

Services for Science

Ian Foster



Computation Institute

Argonne National Lab & University of Chicago



Thanks!

- DOE Office of Science



- NSF Office of Cyberinfrastructure



- National Institutes of Health

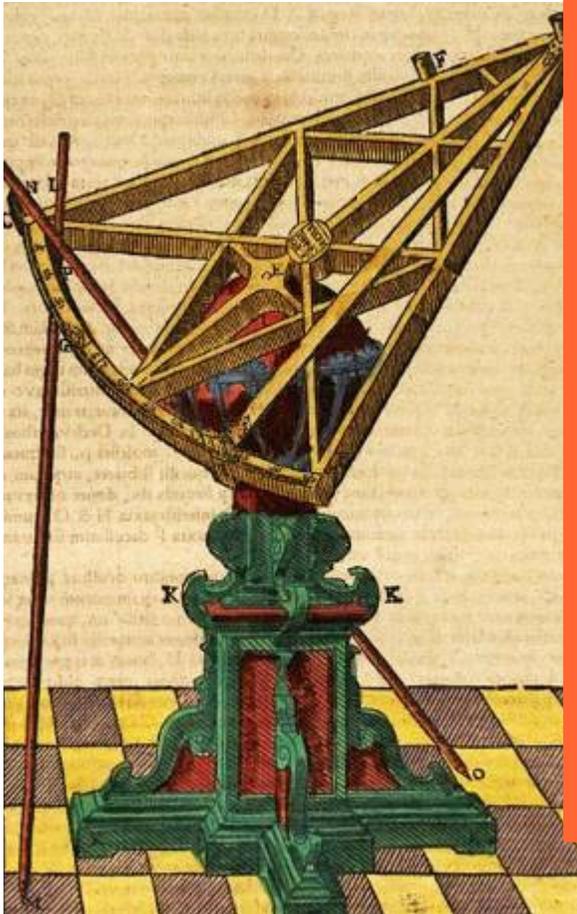
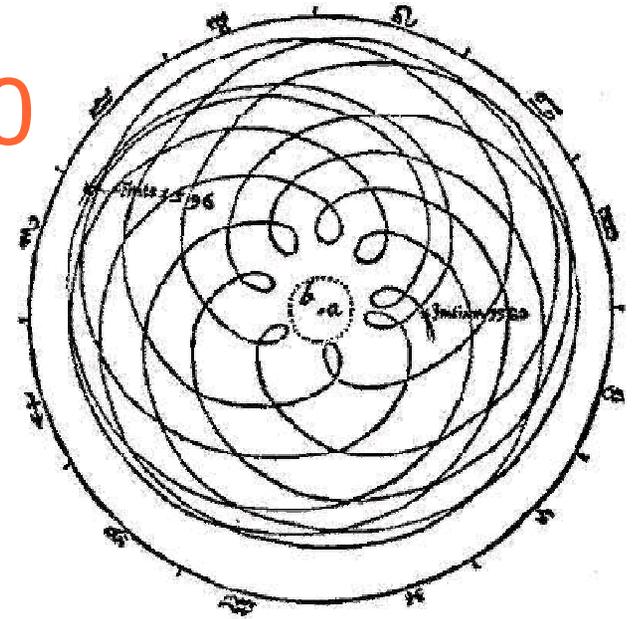


- Colleagues at Argonne, U.Chicago, USC/ISI, OSU, Manchester, and elsewhere



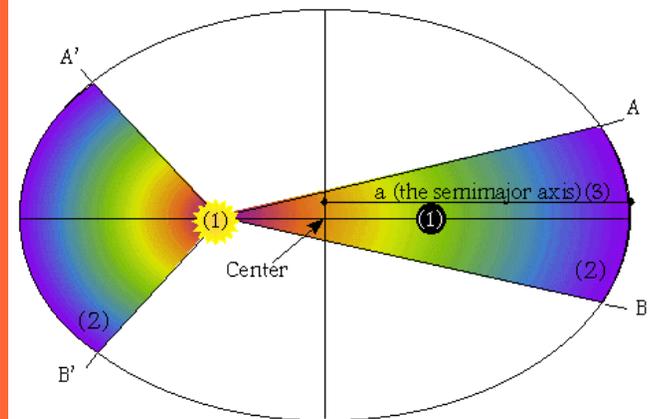
Scientific Communication, ~1600

DE MOTIB. STELLÆ MARTIS



Brahe

Kepler







Scientific Communication, ~2000 Service-Oriented Science

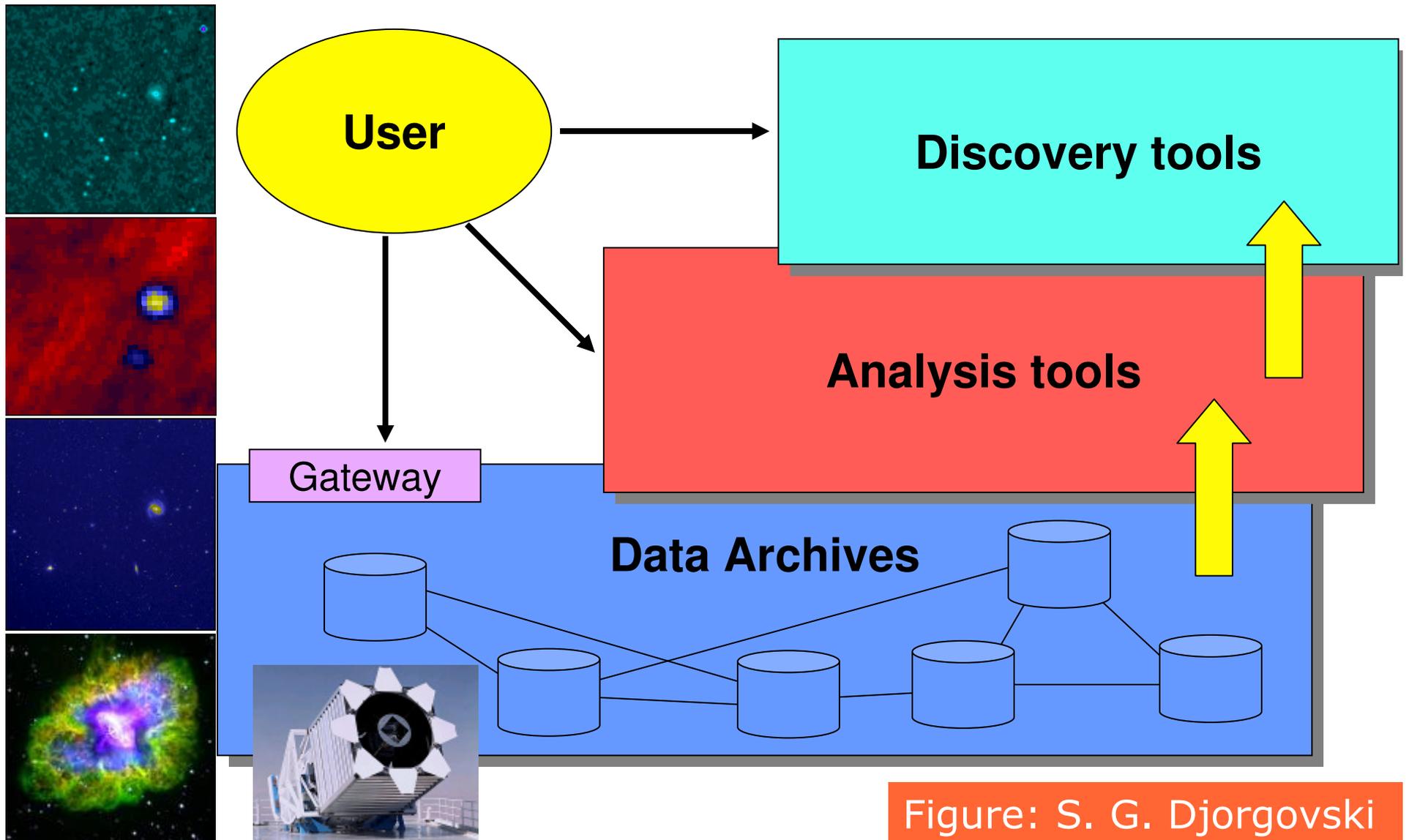
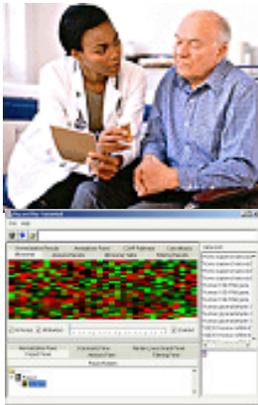


Figure: S. G. Djorgovski

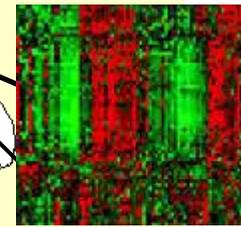
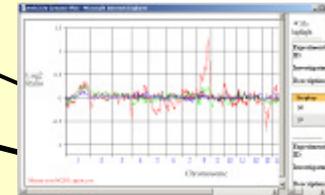
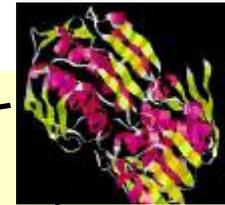
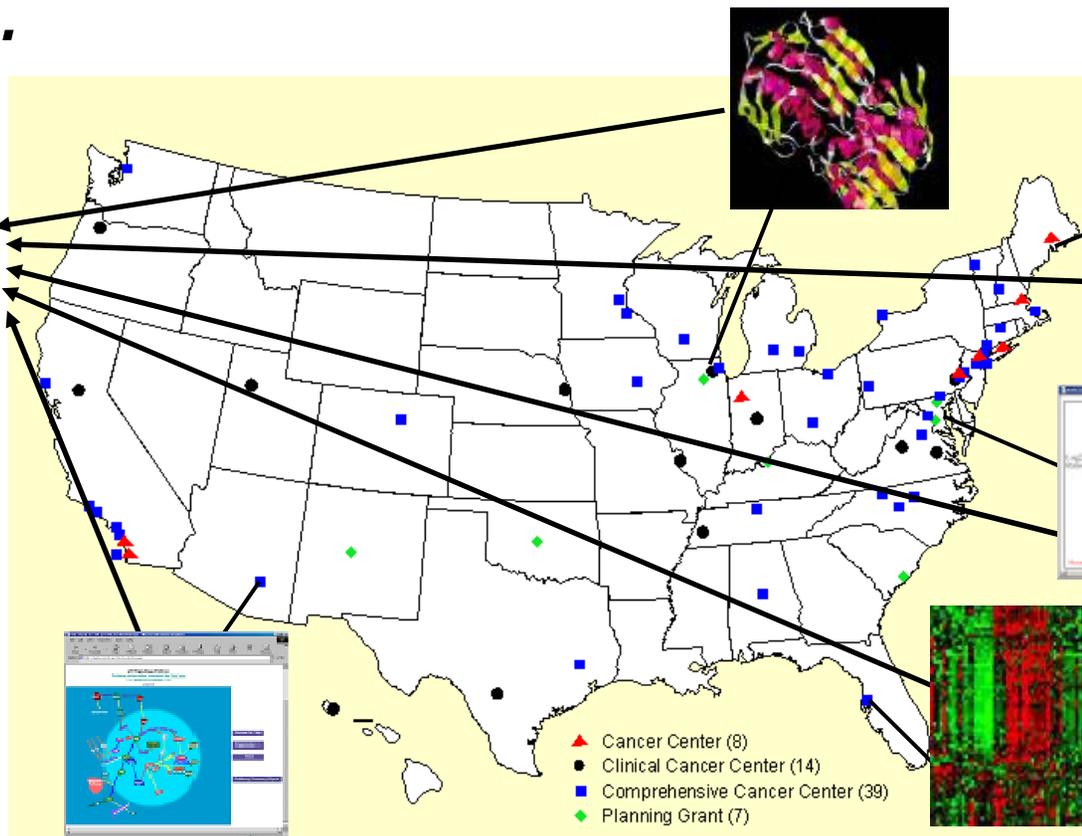


Services & Cancer Biology

caBIG: sharing of infrastructure, applications, and data.



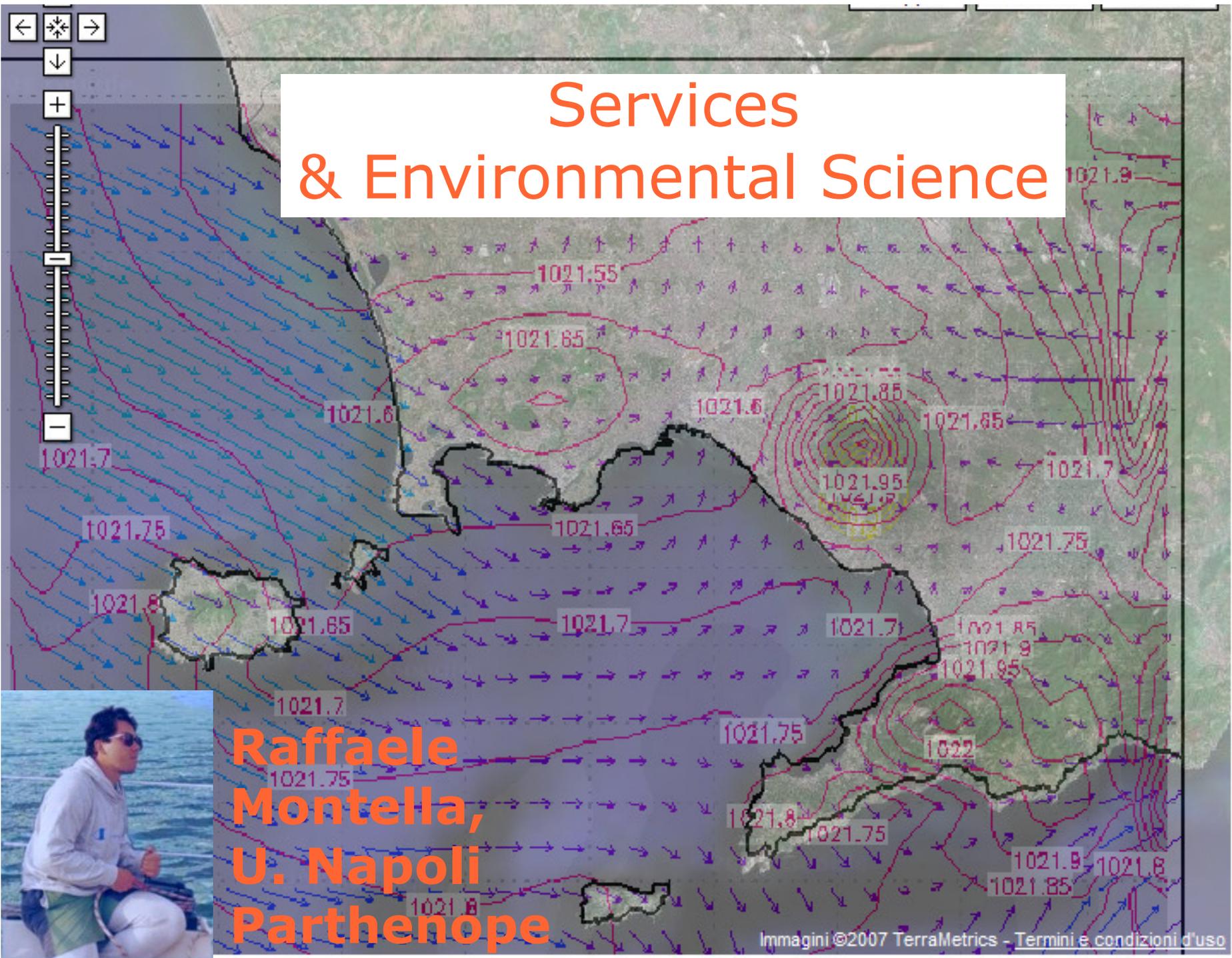
**Data
Integration!**



caBIG cancer Biomedical
Informatics Grid

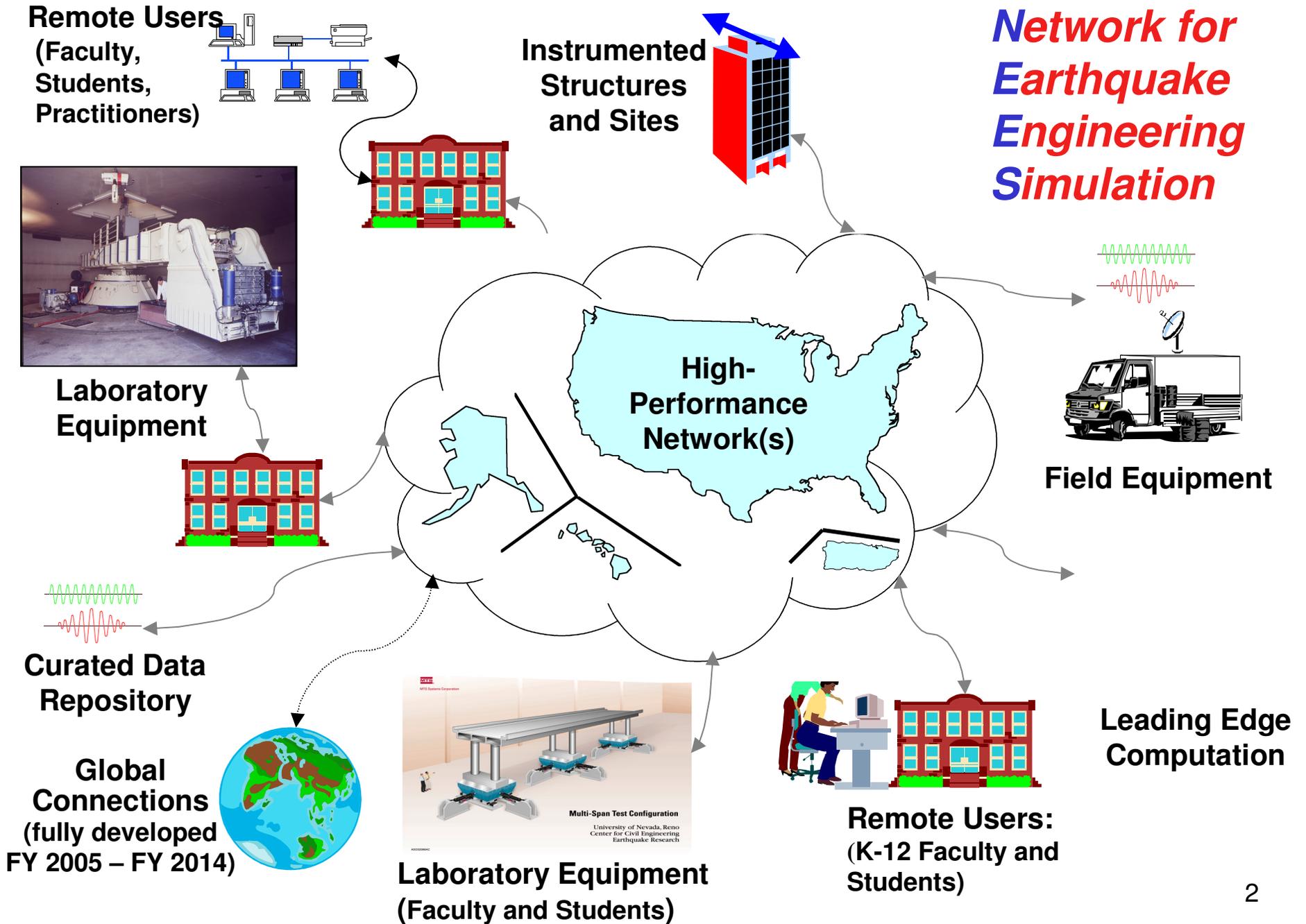


Services & Environmental Science



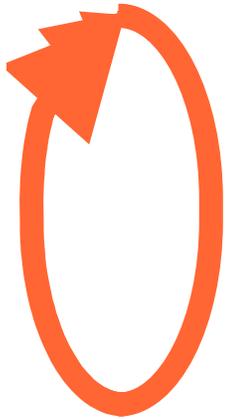
**Raffaele
Montella,
U. Napoli
Parthenope**

Network for Earthquake Engineering Simulation





Service-Oriented Science



People **create** services (data, code, instr.) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ *I find "someone else" to **host** services,
so I don't have to become an expert in
operating services & computers!*



TeraGrid[™]
EMPOWERING DISCOVERY



→ *I hope that this "someone else" can
manage security, reliability, scalability, ...*



"Service-Oriented Science", *Science*, 2005

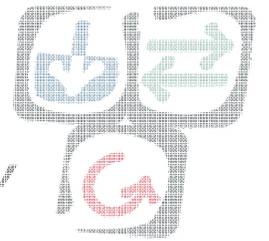


Creating Services

People **create** services (data, code, instr.) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ I find "someone else" to **host** services,
so I don't have to become an expert in
operating services & computers!

→ I hope that this "someone else" can
manage security, reliability, scalability, ...



TeraGrid™
EMPOWERING DISCOVERY

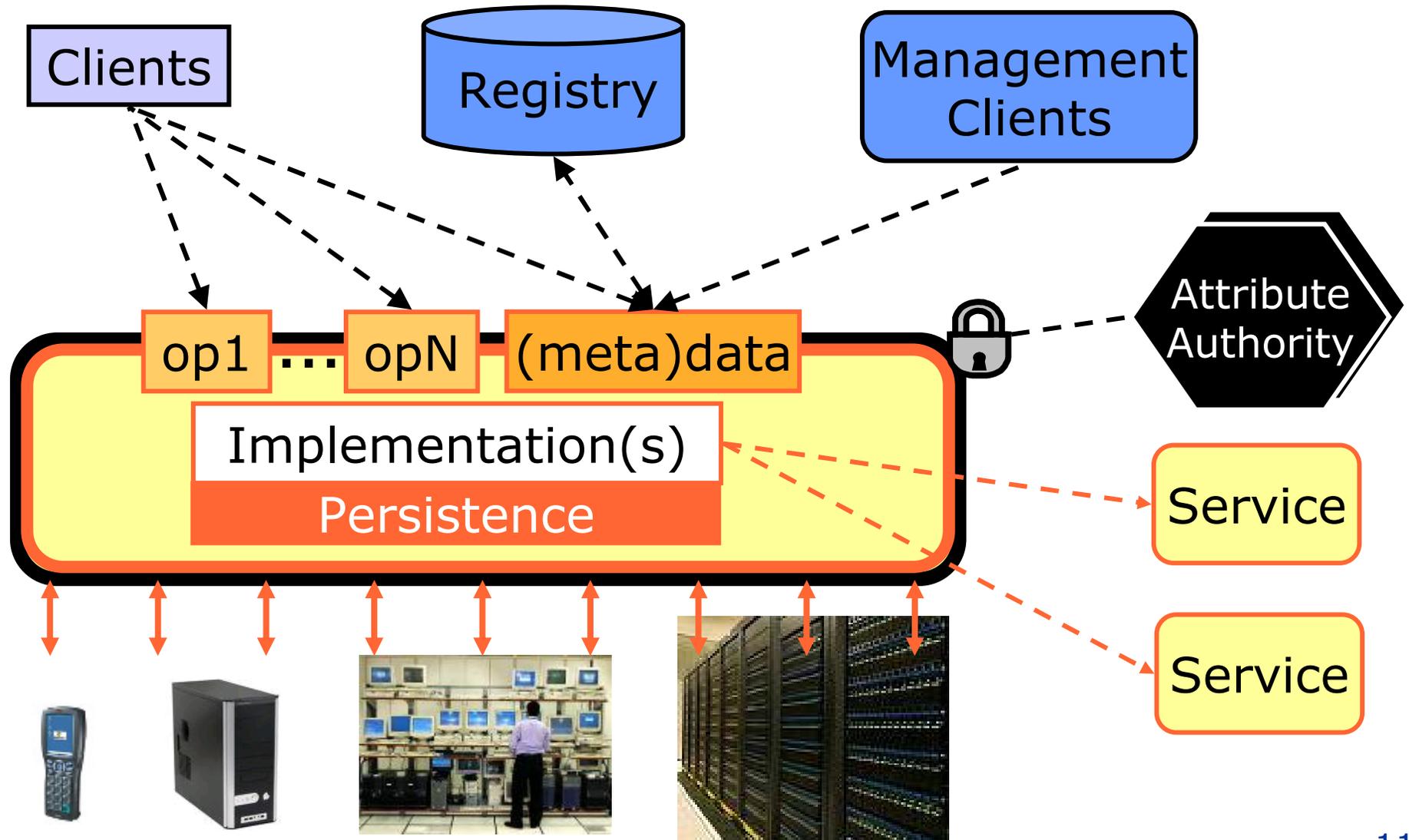


amazon
webservices™

"Service-Oriented Science", *Science*, 2005



Anatomy of a Service





Exposing Service State

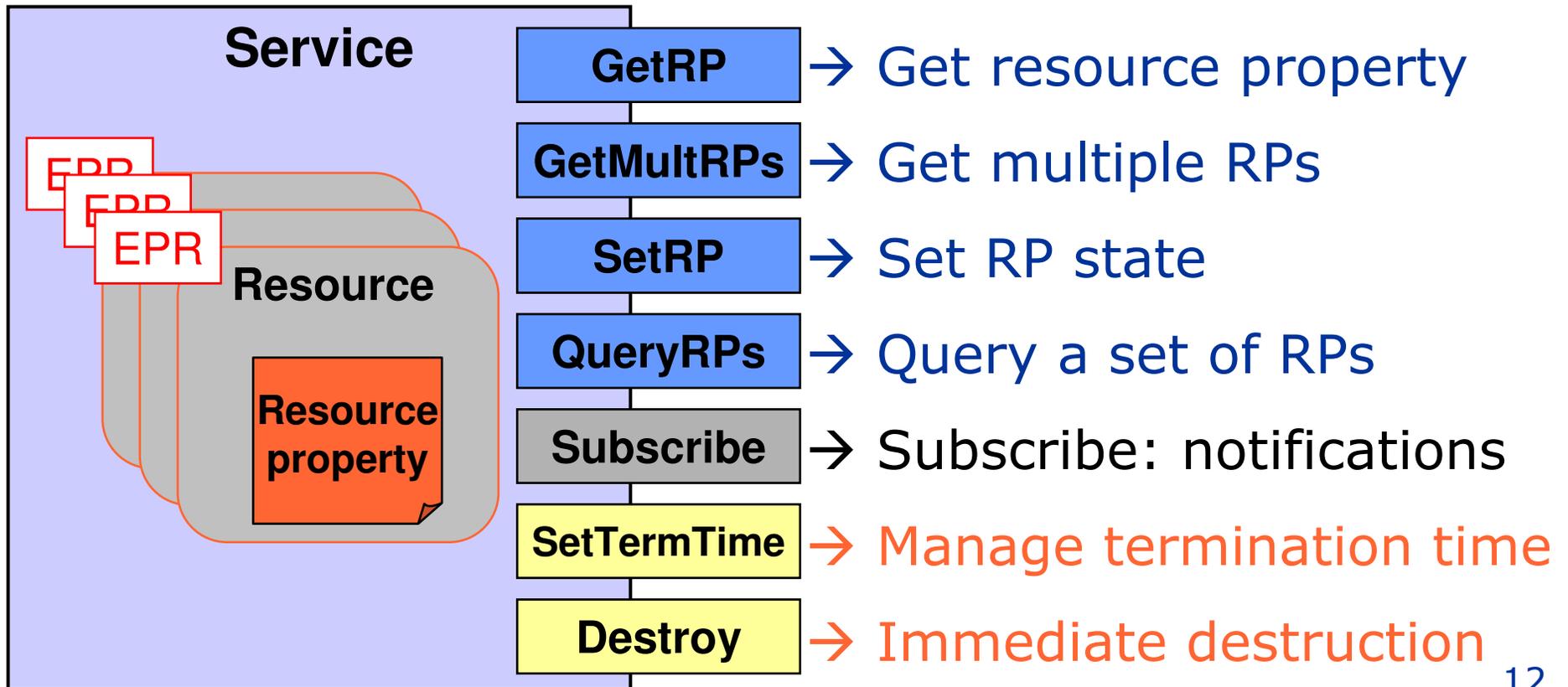
State representation

- ◆ Resource, resource property

State identification

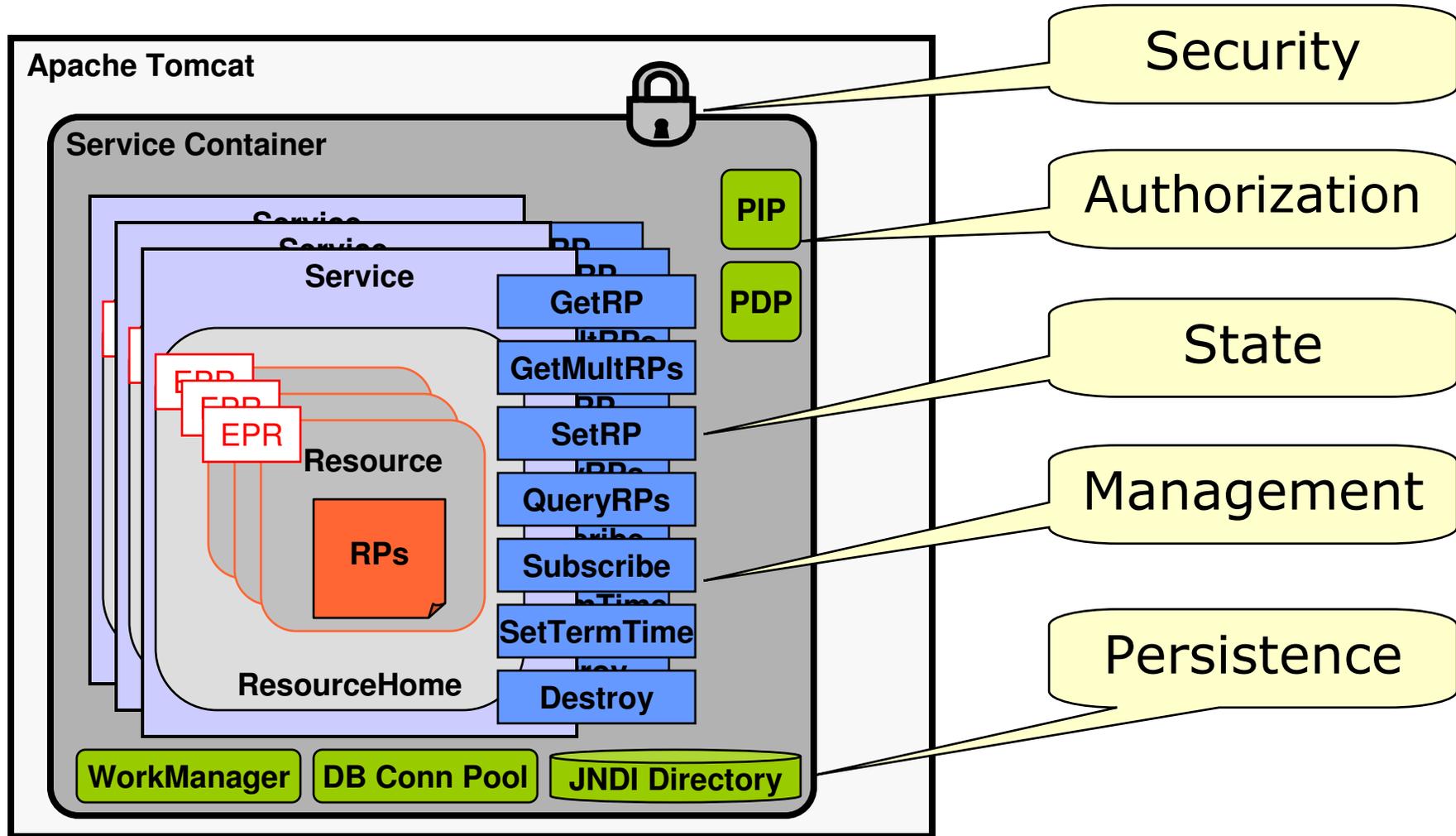
- ◆ Endpoint Reference

State Interfaces





Globus Toolkit Web Services Container





Creating Services (~2005)

“This **full-day tutorial** provides an introduction to programming Java services with the latest version of the Globus Toolkit version 4 (GT4). The tutorial teaches how to build a Java Service that makes use of GT4 mechanisms for state management, security, registry and related topics.”



Creating Services in 2008

Introduce and gRAVI

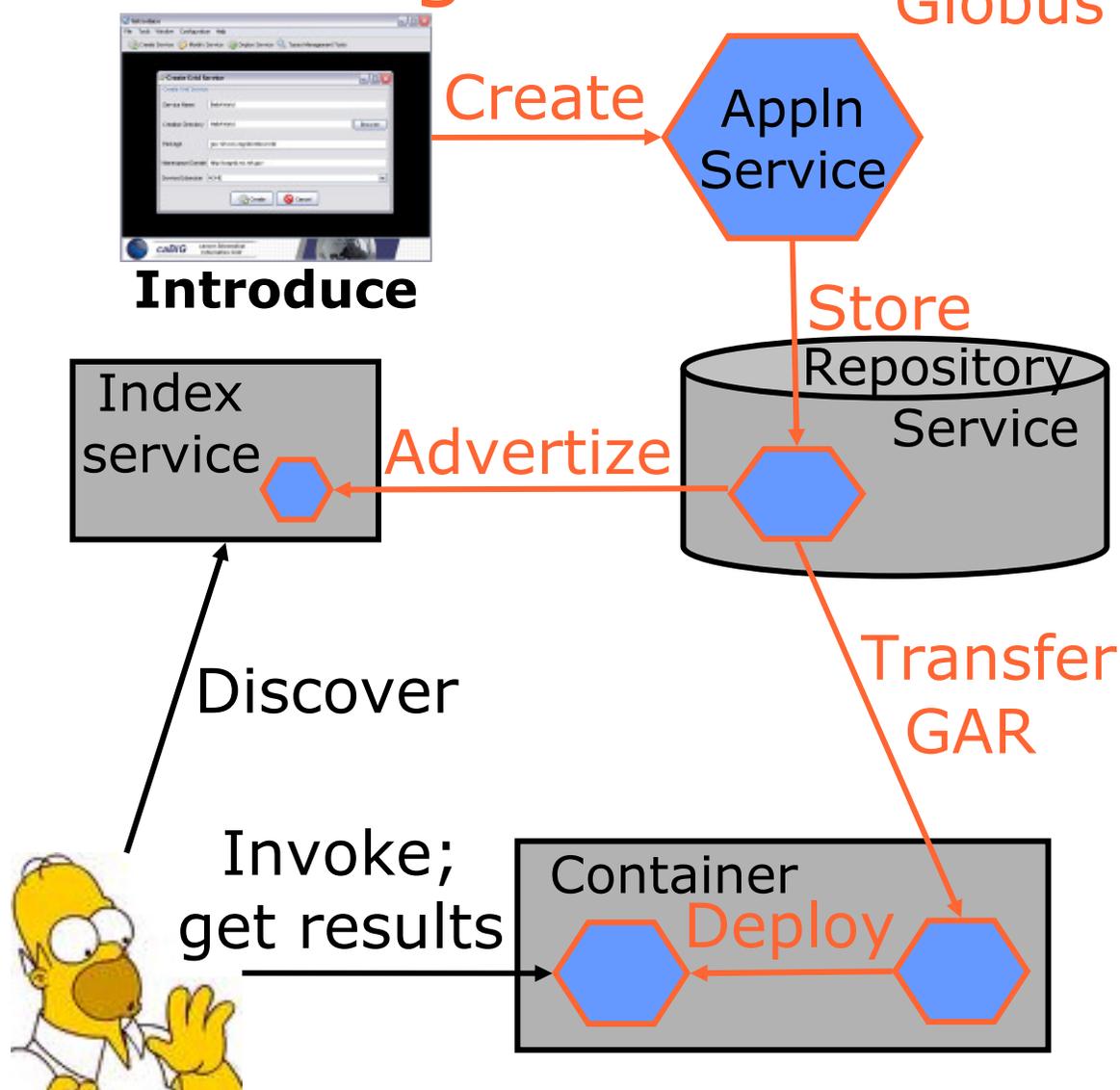


- Introduce

- ◆ Define service
- ◆ Create skeleton
- ◆ Discover types
- ◆ Add operations
- ◆ Configure security

- **Grid Remote Application Virtualization Infrastructure**

- ◆ Wrap executables



Creating Services

Introduce + gRAVI

Shannon Hastings
Scott Oster
David Ervin
Stephen Langella

Kyle Chard
Ravi Madduri

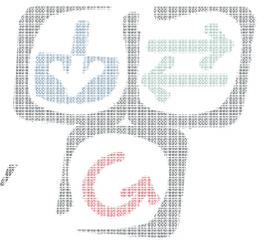


Discovering Services

People **create** services (data or functions) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ *I find "someone else" to **host** services, so I don't have to become an expert in operating services & computers!*

→ *I hope that this "someone else" can **manage** security, reliability, scalability, ...*



TeraGrid
EMPOWERING DISCOVERY



amazon
web services™

"Service-Oriented Science", *Science*, 2005



Discovering Services

Assume success → Billions of services



Semantics → Types, ontologies



Permissions → Can I use it?

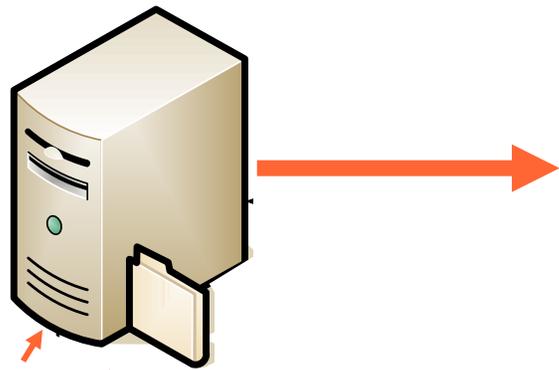


Reputation → The ultimate arbiter?





Discovery (1): Registries



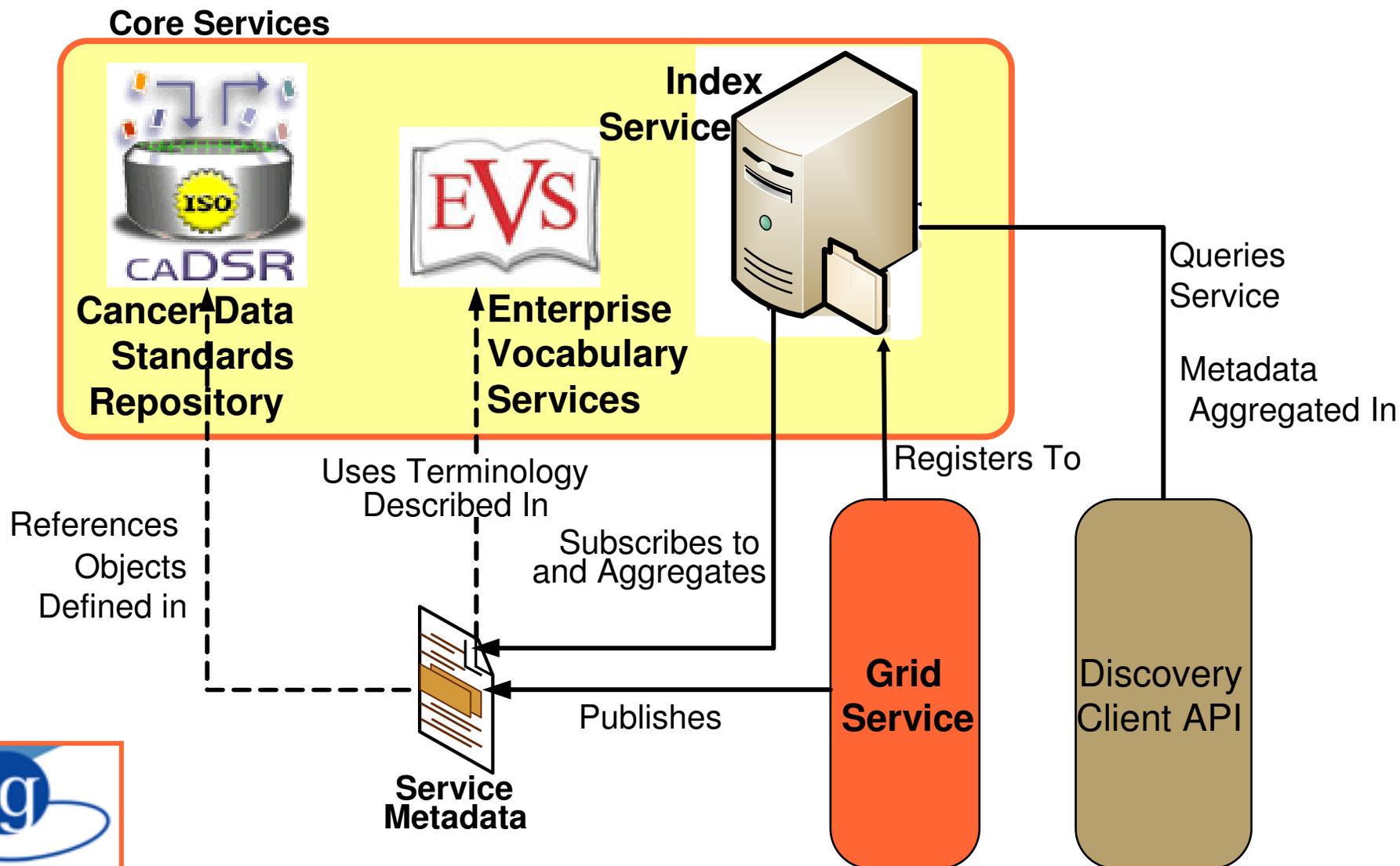
Globus

Search Watch loads

- WSDL @ <http://soap.bind.ca/wsdl/bind.wsdl>
- Soaplab @ <http://www.ebi.ac.uk/soaplab/emboss4/services/>
- GT4 Services@ <http://cagrid-index.nci.nih.gov:8080/wsrf/services/DefaultIndexService>
 - <http://165.112.133.34:42210/wsrf/services/cagrid/CAGWAS>
 - <http://140.254.80.50:50100/wsrf/services/cagrid/OrderQueueService>
 - <http://140.254.80.50:50101/wsrf/services/cagrid/DICOMDataService>
 - <https://cagrid-workflow.nci.nih.gov:8443/wsrf/services/cagrid/WorkflowFactoryService>
 - <http://cagrid-service.nci.nih.gov:8080/wsrf/services/cagrid/EV5GridService>
 - <https://cagrid-workflow.nci.nih.gov:8443/wsrf/services/cagrid/FederatedQueryProcess>
 - <https://cagrid-gts-slave.nci.nih.gov:8443/wsrf/services/cagrid/GTS>
 - <https://cagrid-gts-master.nci.nih.gov:8443/wsrf/services/cagrid/GTS>
 - <http://140.254.80.50:50101/wsrf/services/cagrid/GridImageResultService>
 - <http://sonicserver-dev.northwestern.edu:9000/wsrf/services/cagrid/NHISAS>
 - <https://caasers.mayo.edu:8443/wsrf/services/cagrid/CaaersDataService>
 - <http://cpas.nci.nih.gov:8080/wsrf/services/cagrid/CpasSvc>
 - <http://sonicserver-dev.northwestern.edu:9000/wsrf/services/cagrid/NHISDS>
 - <http://137.187.182.60:42210/wsrf/services/cagrid/CAGWAS>
 - <http://18.103.11.105:6060/wsrf/services/cagrid/PreprocessDatasetMAGEService>
 - <http://cananolab.nci.nih.gov:80/wsrf-canano/services/cagrid/CaNanoLabSvc>
 - <http://140.254.80.50:50010/wsrf/services/cagrid/PathologyDataService>
 - <https://cagrid-cds.nci.nih.gov:8443/wsrf/services/cagrid/CredentialDelegationService>
 - <http://cabio-gridservice.nci.nih.gov:80/wsrf-cabio/services/cagrid/CaBIOSvc>
 - <https://cagrid-auth.nci.nih.gov:8443/wsrf/services/cagrid/AuthenticationService>
 - <http://cgwb.nci.nih.gov:80/wsrf/services/cagrid/CGWBDataService>
 - <http://cagrid-service.nci.nih.gov:8080/wsrf/services/cagrid/CaDSRService>
 - <http://140.254.80.50:50100/wsrf/services/cagrid/DICOMDataService>
 - <http://cananolab.abcc.ncifcrf.gov:8080/wsrf-canano/services/cagrid/CaNanoLabSvc>
 - <http://cacti-g5-12.wustl.edu:8080/wsrf-canano/services/cagrid/CaNanoLabSvc>
 - <http://128.252.227.214:18080/wsrf/services/cagrid/CaArraySvc>
 - <https://137.187.182.58:58091/wsrf/services/cagrid/CaTissueCore>
 - <https://137.187.182.60:49210/wsrf/services/cagrid/CaTissueCore>
 - <http://140.254.80.50:50101/wsrf/services/cagrid/ImageDataService>
 - <https://cagrid-dorian.nci.nih.gov:8443/wsrf/services/cagrid/Dorian>
 - <http://cagrid-service.nci.nih.gov:8080/wsrf/services/cagrid/GlobalModelExchange>
 - <https://cagrid-gridgrouper.nci.nih.gov:8443/wsrf/services/cagrid/GridGrouper>



Discovery (2): Standardized Vocabularies



Globus



Discovery (3): Tagging & Social Networking

Section II. Risk Response

Data Set Name: [j02tab.dta](#)

Variables:

`line_id:rr7_id:rr7a:rr7b:rr7oth:rr8:`

- Case Count: 2,320
- Variable Count: 14
- Size: 6,712KB
- Unit of Observation: Household action
- Record per Case: 37 (total records 85,840)

Section IV. Household Composition

Data Set Name: [j05.dta](#)

Variables:

`line_id:hc_id:hc2:hc3:hc3o:hc4:hc5:hc6:hc7:hc7o:hc8:hc8a:hc8b:`

- Case Count: 2,87
- Variable Count: 2
- Size: 1,555KB
- Unit of Observation:
- Record per Case:

Section V. Occup

Data Set Name: [j0](#)

- Case Count: 2,87
- Variable Count: 21
- Size: 2,283KB
- Unit of Observation: Household member
- Record per Case: varied (total records 10,602)

hc3	
<input type="checkbox"/> select	
spitzname handle	
nick name nickname	
moniker	
Enter your tag here...	
TagIt	Close

GLOSS:
Generalized
Labels Over Scientific
data Sources
(Foster, Nestorov)



Discovery (3): Tagging & Social Networking



David de Roure,
Carole Goble,
et al.

Use myExperiment to...

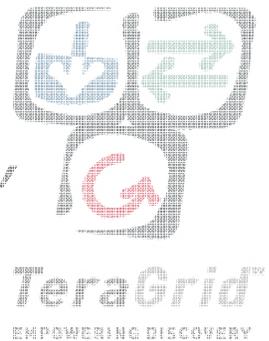
-  Find **Workflows**
-  Find **Files**
-  Share Your **Workflows and Files**
-  Join **Groups**
-  Find **People** and Make Friends
-  Send **Messages**
-  Get **Feedback**
-  Tag and Rate things
-  Write **Reviews and Comments**
-  Build your **Profile and Reputation**



Composing Services

People **create** services (data or functions) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ *I find "someone else" to **host** services,
so I don't have to become an expert in
operating services & computers!*



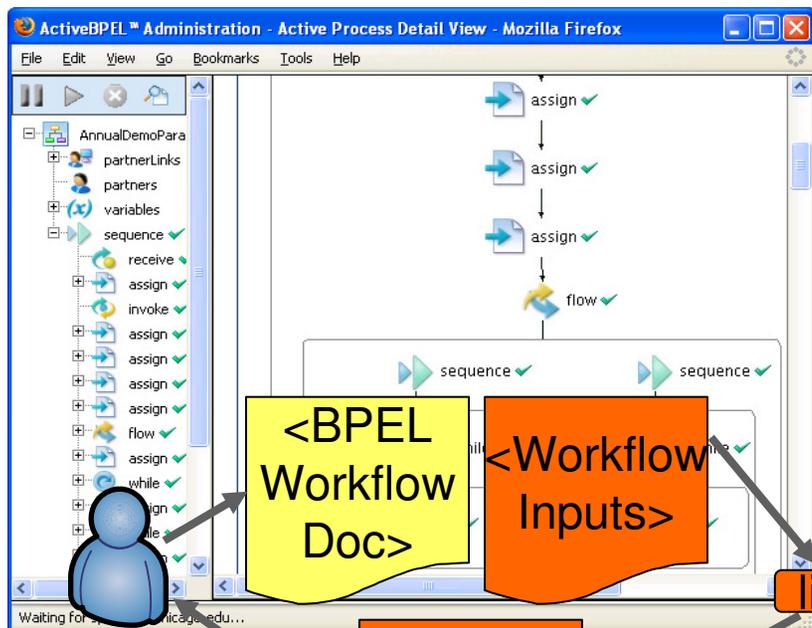
→ *I hope that this "someone else" can
manage security, reliability, scalability, ...*



"Service-Oriented Science", *Science*, 2005



Composing Services: E.g., BPEL Workflow System



Researcher
Or Client App

<BPEL
Workflow
Doc>

<Workflow
Inputs>

<Workflow
Results>

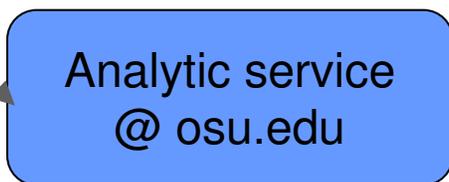
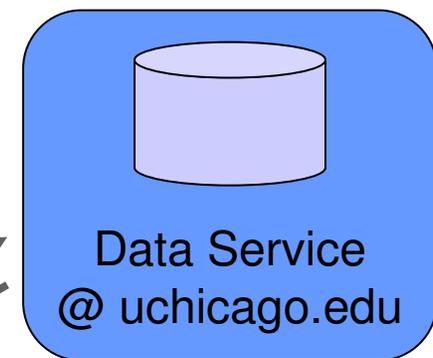
link

BPEL
Engine

link

link

link



See also Kepler & Taverna



caBIG™ cancer Biomedical Informatics Grid™

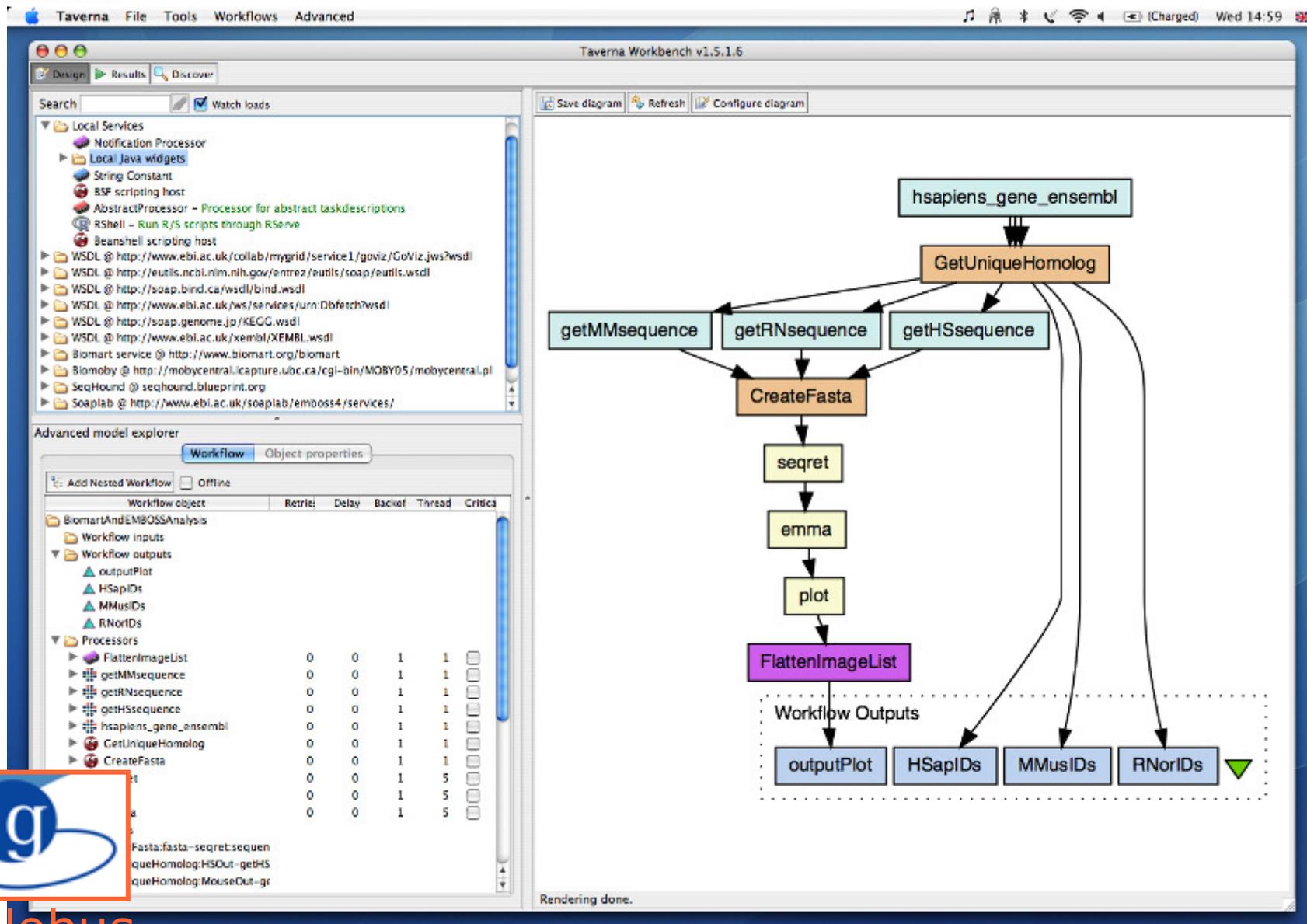
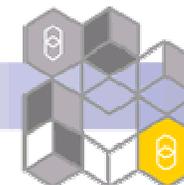
an initiative of the National Cancer Institute





Composing Services

myGrid



Globus

Composing Services

Taverna + GT4

Taverna team

Wei Tan

Ravi Madduri

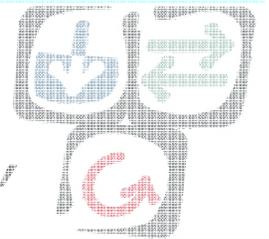


Publishing Services

People **create** services (data or functions) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ I find "someone else" to **host** services,
so I don't have to become an expert in
operating services & computers!

→ I hope that this "someone else" can
manage security, reliability, scalability, ...



TeraGrid[™]
EMPOWERING DISCOVERY



"Service-Oriented Science", *Science*, 2005



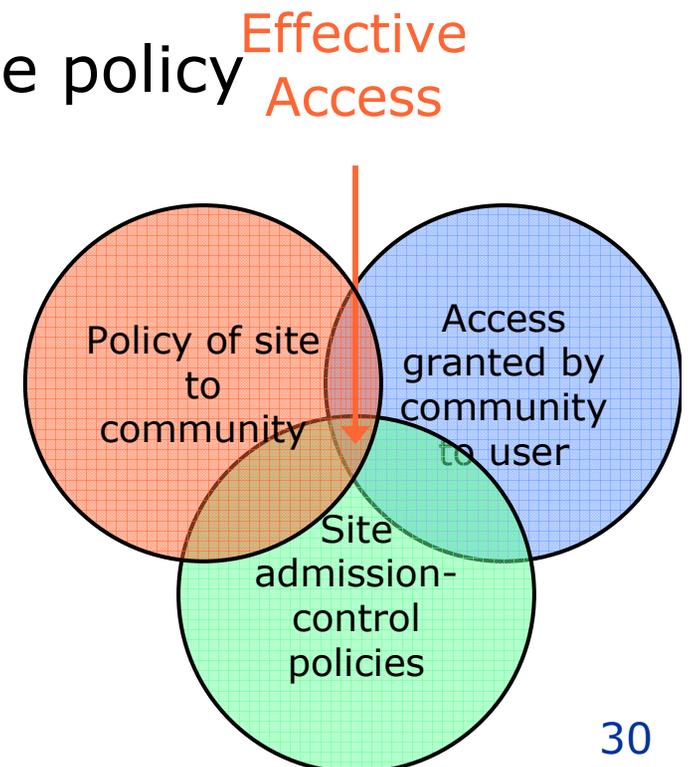
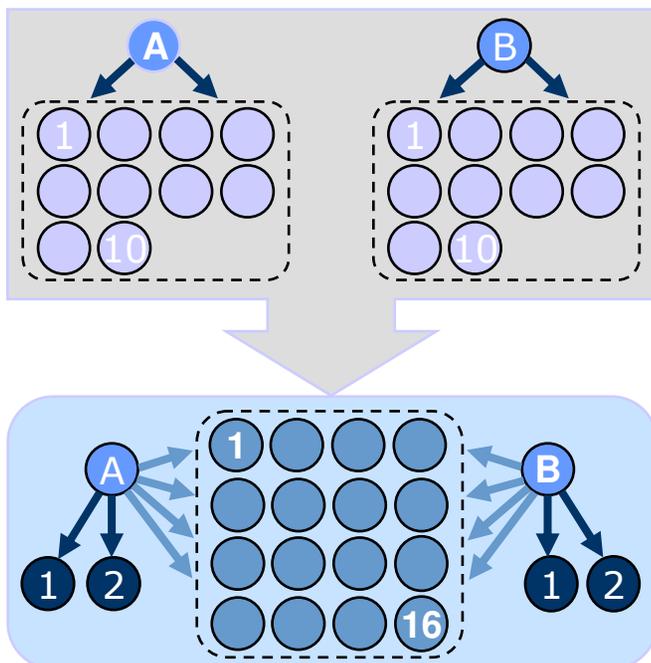
Publishing Services

- Description → Syntax, semantics
- State → Availability, load, ...
- Policies → Who, what, when, ...
- Hosting → Location, scalability, ...



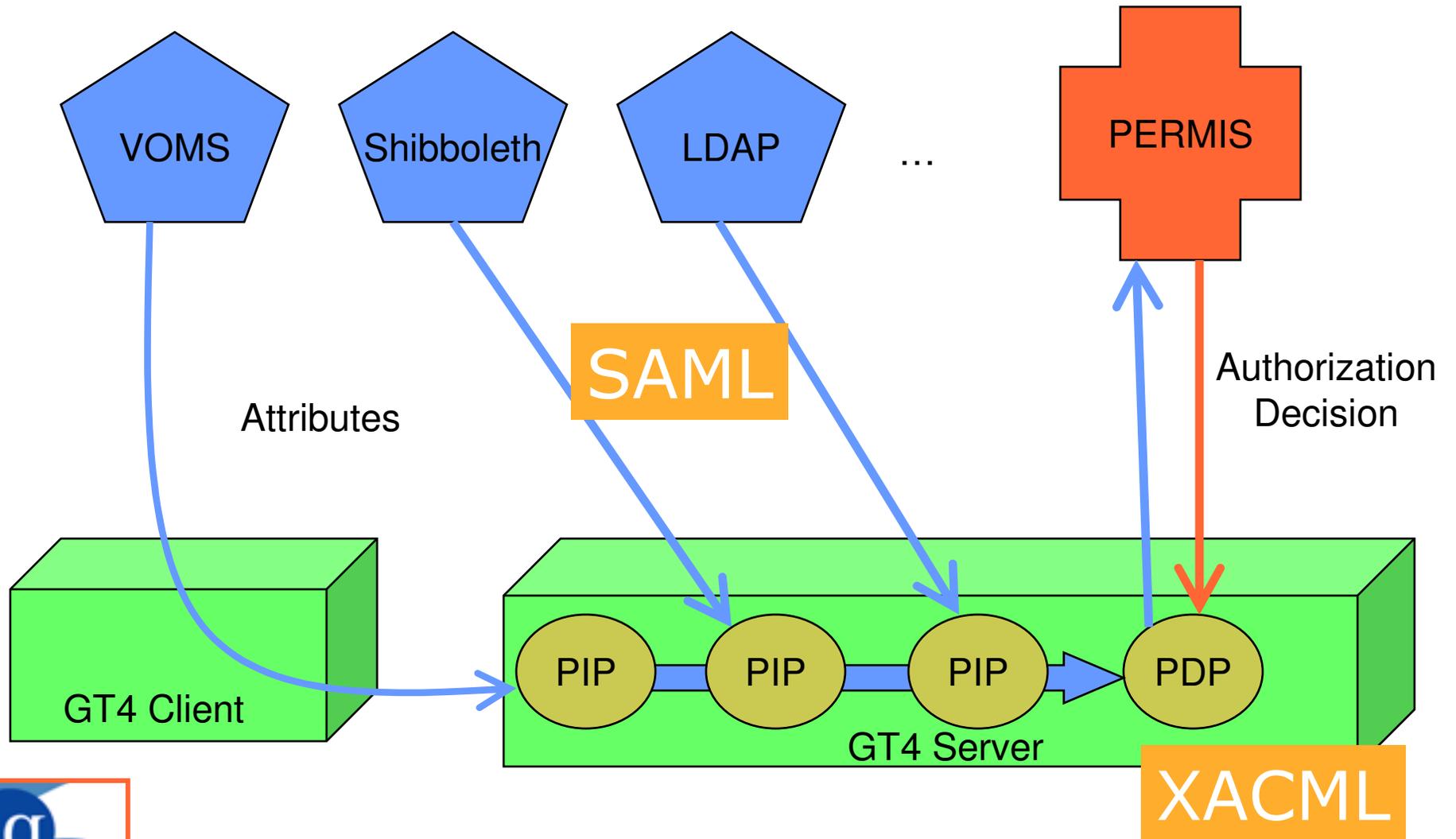
Defining Community: Membership and Laws

- Identify VO participants and roles
 - ◆ For people and services
- Specify and control actions of members
 - ◆ Empower members → delegation
 - ◆ Enforce restrictions → federate policy



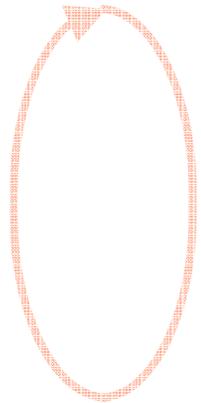


Authorization: SAML & XACML





Hosting Services



People **create** services (data or functions) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ I find "someone else" to **host** services,
so I don't have to become an expert in
operating services & computers!



TeraGrid™
EMPOWERING DISCOVERY



→ I hope that this "someone else" can
manage security, reliability, scalability, ...



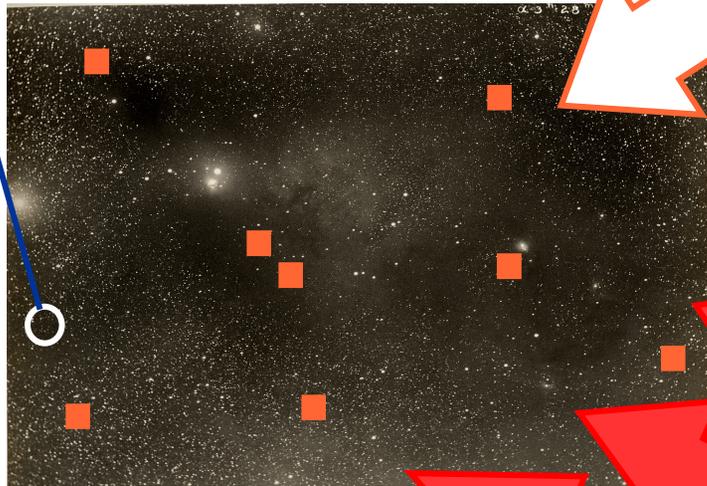
"Service-Oriented Science", *Science*, 2005



The Importance of "Hosting" and "Management"

Tell me about this star

Tell me about these 20K stars



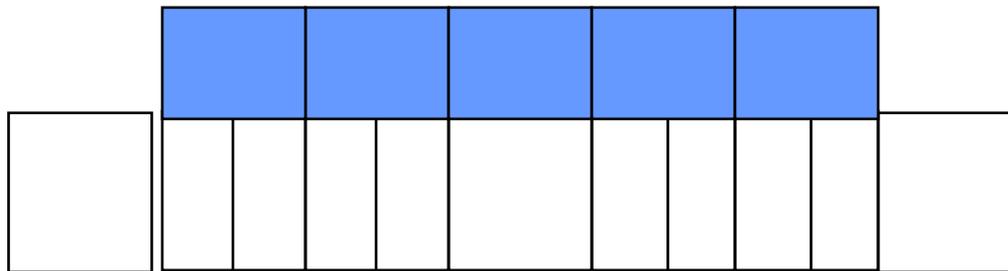
Support 1000s of users



E.g., Sloan Digital Sky Survey, ~10 TB; others much bigger



The Two Dimensions of Service-Oriented Science



Function
Resource

- **Decompose** across network

Clients **integrate** dynamically

- ◆ Select & compose services
- ◆ Select "best of breed" providers
- ◆ Publish result as new services

- Decouple **resource** & **service** providers

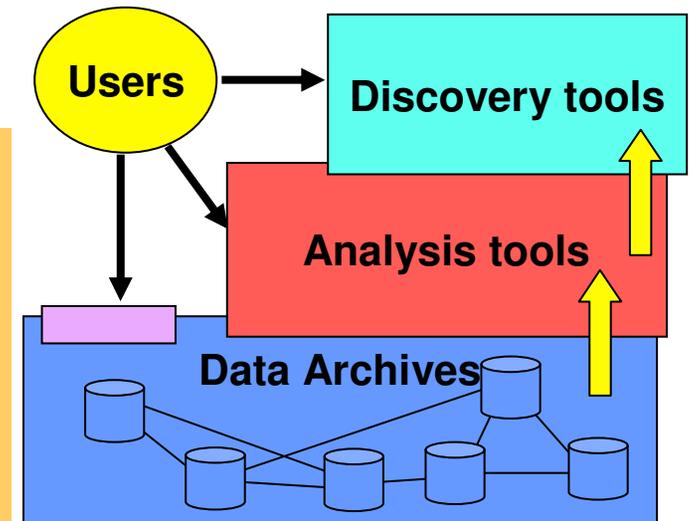


Fig: S. G. Djorgovski
34

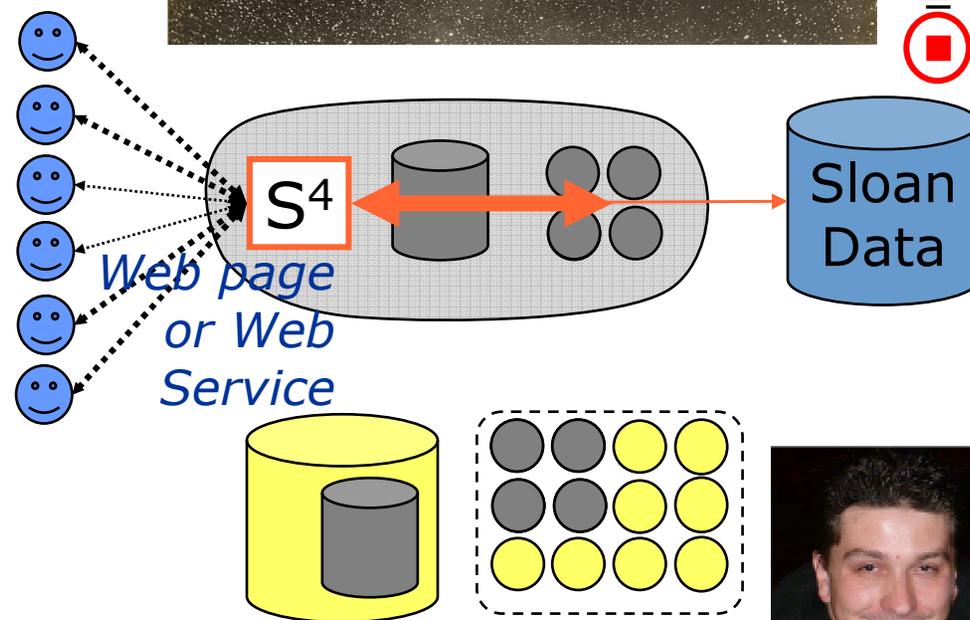
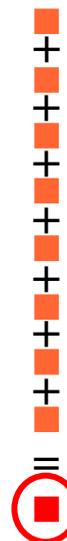
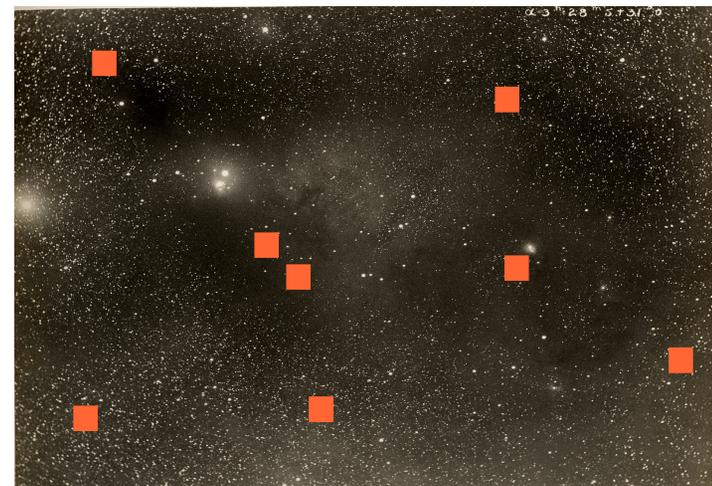


Dynamic Provisioning: Falkon Applied to Stacking



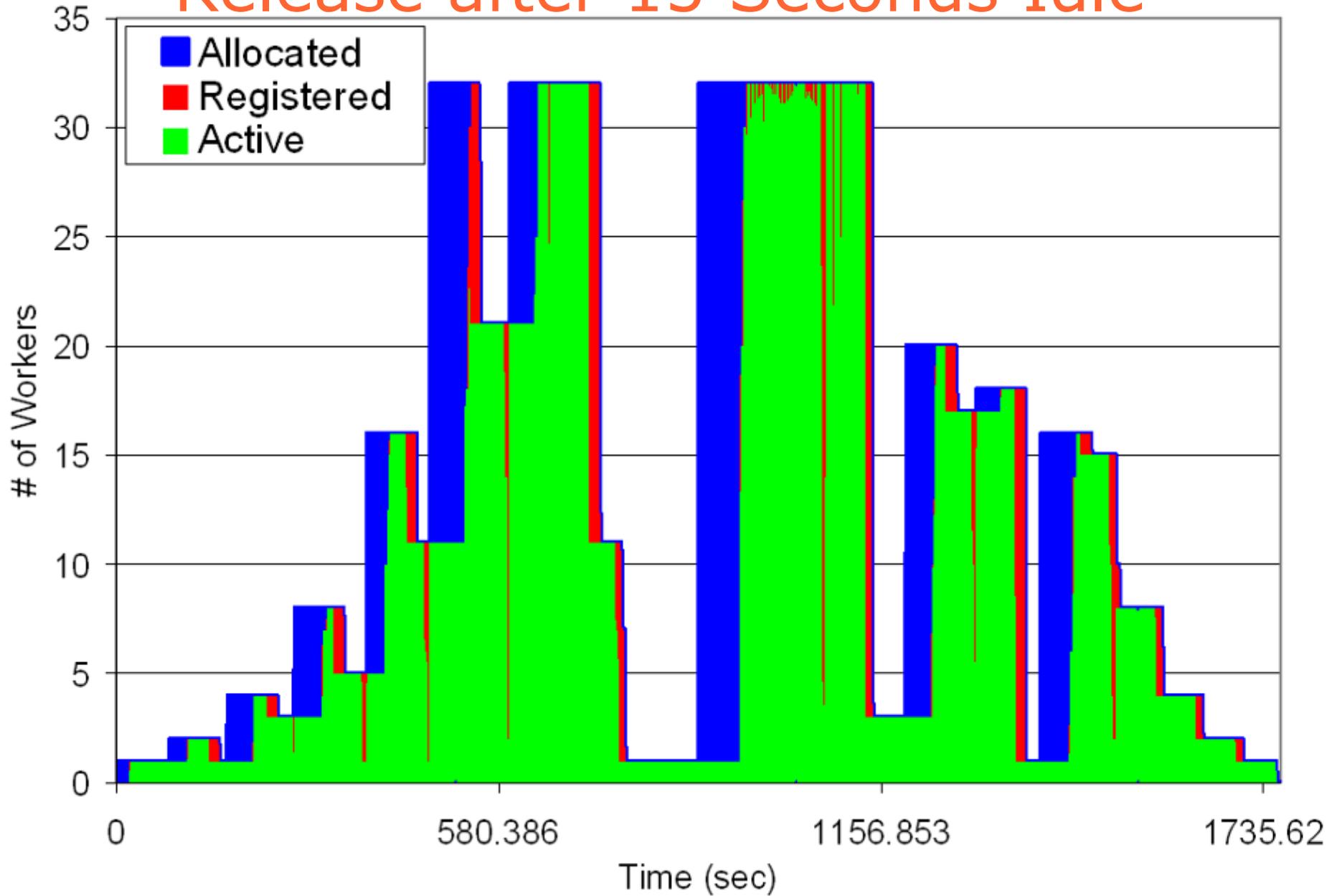
Globus

- Purpose
 - ◆ On-demand “stacks” of random locations within ~10TB dataset
- Challenge
 - ◆ Rapid access to 10-10K “random” files
 - ◆ Time-varying load
- Solution
 - ◆ Dynamic acquisition of compute, storage

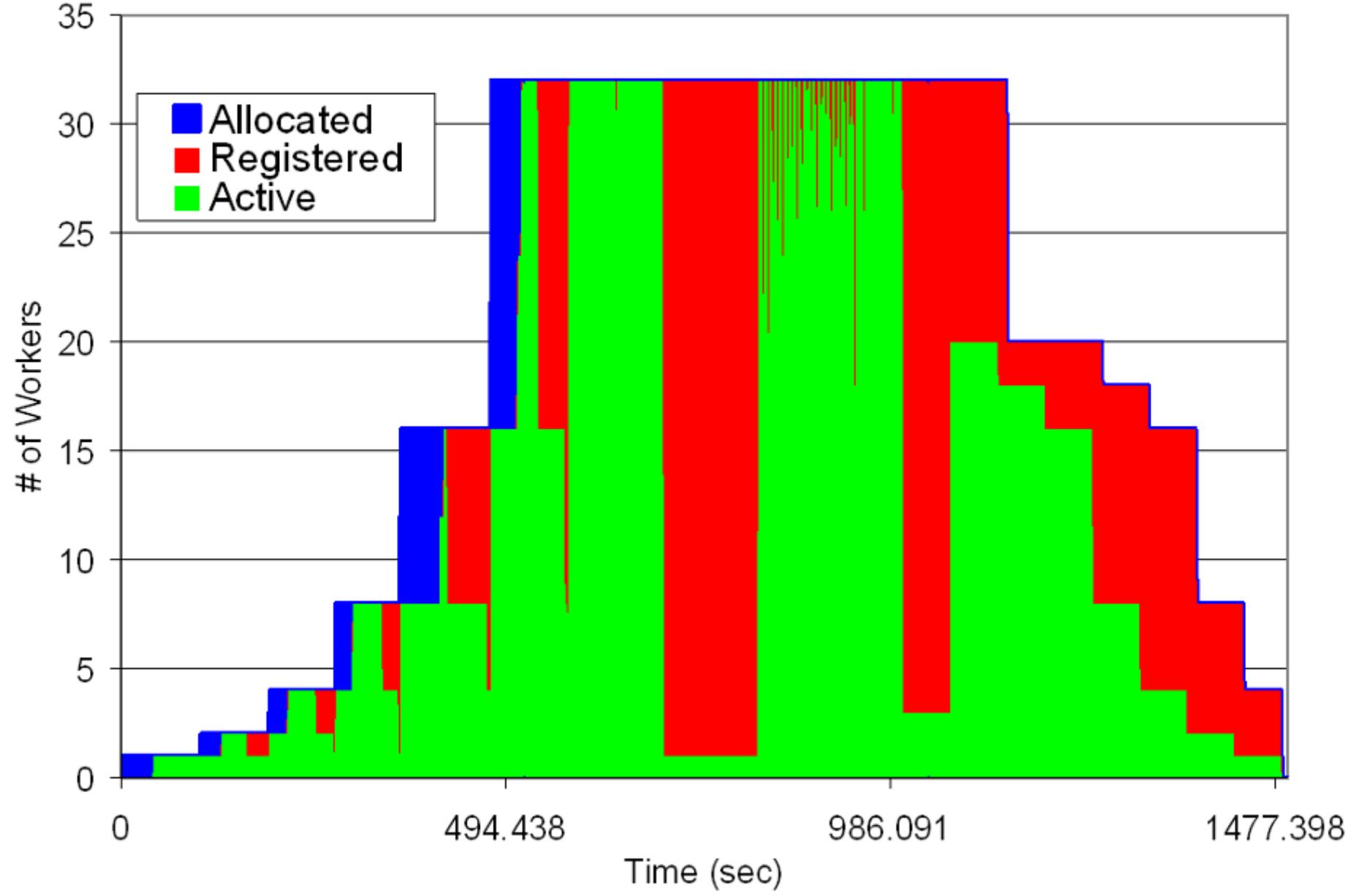


Joint work with Ioan Raicu & Alex Szalay

Dynamic Provisioning with Falkon: Release after 15 Seconds Idle

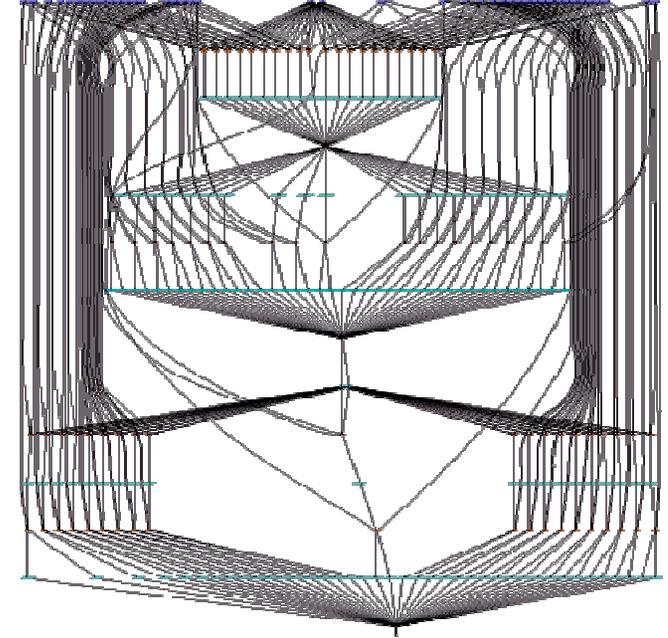
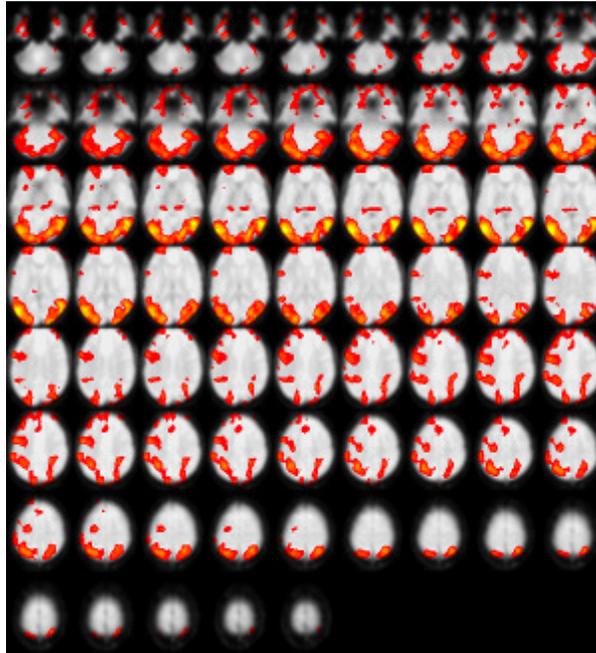
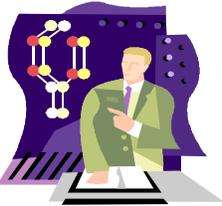


Dynamic Provisioning with Falkon: Release after 180 Seconds Idle





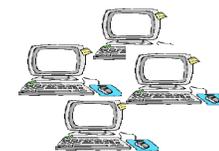
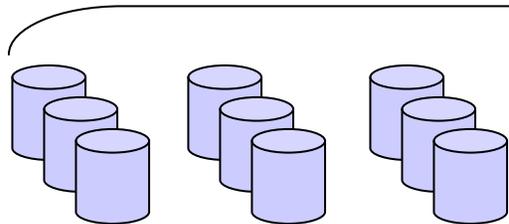
Building Scalable Service Implementations



Functional MRI



Ben Clifford,
Mihael Hatigan,
Mike Wilde,
Yong Zhao





AIRSN Program Definition

```
(Run snr) functional ( Run r, NormAnat a,  
                      Air shrink ) {
```

```
Run yroRun = reorientRun( r , "y" );
```

```
Run roRun = reorientRun( yroRun , "x" );
```

```
Volume std = roRun[0];
```

```
Run rndr = random_select( roRun, 0.1 );
```

```
AirVector rndAirVec = align_linearRun( rndr, std, 12, 1000, 1000, "81 3 3" );
```

```
Run reslicedRndr = resliceRun( rndr, rndAirVec, "o", "k" );
```

```
Volume meanRand = softmean( reslicedRndr, "y", "null" );
```

```
Air mnQAAir = alignlinear( a.nHires, meanRand, 6, 1000, 4, "81 3 3" );
```

```
Warp boldNormWarp = combinewarp( shrink, a.aWarp, mnQAAir );
```

```
Run nr = reslice_warp_run( boldNormWarp, roRun );
```

```
Volume meanAll = strictmean( nr, "y", "null" )
```

```
Volume boldMask = binarize( meanAll, "y" );
```

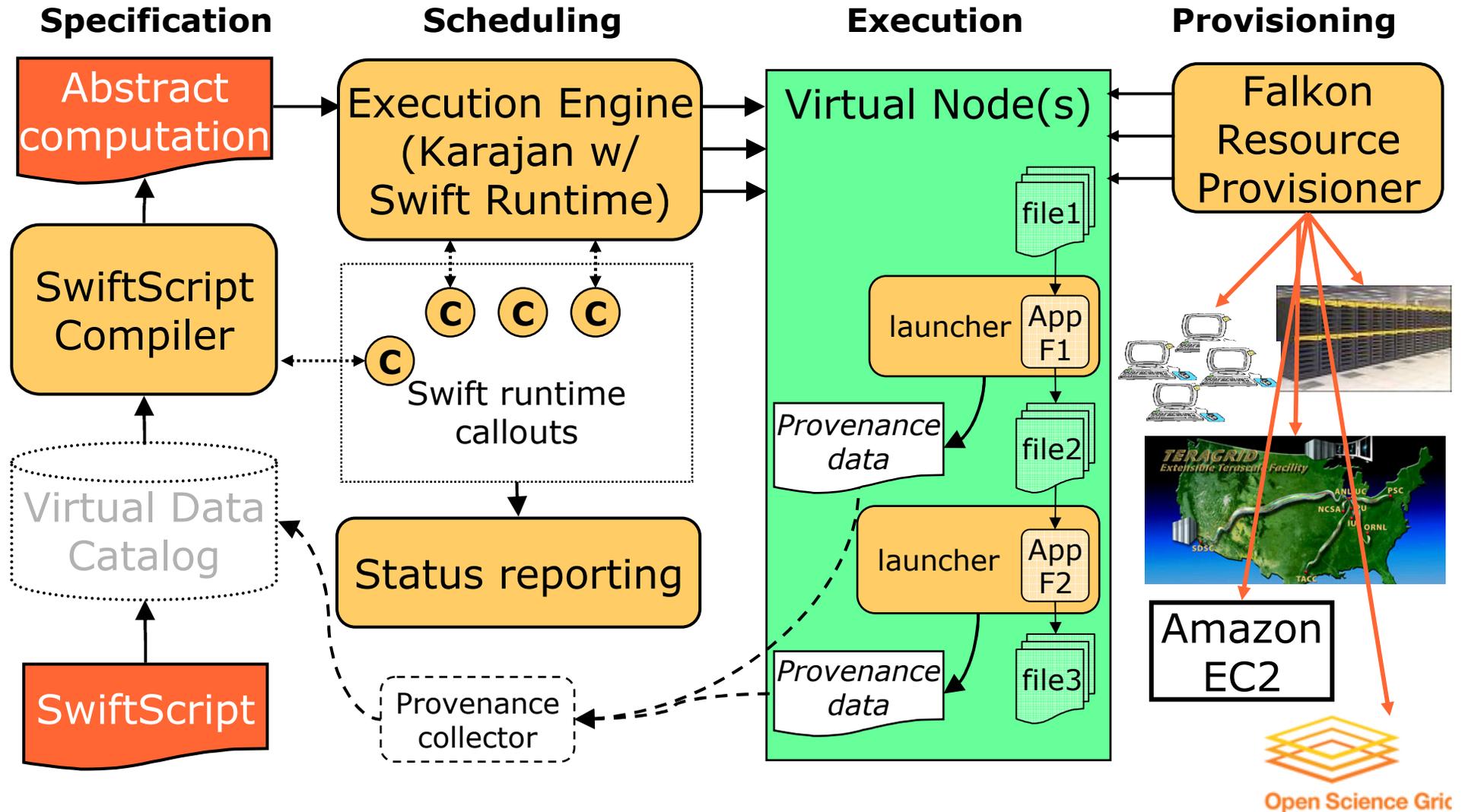
```
snr = gsmoothRun( nr, boldMask, "6 6 6" );
```

```
}
```

```
(Run or) reorientRun (Run ir,  
                      string direction) {  
  foreach Volume iv, i in ir.v {  
    or.v[i] = reorient(iv, direction);  
  }  
}
```



Dynamic Provisioning: Swift Architecture





Services for Science

- They're **new**
 - ◆ A new approach to communicating
 - ◆ A (not-so new) approach to structuring systems
- They're **real**
 - ◆ Excellent infrastructure and tools (Globus, Introduce, gRAVI, Taverna, Swift, etc.)
 - ◆ Substantial numbers of services out there
- They're **challenging**
 - ◆ Sociology: incentives, rewards
 - ◆ Infrastructure: hosting
 - ◆ Provenance: justifying "results"
 - ◆ Scaling: services, requests