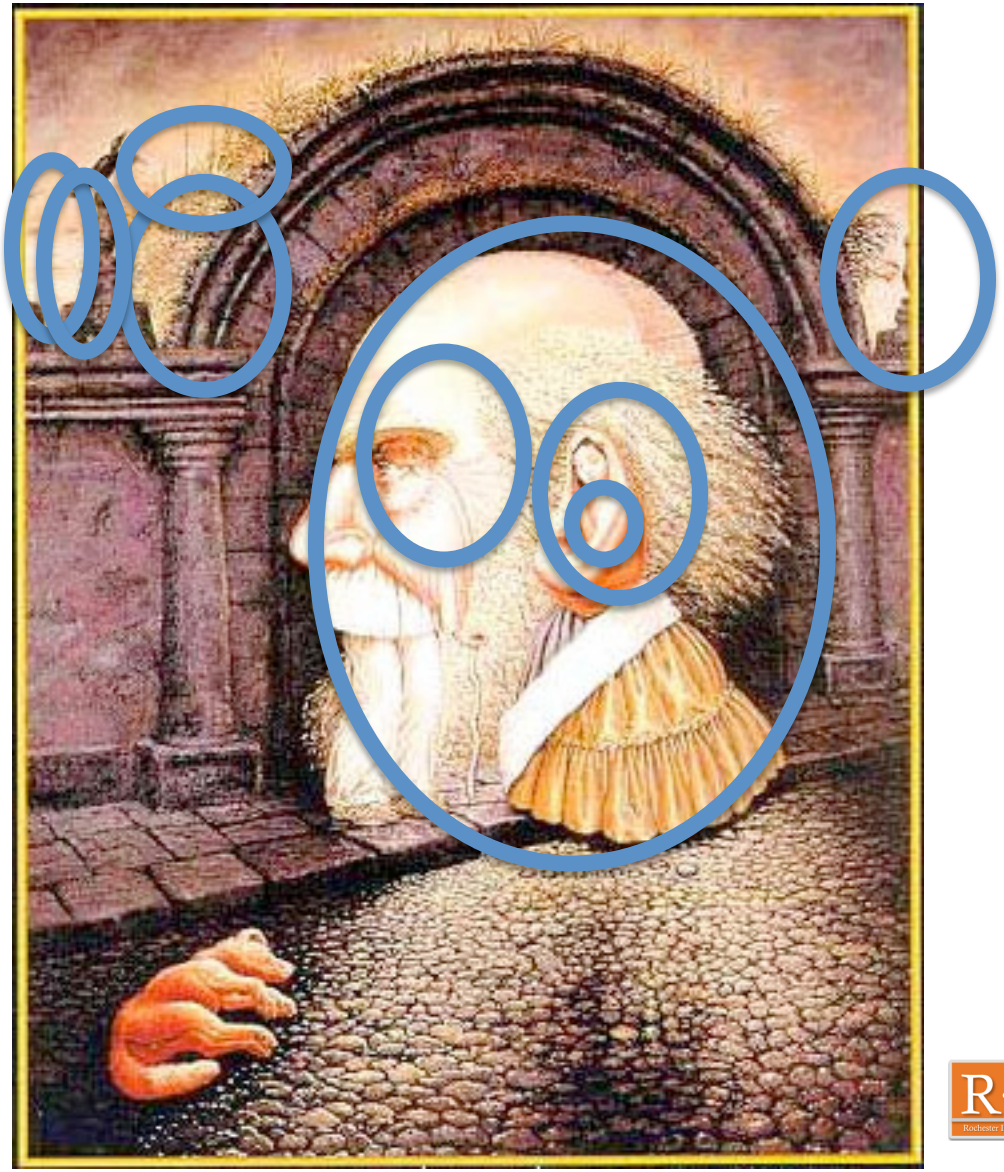
Open Grid Forum

Note: This is the impression Gvl Was with the GT Project from 1996-2007

Help me if the dates are wrong

The wife and the mother-in-law

1. old man profile
2. Old man standing
3. Lady
4. face on the right side of the pillar (profile)
5. face on the left side of the pillar (profile)
6. another gestalt shape, at the opposite to the left pillars face.
7. a front facing extreme left in the sky.
8. face near a crow.
9. baby with the lady.

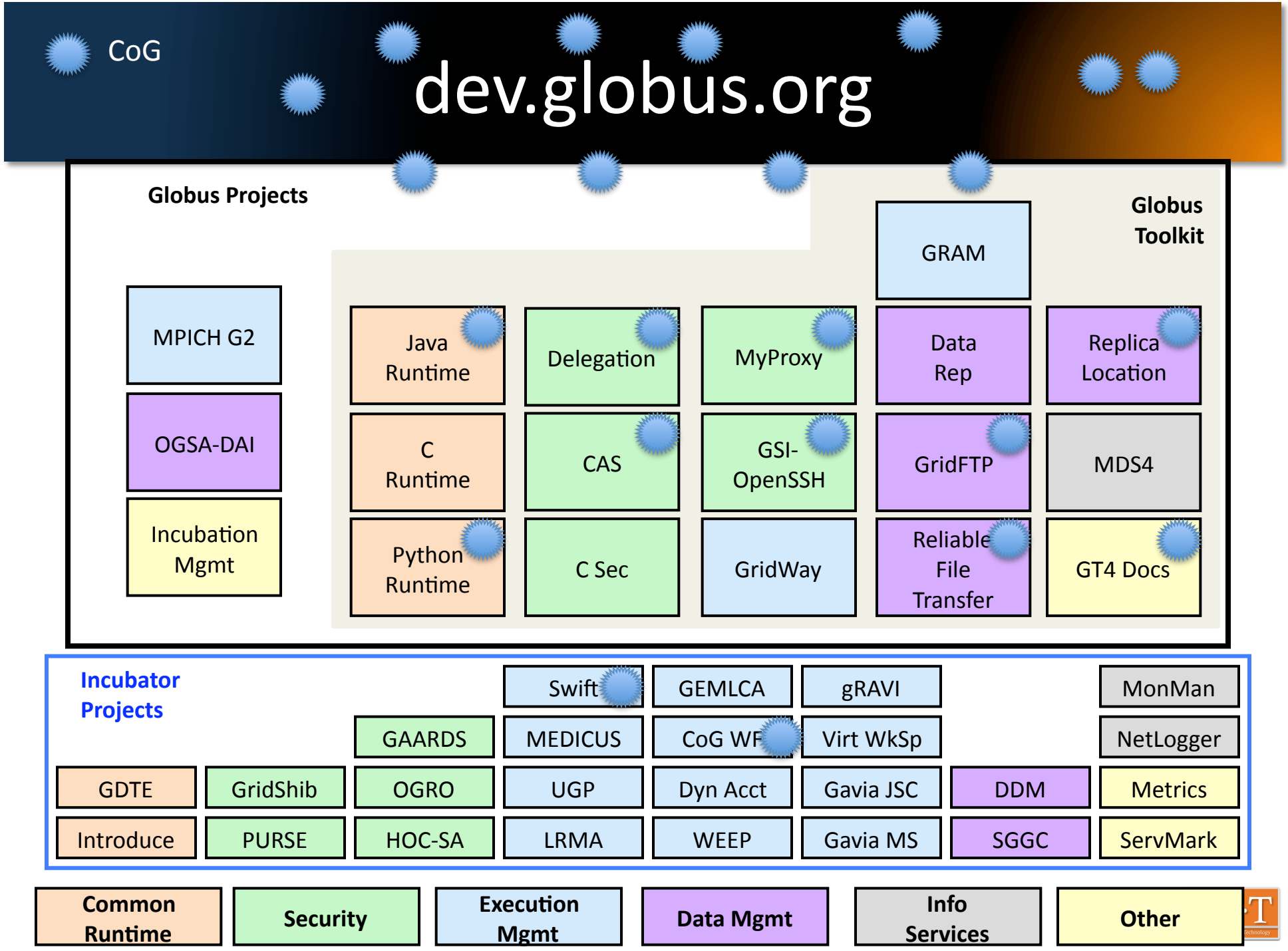


Fast Forward to Today

R·I·T
Rochester Institute of Technology

Today: Globus 4

- Contains pre-WS GT2, many claim they do GT4 but that may not mean they use web services
- Contains WSRF style web services
  - Improved job scalability
  - Improved file transfer
  - Many components added by the community
- Lots of useful components
- Lots of development of additional components


CoG
dev.globus.org
Globus Projects
Globus Toolkit
MPICH G2
OGSA-DAI
Incubation Mgmt
Java Runtime
C Runtime
Python Runtime
Delegation
CAS
C Sec
MyProxy
GSI-OpenSSH
GridWay
GRAM
Data Rep
GridFTP
Reliable File Transfer
Replica Location
MDS4
GT4 Docs
Incubator Projects
Swift
GEMLCA
gRAVI
MonMan
GAARDS
MEDICUS
CoG WF
Virt WkSp
NetLogger
GDTE
GridShib
OGRO
UGP
Dyn Acct
Gavia JSC
DDM
Metrics
Introduce
PURSE
HOC-SA
LRMA
WEEP
Gavia MS
SGGC
ServMark
Common Runtime
Security
Execution Mgmt
Data Mgmt
Info Services
Other

Job Submission

R·I·T
Rochester Institute of Technology

GRAM4 Scalability

- Job Submission
- Scalability a major focus of GRAM's design
  - GRAM4 can manage 32,000 active jobs
  - Ability to manage load on control node
  - GRAM4 can handle bursts of up to 50 job submissions
  - Each job requires ~2s to process
- Are the error conditions acceptable?
  - Job should be rejected or timeout before overloading the service container or service host

Job Submission Features

Globus has…

- Web service for job submission and control
- Cmd line and Java clients
- Data stage-in/stage-out
- Client notification support
- Plug-in interface for local resource managers
- Support for popular resource managers
- Plug-in interface for authorization decisions
- Advance reservation support (GARS)
- Standards compliance
- Supported on natl. grids

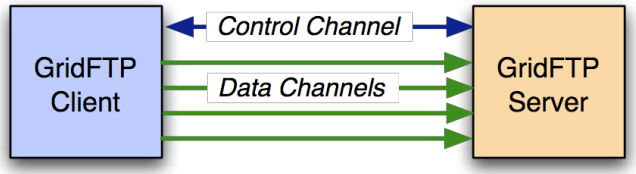
Globus doesn't have…

- Fancy submission tools (see Condor-G)
- Portlets (see CoG, OGCE)
- Scheduling/queuing (see GridWay, PBS, LSF…)
- Co-scheduling (see GARS, HARC, GUR)
- Support for every resource manager
- Support for complicated authorization decisions (see VOMS, CAS, GridShib)
- CoG Kit Karajan Workflow

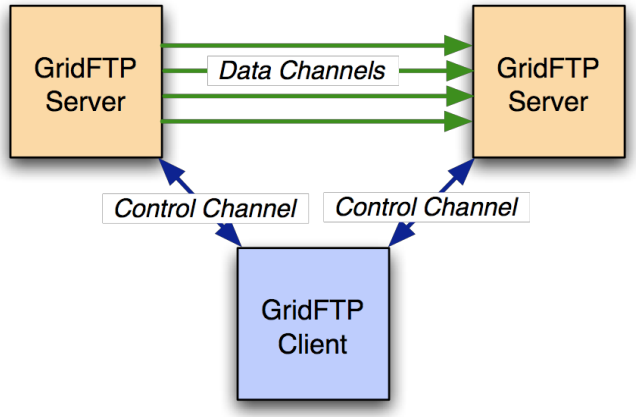
Filetransfer

GridFTP Features

- A high-performance, secure data transfer service optimized for high-bandwidth wide-area networks
  - FTP with extensions
  - Uses basic Grid security (control and data channels)
  - Multiple data channels for parallel transfers
  - Partial file transfers
  - Third-party (direct server-to-server) transfers

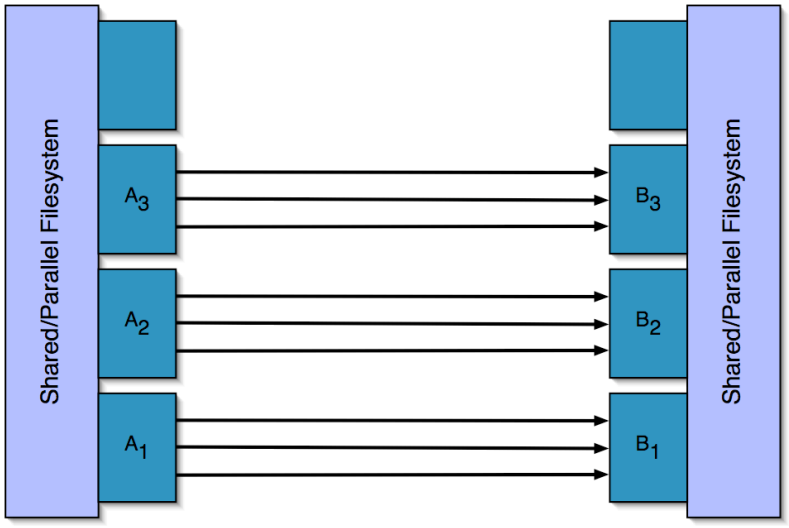


Basic Transfer
One control channel, several parallel data channels



Third-party Transfer
Control channels to each server, several parallel data channels between servers

Striped GridFTP



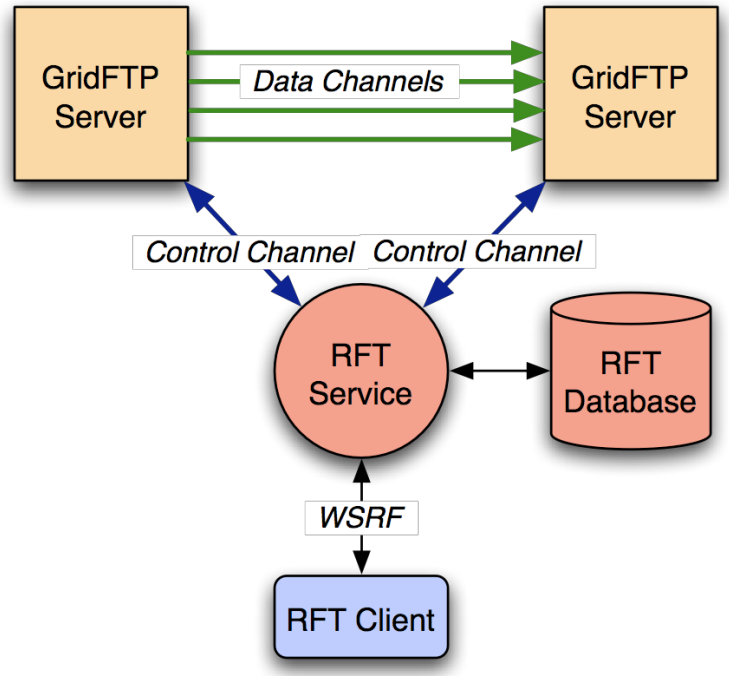
- GridFTP supports a striped (multi-node) configuration
  - Establish control channel with one node
  - Coordinate data channels on multiple nodes
  - Allows use of many NICs in a single transfer
- Requires shared/parallel filesystem on all nodes
  - On high-performance WANs, aggregate performance is limited by filesystem data rates

globus-url-copy

- Command-line client for GridFTP servers
  - Text interface
  - No “interactive shell” (single command per invocation)
- Many features
  - Grid security, including data channel(s)
  - HTTP, FTP, GridFTP
  - Server-to-server transfers
  - Subdirectory transfers and lists of transfers
  - Multiple parallel data channels
  - TCP tuning parameters
  - Retry parameters
  - Transfer status output

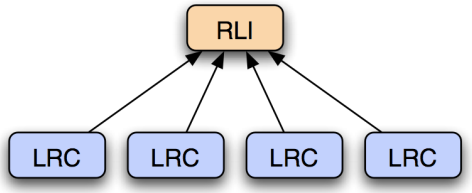
RFT – Reliabal File Transfer Service

- A WSRF service for queuing file transfer requests
  - Server-to-server transfers
  - Checkpointing for restarts
  - Database back-end for failovers
- Allows clients to request transfers and then "disappear"
  - No need to manage the transfer
  - Status monitoring available if desired

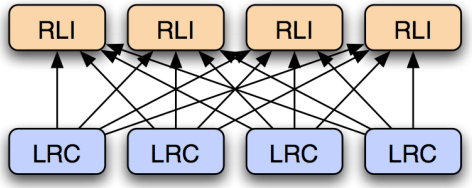


RLS - Replica Location Service

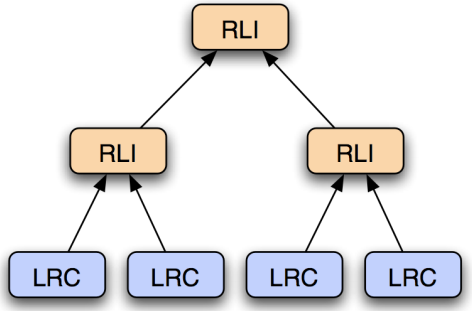
- A distributed system for tracking replicated data
  - Consistent local state maintained in Local Replica Catalogs (LRCs)
  - Collective state with relaxed consistency maintained in Replica Location Indices (RLIs)
- Performance features
  - Soft state maintenance of RLI state
  - Compression of state updates
  - Membership and partitioning information maintenance



Simple Hierarchy
The most basic deployment of RLS



Fully Connected
High availability of the data at all sites



Tiered Hierarchy
For very large systems and/or very Large collections

pyGlobus*

- High-level, object-oriented interface in Python to GT4 Pre-WS APIs.
  - GSI security
  - GridFTP
  - GRAM
  - XIO
  - GASS
  - MyProxy
  - RLS
- Also includes tools and services
  - GridFTP server
  - GridFTP GUI client
  - Other GT4 clients

* pyGlobus contributed to GT4 by the Distributed Computing Department of LBNL

Globus Toolkit Features

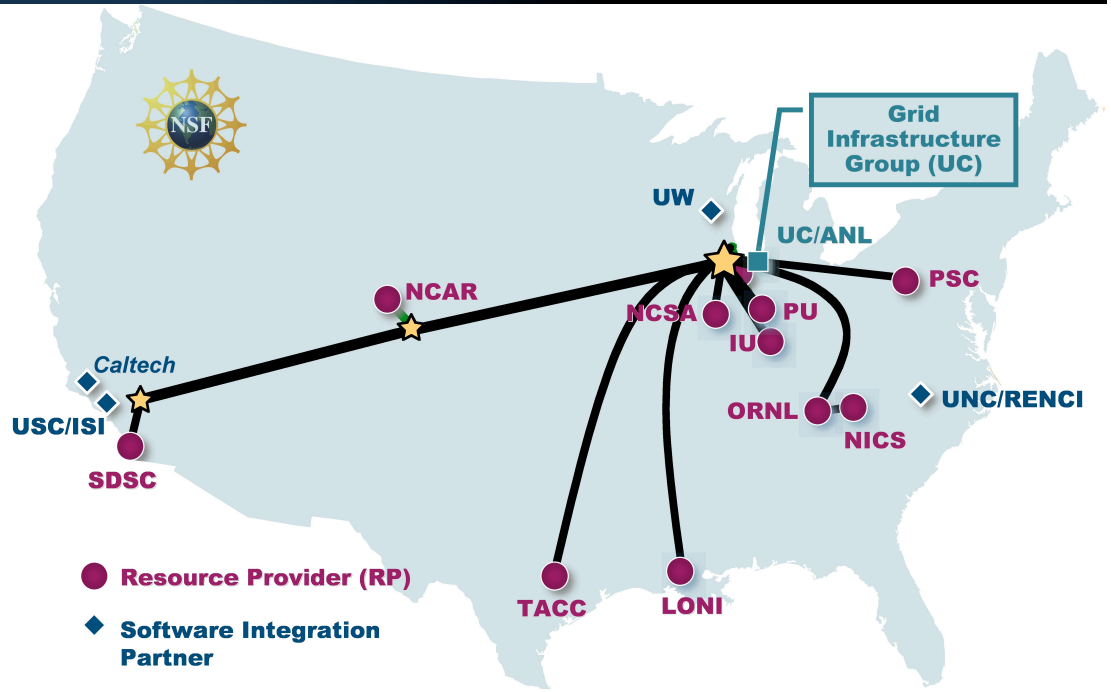
Globus has…

- Modular architecture
- APIs
- Embeddable libraries
- Web service interfaces
- Globus-enabled frameworks for MPI, RPC, parallel jobs, etc.
- Globus support on national infrastructure

Globus doesn't have…

- Your application already Grid-enabled
- A tool to automatically adapt your code
- Domain-specific frameworks

Deployment: NSF TeraGrid



- TeraGrid DEEP: Integrating NSF's most powerful computers (60+ TF)
  - 2+ PB Online Data Storage
  - National data visualization facilities
  - World's most powerful network (national footprint)
- TeraGrid WIDE Science Gateways: Engaging Scientific Communities
  - 90+ Community Data Collections
  - Growing set of community partnerships spanning the science community.
  - Leveraging NSF ITR, NIH, DOE and other science community projects.
  - Engaging peer Grid projects such as Open Science Grid in the U.S. as peer Grids in Europe and Asia-Pacific.
- Base TeraGrid Cyberinfrastructure: Persistent, Reliable, National
  - Coordinated distributed computing and information environment
  - Coherent User Outreach, Training, and Support
  - Common, open infrastructure services

A National Science Foundation Investment in Cyberinfrastructure

$100M 3-year construction (2001-2004)

$150M 5-year operation & enhancement (2005-2009)

* Slide courtesy of Ray Bair, Argonne National Laboratory

Security

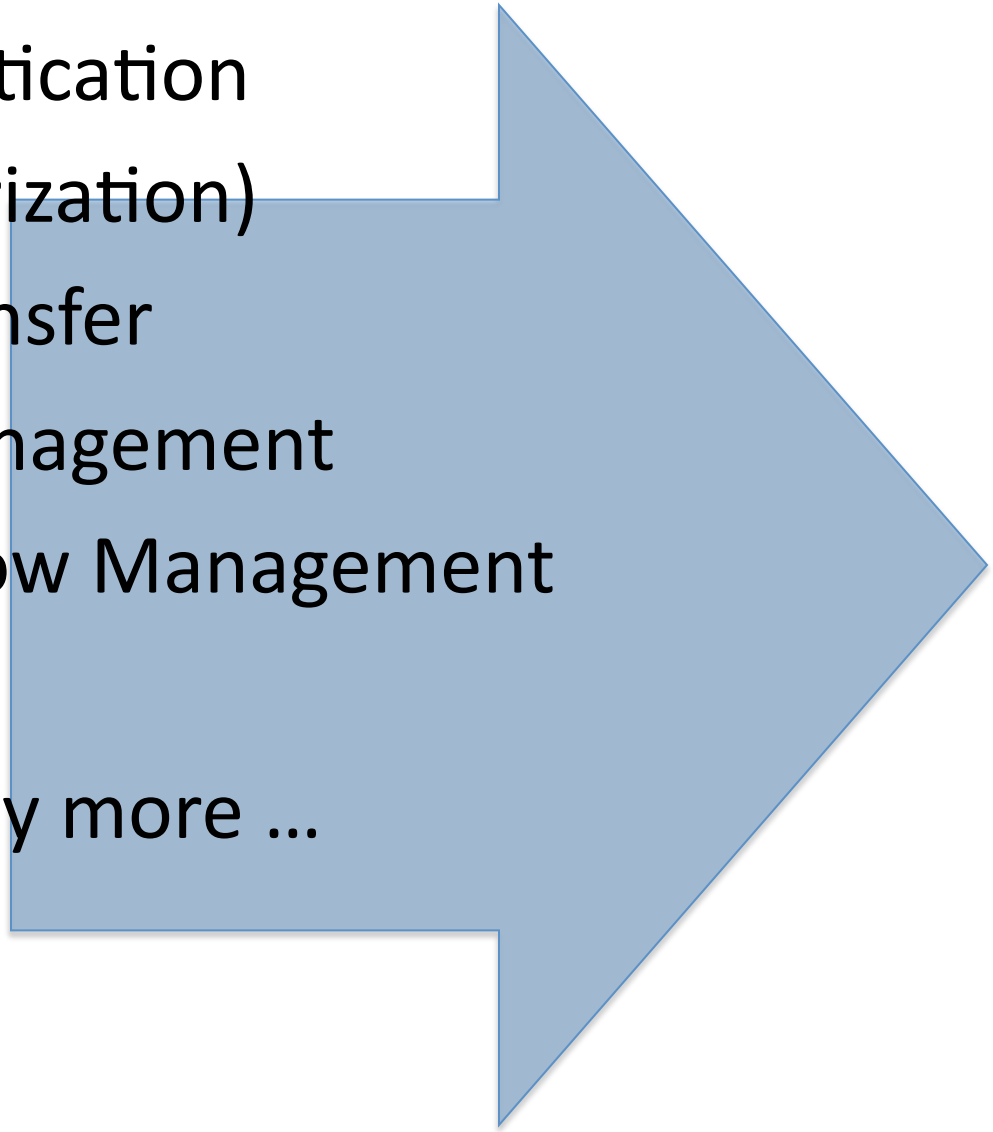
I am leaving this out as this will fill another hour, day, week, …

GSI authentication:
delegation, time restricted proxys

Which Services Do You need?

- Authentication
- (Authorization)
- File Transfer
- Job Management
- Workflow Management

- Probably more …



CoG Kit

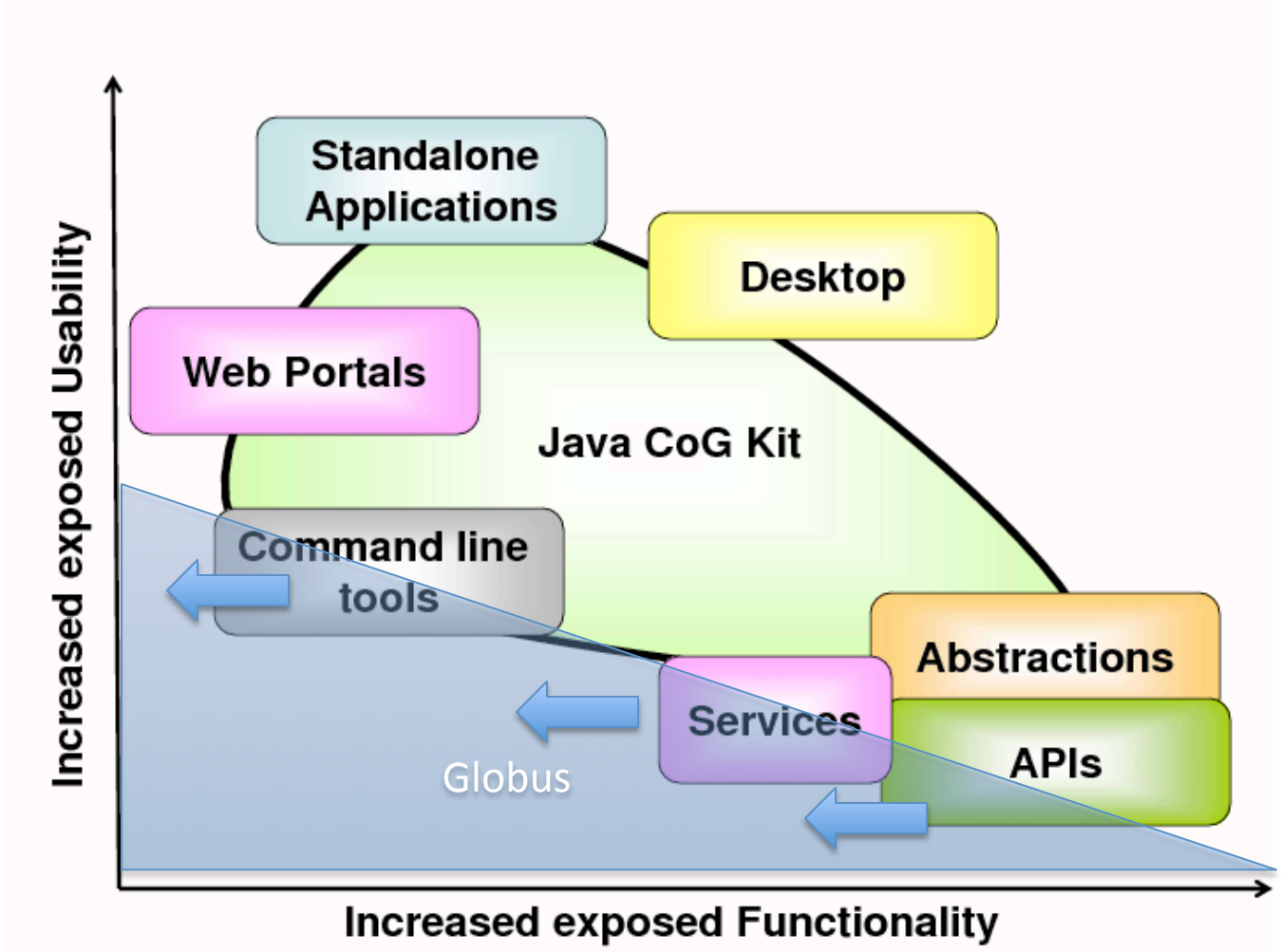
Condor

Globus

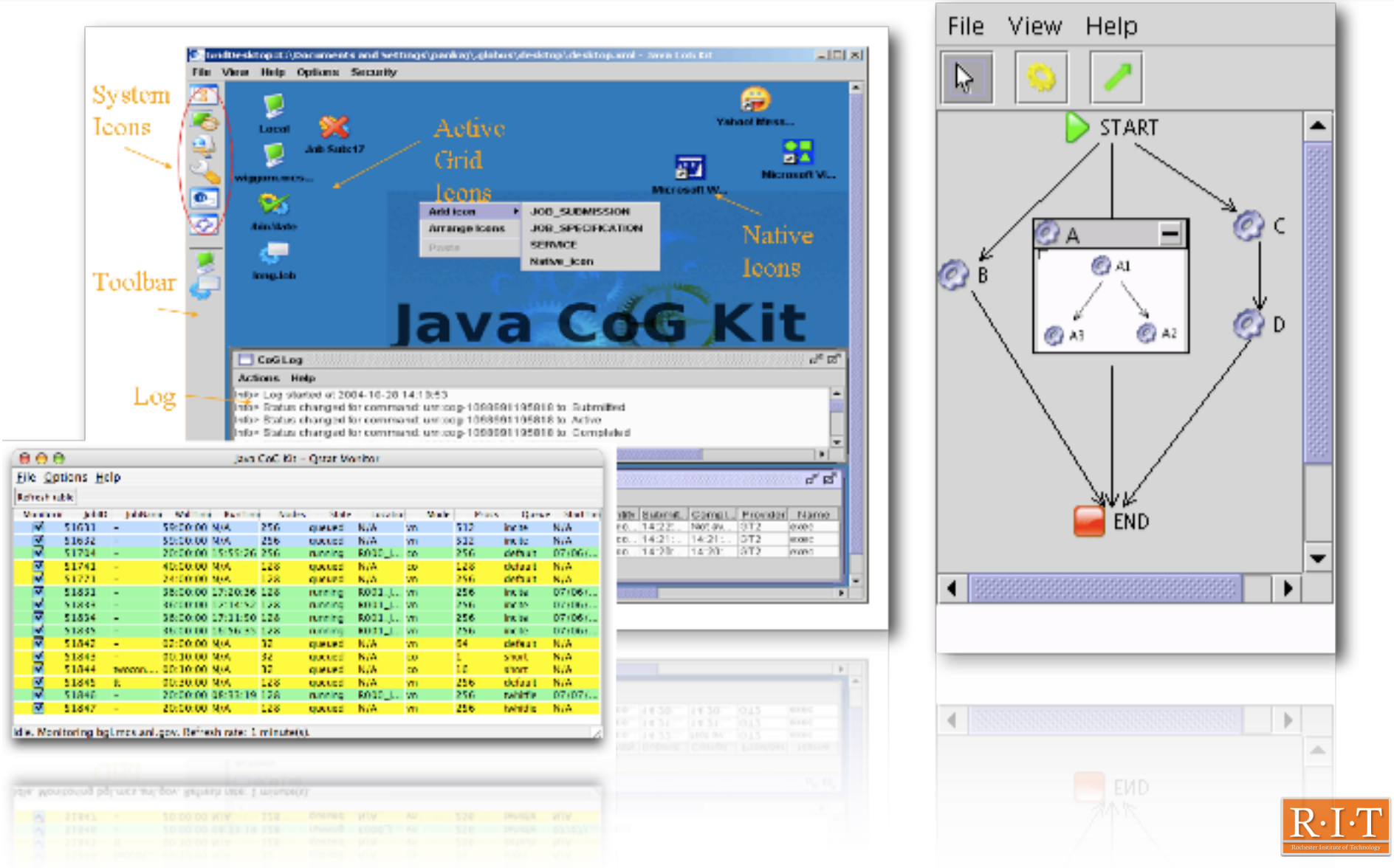
Others

?

CoG Kit


Standalone Applications
Desktop
Web Portals
Java CoG Kit
Command line tools
Abstractions
Services
APIs
Globus
Increased exposed Usability
Increased exposed Functionality

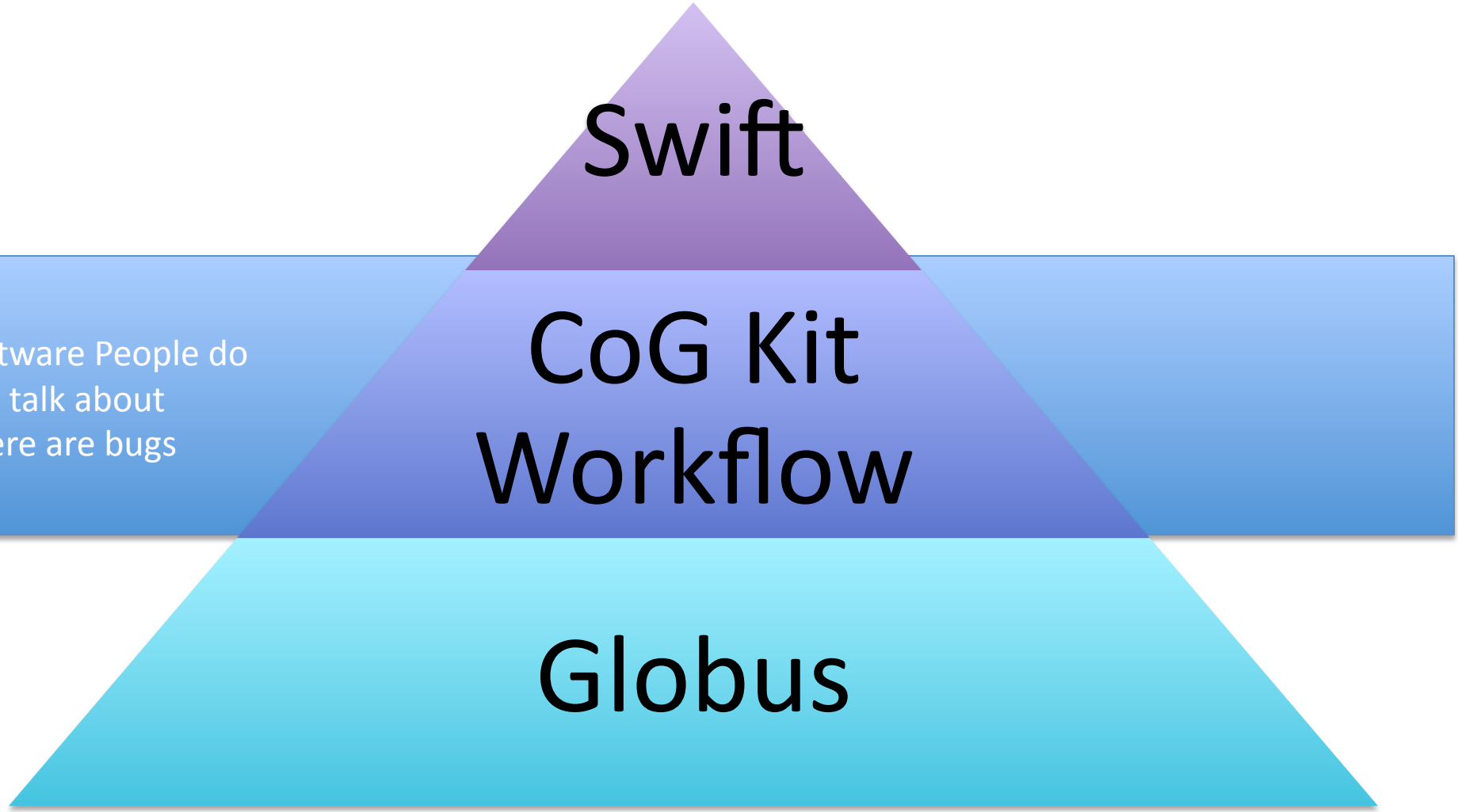
Visual Interfaces



CoG Karajan Workflow

```
<project>
  <include file="cogkit.xml"/>
  <execute executable="/bin/date"
           stdout="thedate"
           host="hot.mcs.anl.gov" provider="GT2"/>
  <echo message="Job completed. Transferring the output"/>
  <transfer srchost="hot.mcs.anl.gov" srcfile="thedate"
           desthost="localhost" provider="gridftp"/>
  <echo message="Transfer complete"/>
  <set name="date">
    <readFile file="thedate"/>
  </set>
```

Swift


Swift
CoG Kit
Workflow
Globus
Software People do
not talk about
There are bugs

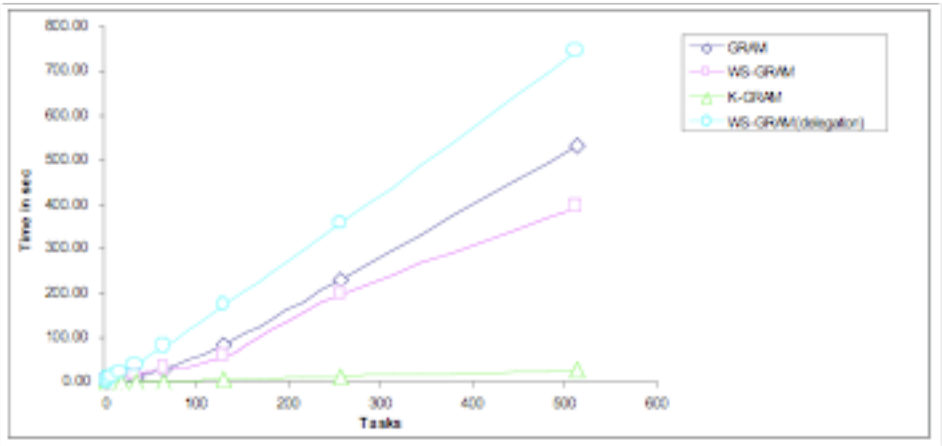
Task vs Job

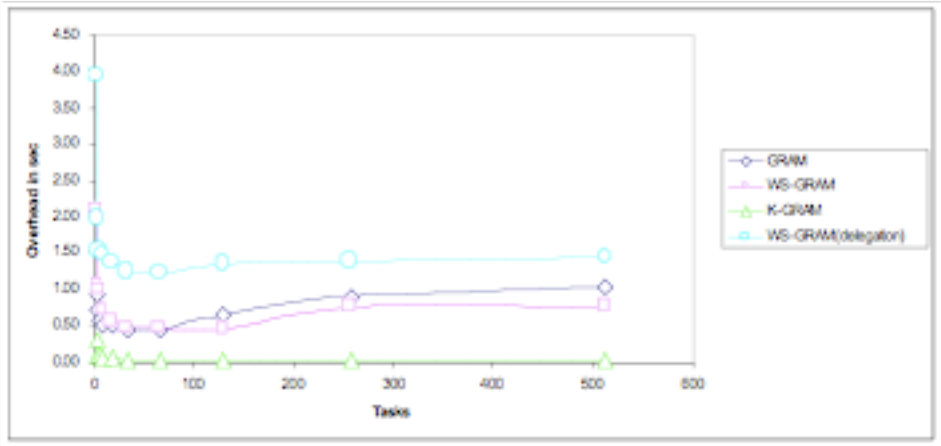
- We have developed frameworks doing millions of tasks before
- Because the Grid is available, some now try to use the Grid to map a task 1-to-1 to jobs
- Some call this Many-Task-Computing
  - This is not new
  - The basic construct of the CoG Kit is a “Task” not a job
- What is new:
  - Definition of a term
  - Wrong approach to Get Things Done
  - Use CoG Kits binning ability
  - CoG Kit Castor may replace Falcon

Job Binning

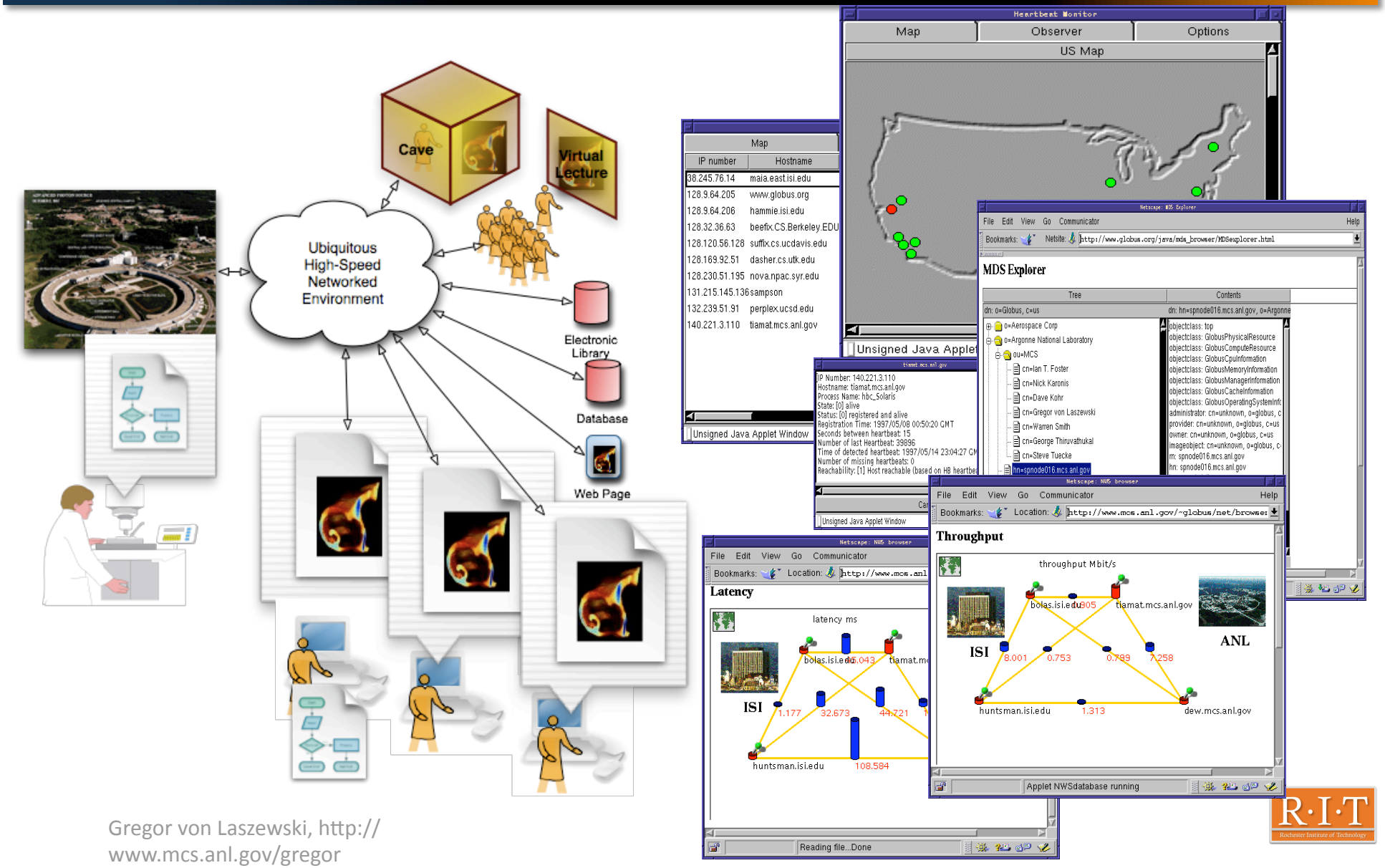
- Example: Executing many jobs in parallel, a typical case in parameter studies
- CoG is 30 times faster than WS-GRAM
- For 512 0.05 seconds instead of 1.46 per job

- GRAM
  - WS-GRAM performance in second range per job
  - CoG Execution service for multiple jobs: milliseconds per job (scales well through light weight threading)





Real Time Data Analysis

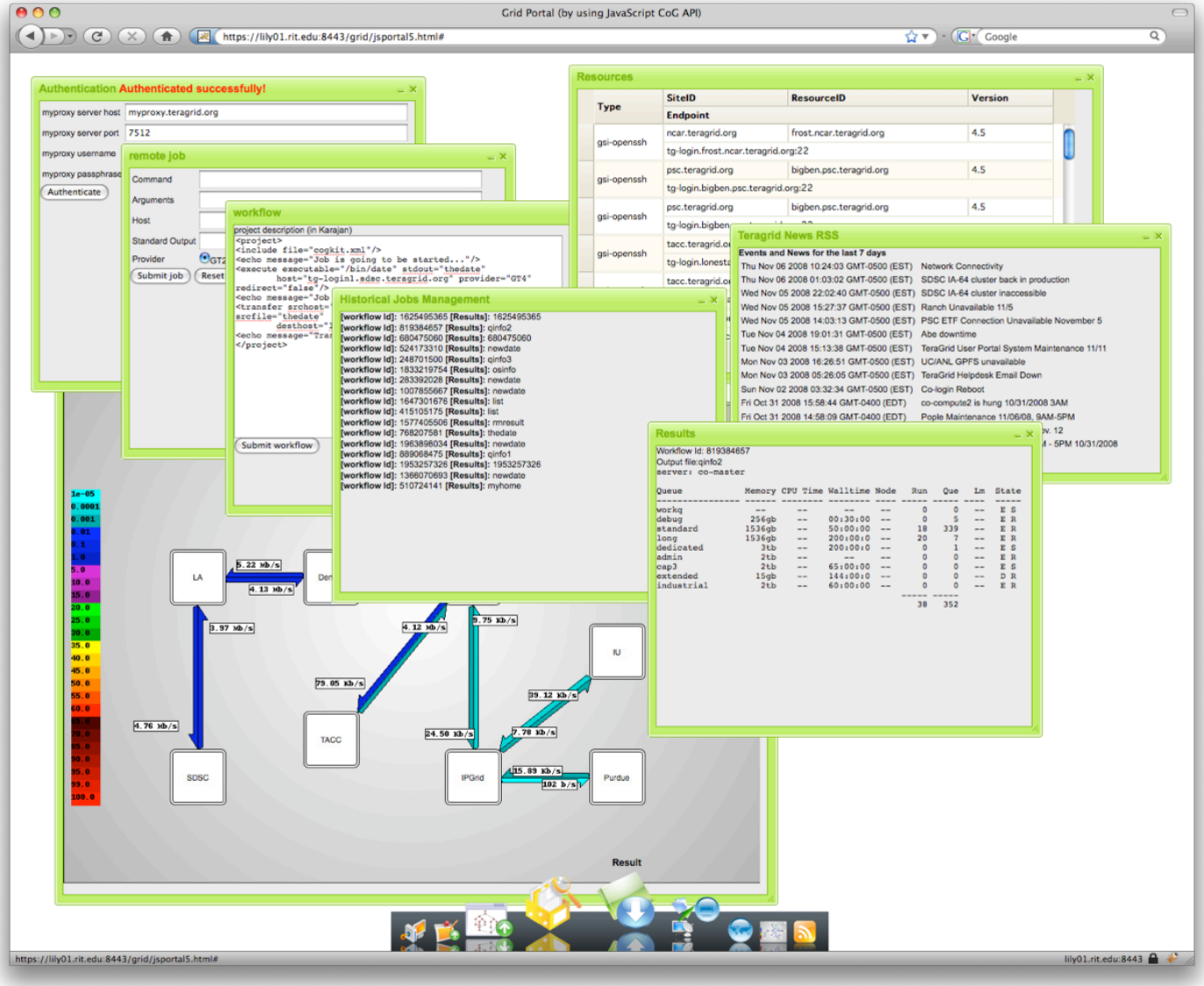


Gregor von Laszewski, http://www.mcs.anl.gov/gregor

If we only knew than …

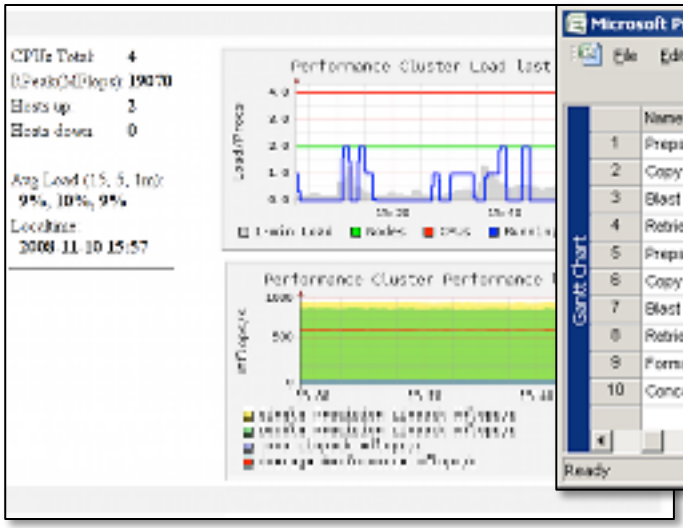
- what we know now …
  - The Future
  - We are working on some of the components as we speak, while others are already available in prototype form.

2008: Cyberaide Web 2.0

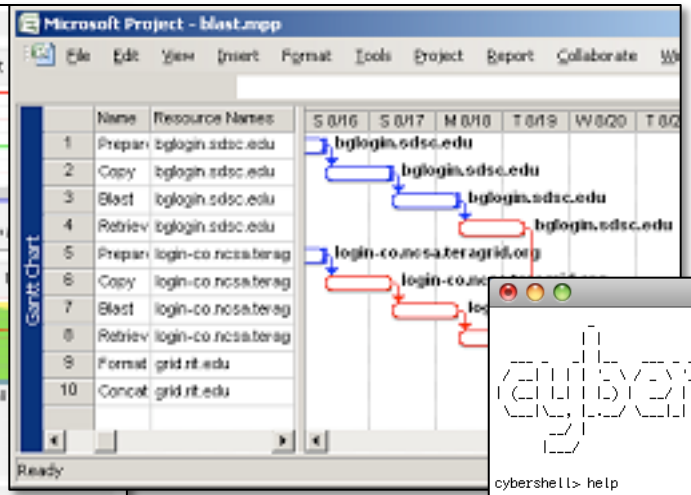


Cyberaide

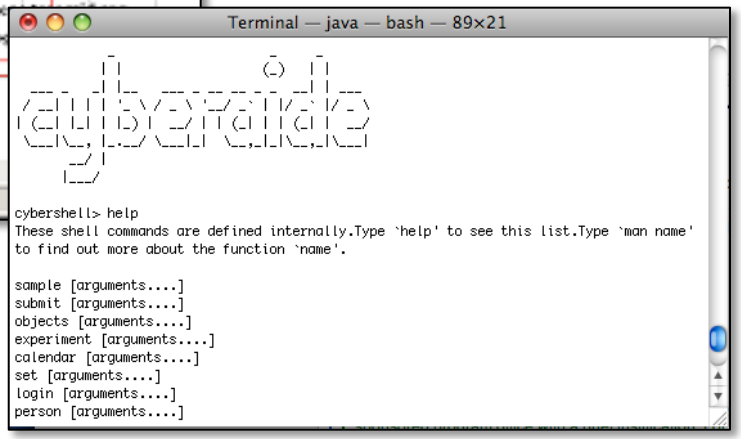
- Access advanced cyberinfrastructure easily



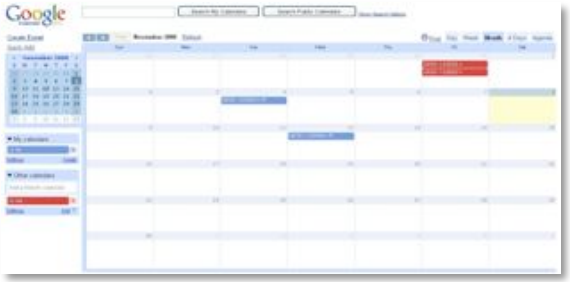
Application Monitoring



Workflow

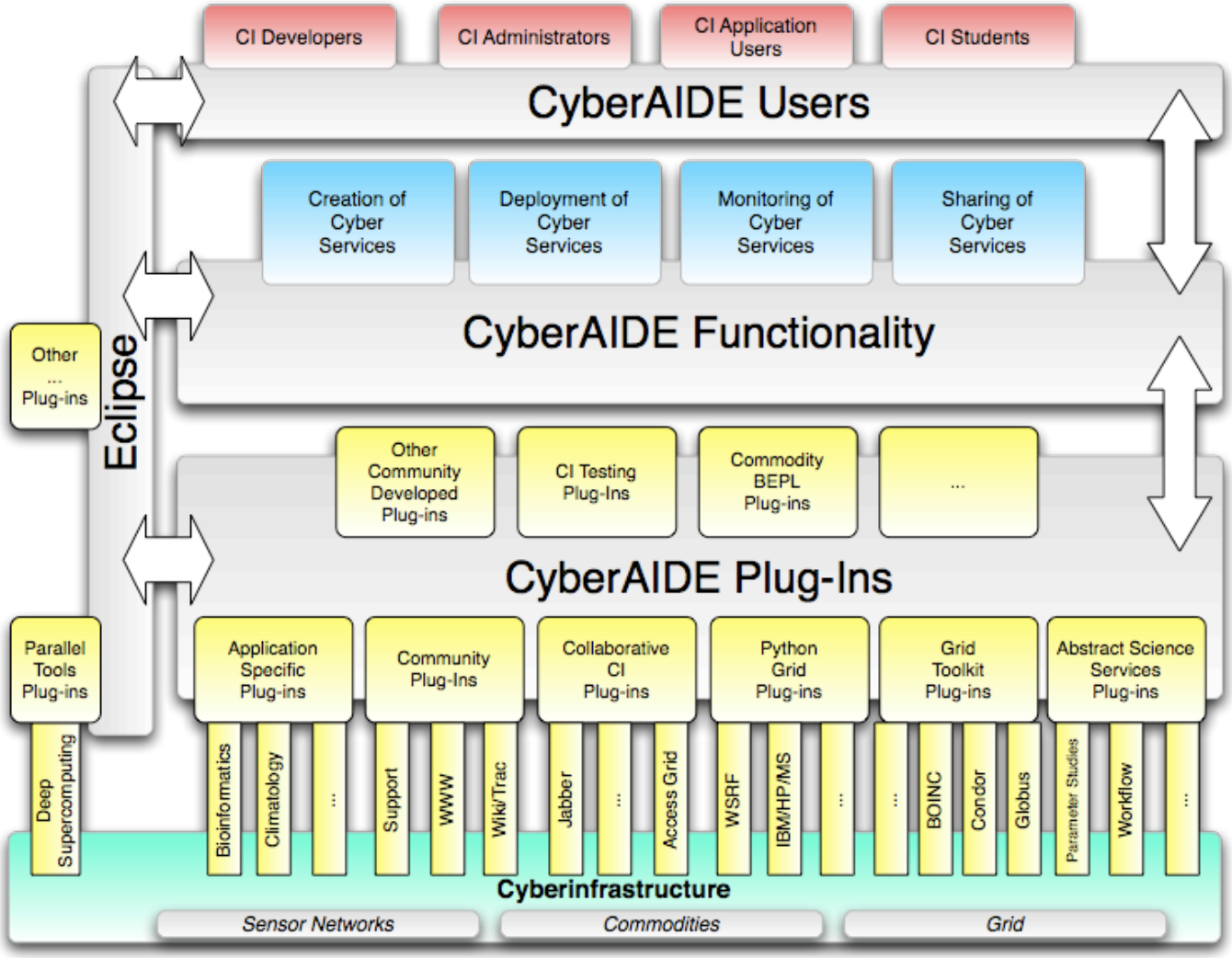


Object Shell

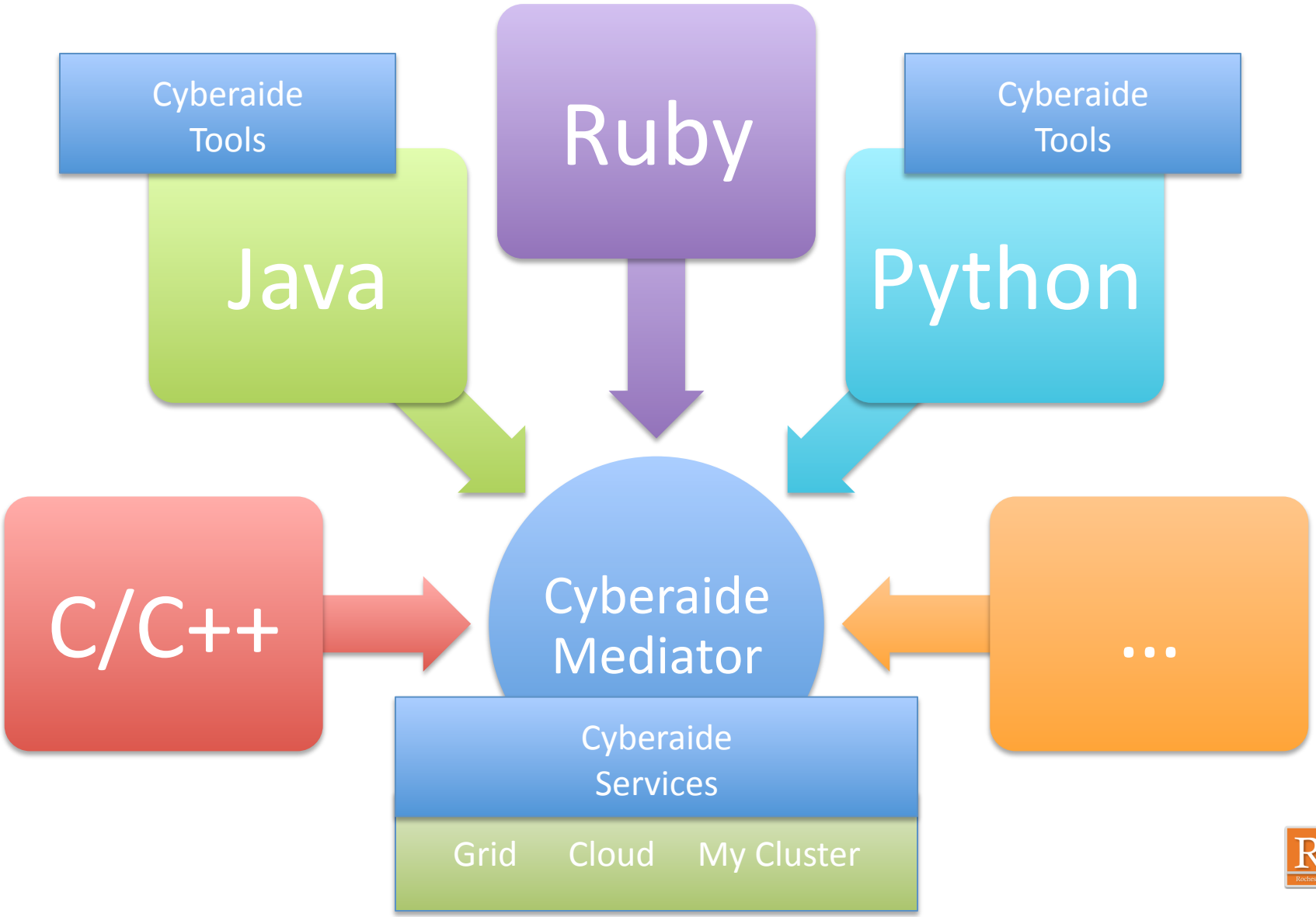


Shared Resource Calendar

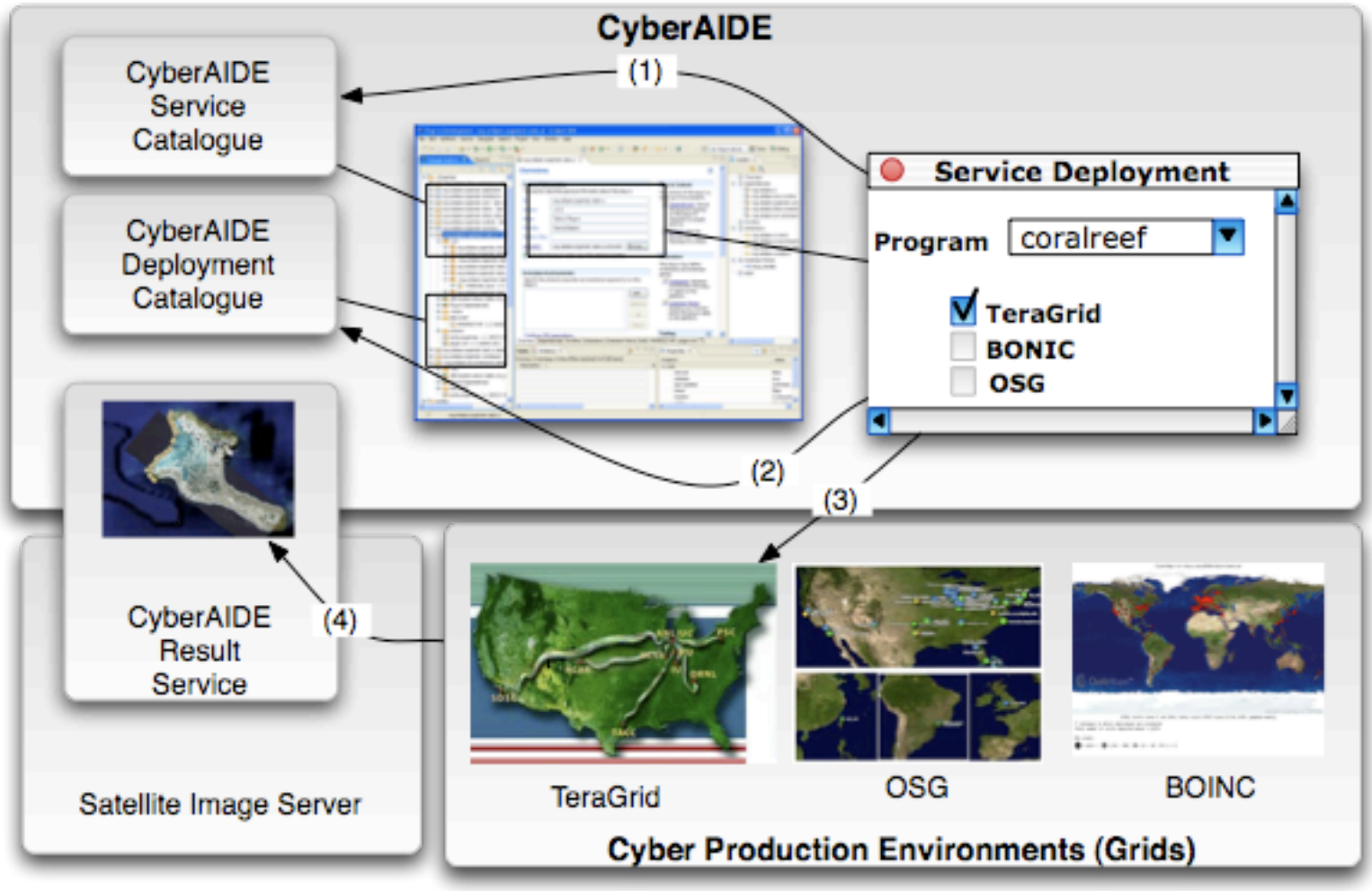
Cyberaide Framework



Mediating to the Future



Real Time Science Infrastructure



Politics, Administration, and other unfortunate hurdles

Administration

- Separate domains
  - Don't touch my stuff
- Network
  - Network between buildings
  - Network to sectors
  - It was faster to send data from one building to the other via Internet2 ;-)
- Accounting
  - Account administration
  - It took 3 month and 12 administrators to get me an account, which included rewriting my job description for doing the same thing ;-)
  - Dans group needed after me only 3 days ;-) and 1 admin
- Security
  - Firewalls
    - Firewalls are expensive
    - Limit bandwidth bellow what is possible
  - HW/SW Approval process
    - Software development process is faster than approval process -> always old version is installed
    - HW approval process was slow
    - No money for new HW
      - Here, let me give you my Windows 3.1 OS box that I have not used anymore for that last x years but its still here
      - That is no good

User Demand Contradictions

- Easy access and no security
  - vs. don't give my data out
- I need lots of data vs.
  - I can not analyze the data due to lack of compute power
- Privacy
  - vs. I don't give my data but you sure should give me yours
- Do things quickly for me
  - vs. do things right for many users
- Let me focus on science
  - vs. but I can program this all myself

Conclusion

- Globus is a Middleware Toolkit
  - We need more
- We need upperware
  - Cyberaide is just one example
- Administrative hurdles are significant
- Security is underestimated
- Networks are too slow
  - If you can string one wire, why not just put in 5 ;-)

Collaborate

- We are interested to collaborate with you.
- Send e-mail to laszewski@gmail.com
- http://www.cyberaide.org