



globus online



Computation Institute

Cloud-based services for (reproducible) science

Ian Foster

Computation Institute

University of Chicago and Argonne National Laboratory



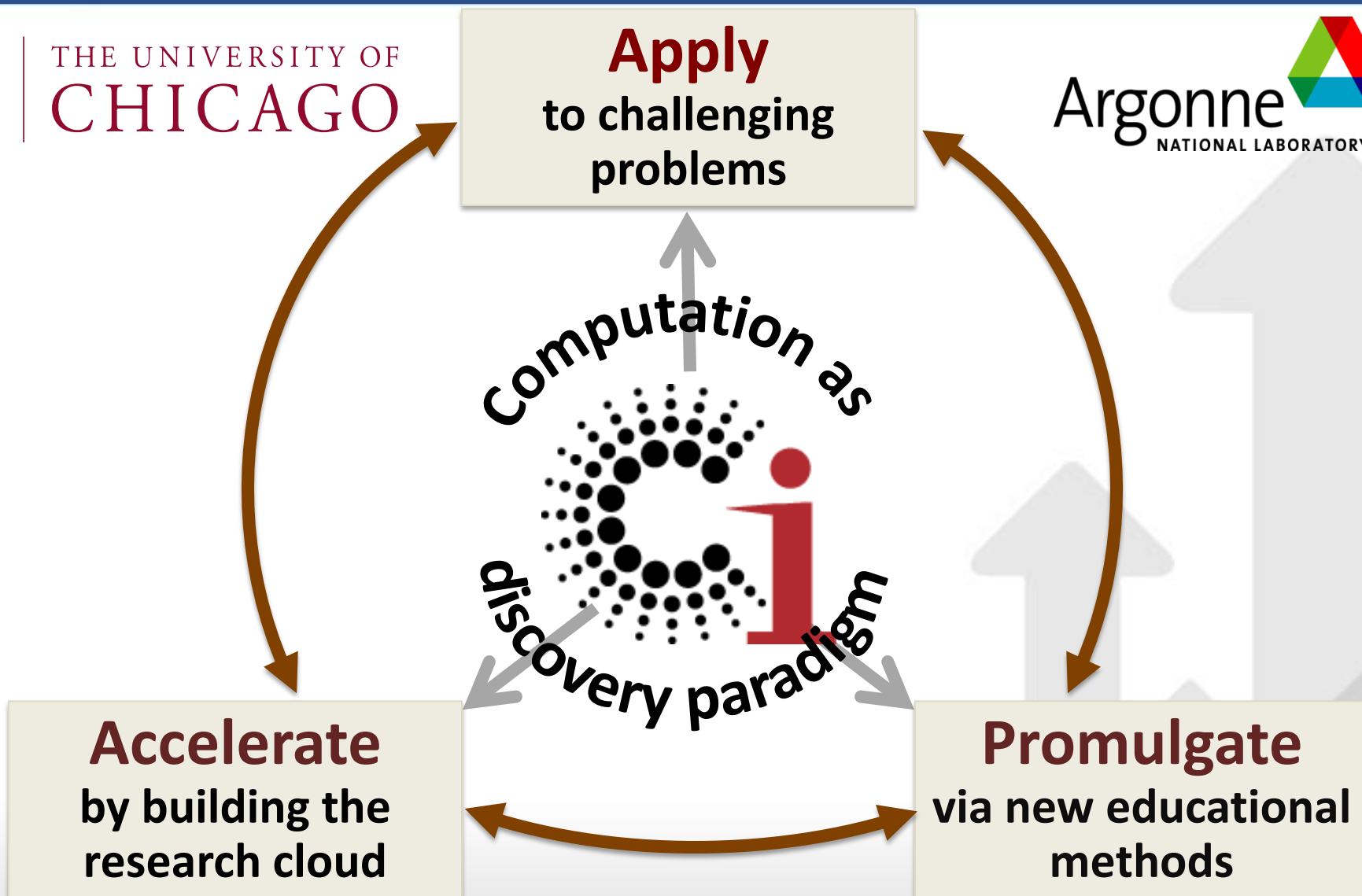
Computation Institute (CI)



THE UNIVERSITY OF
CHICAGO



Argonne
NATIONAL LABORATORY





Cloud Layers

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

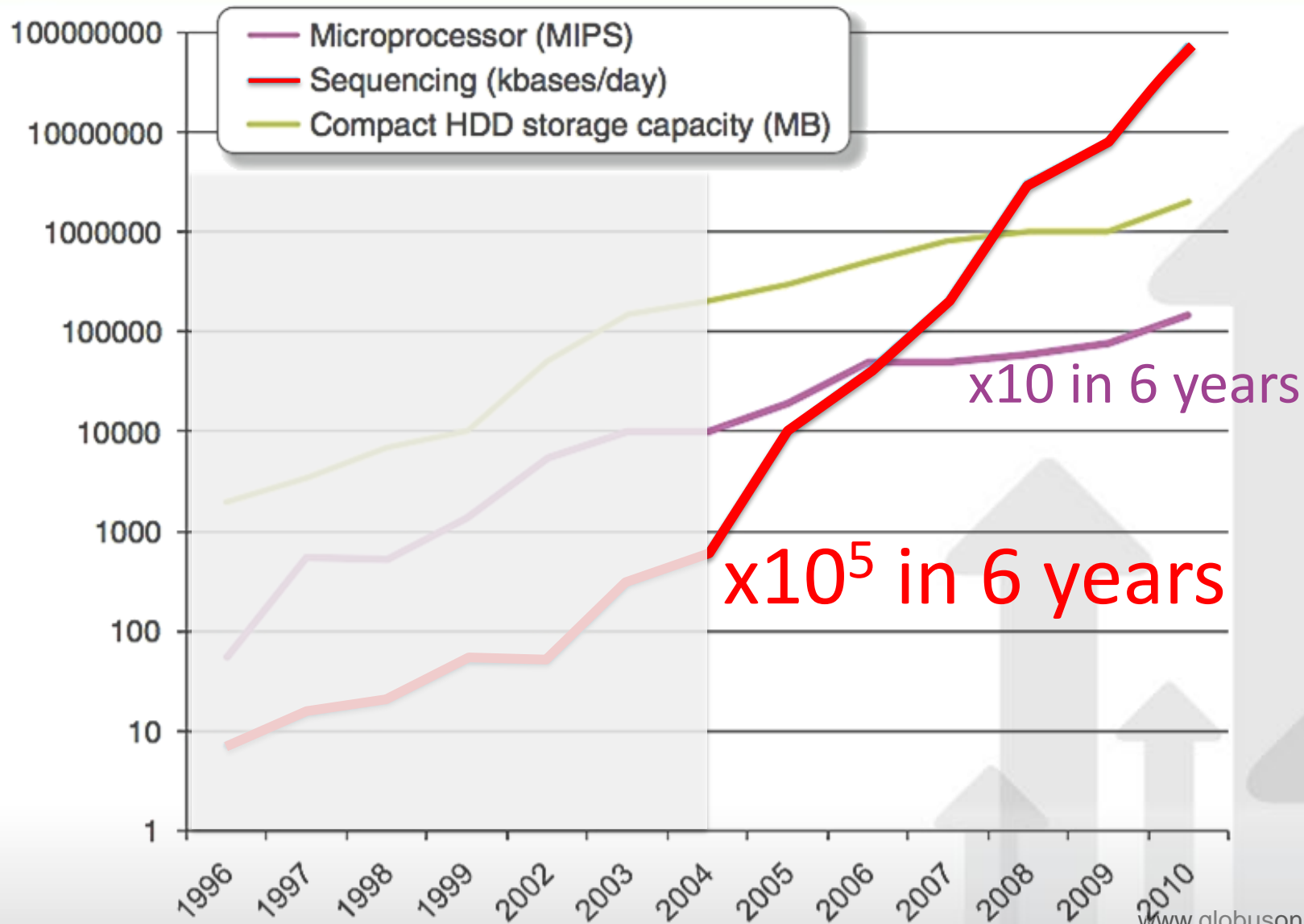
Infrastructure-as-a-Service (IaaS)

The data deluge



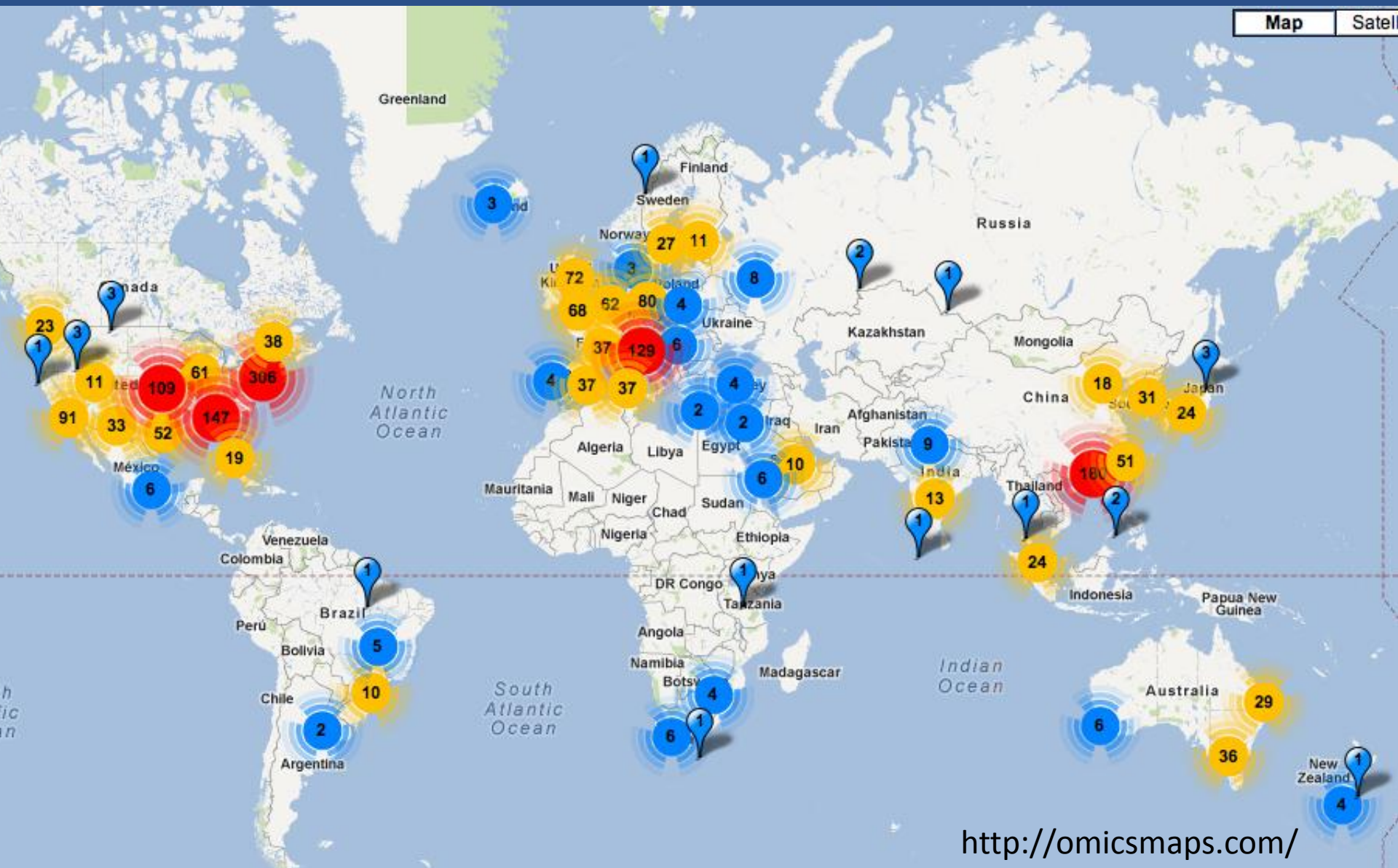


The data deluge in biology



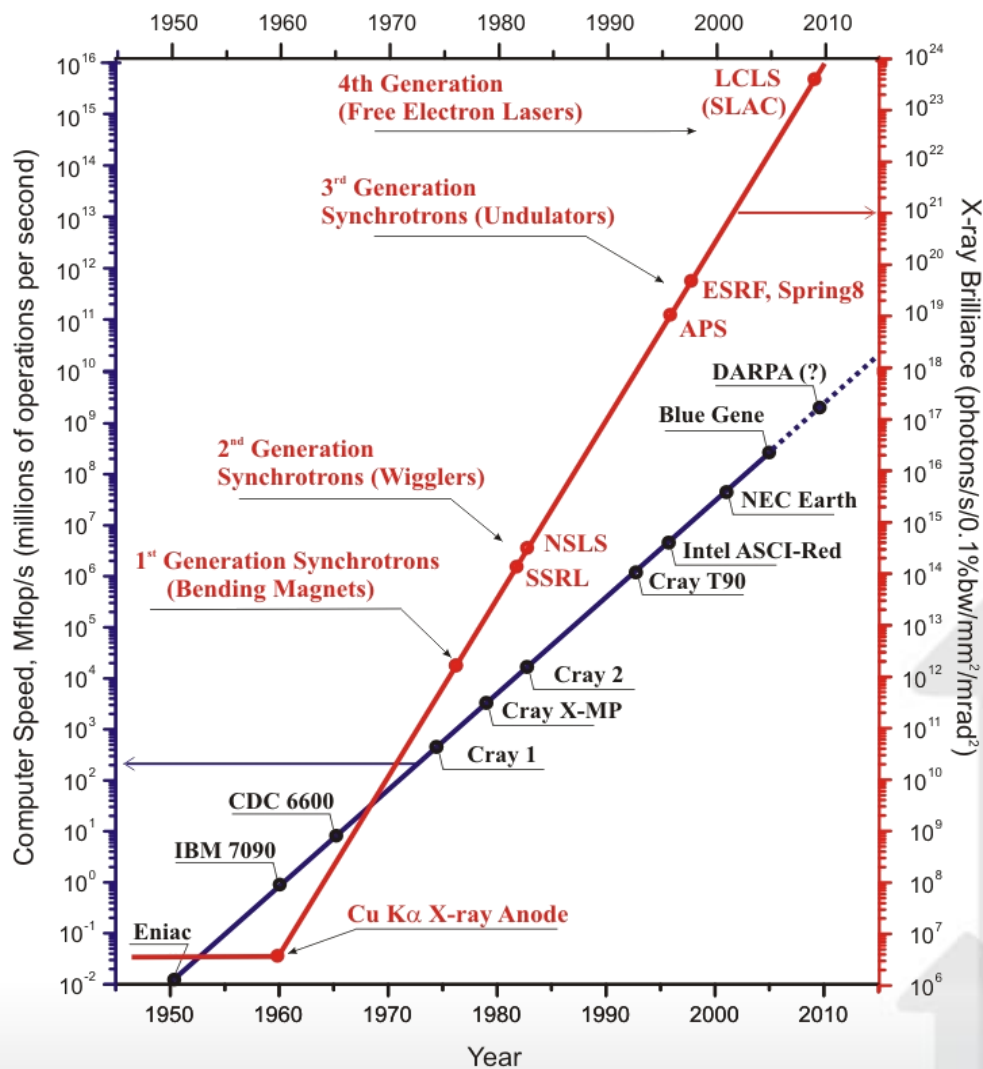


Number of sequencing machines





Moore's Law for X-ray sources



12 orders of magnitude in 6 decades

18 orders of magnitude in 5 decades!

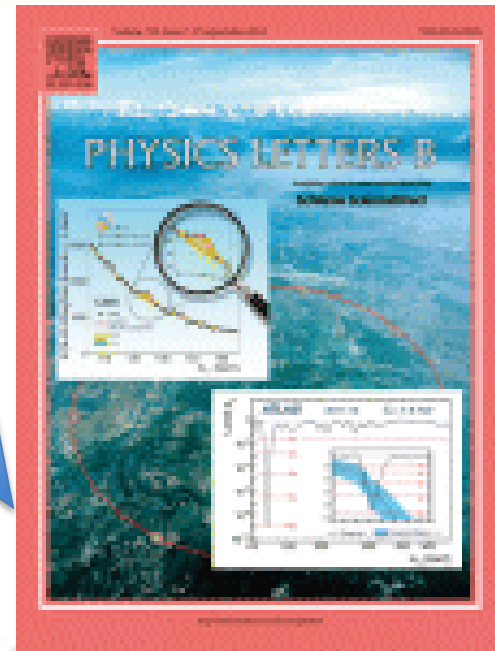


Big science has been successful

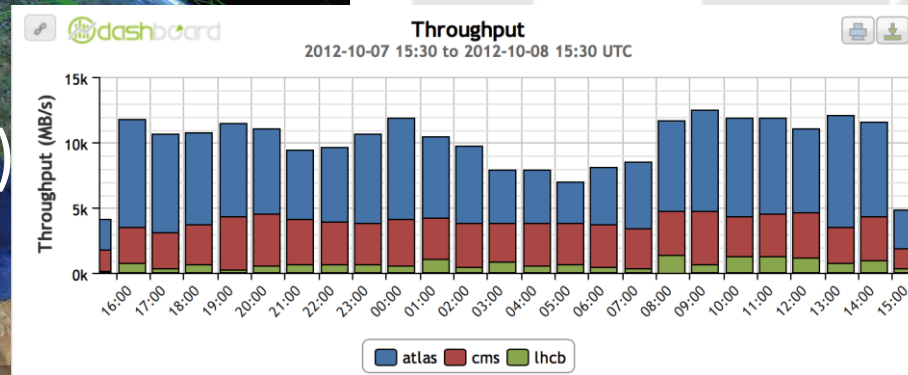
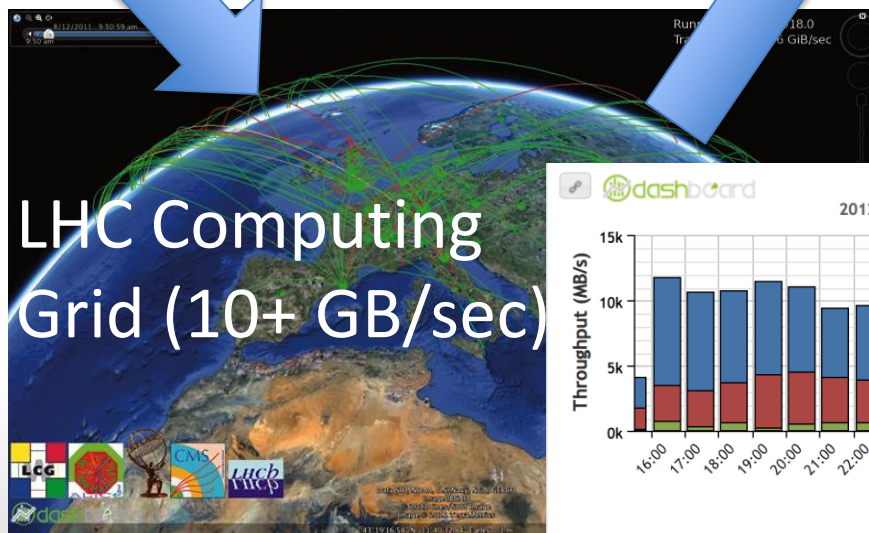
Large Hadron Collider



Globus GridFTP
300K files/day
0.3 PB/day
... just for LHC



15 PB/year
173 TB/day
500 MB/sec

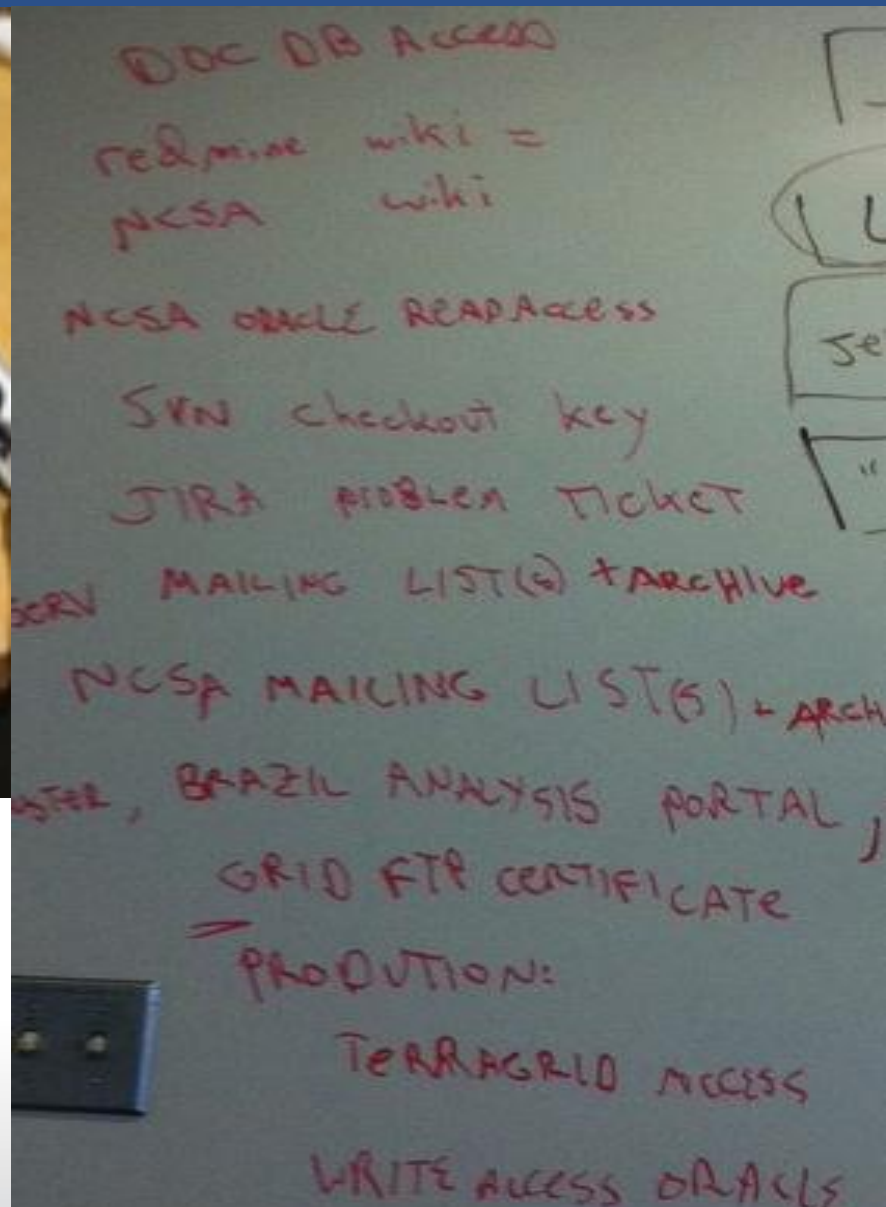




But small and medium science is suffering



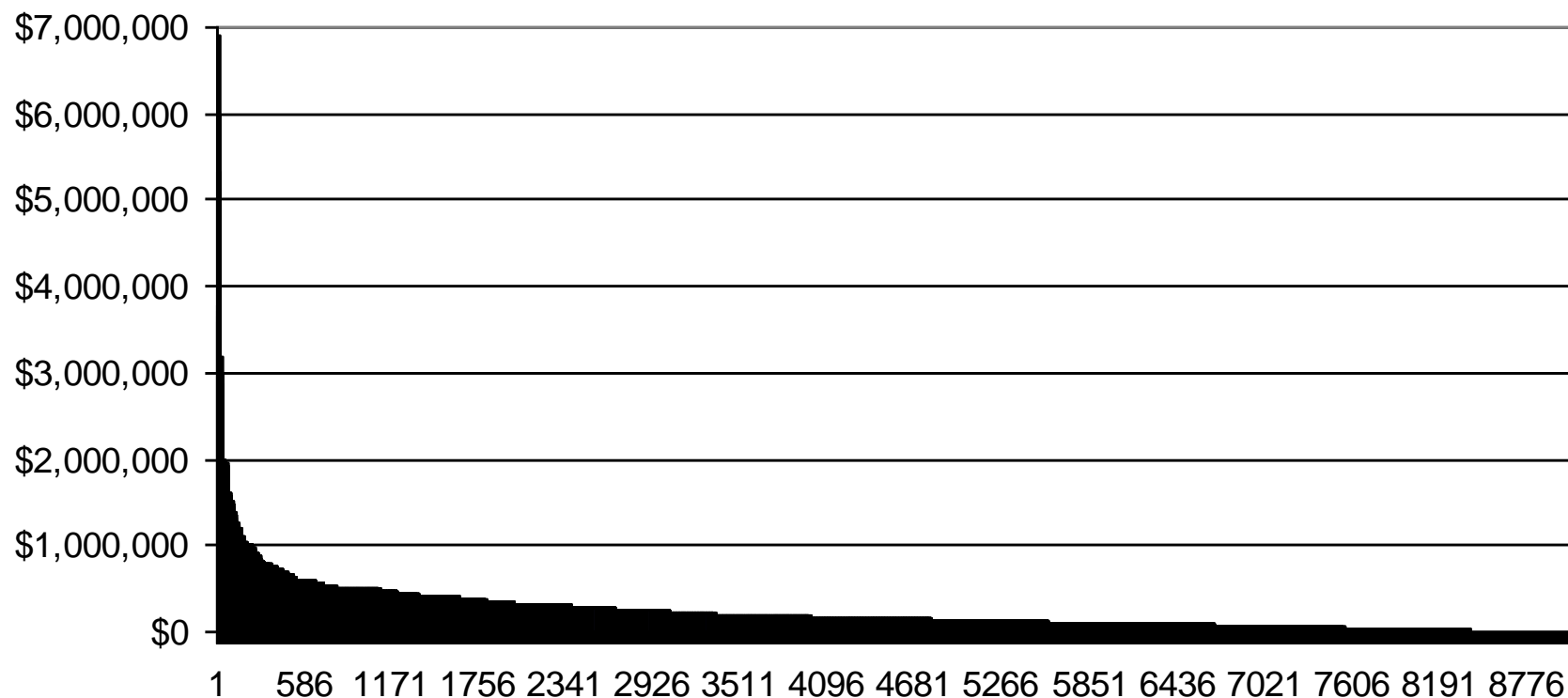
- Data deluge
- Ad-hoc solutions
- Inadequate software, hardware & IT staff





Dark data in the long tail of science

Awarded Amount 2007



NSF grant awards, 2007 (Bryan Heidorn)



A crisis that demands new approaches

- We have **exceptional infrastructure for the 1%** (e.g., supercomputers, Large Hadron Collider, ...)
- But **not for the 99%** (e.g., the vast majority of the 1.8M publicly funded researchers in the EU)

We need new approaches to providing research cyberinfrastructure, that:

- Provide advanced functionality
- Reduce barriers to entry
- Are cheaper
- Are sustainable



Run a research project from a coffee shop?

STARBUCKS COFFEE





Rethinking how we provide **research IT**

Accelerate discovery and innovation worldwide by providing **research IT as a service**



Leverage the **cloud** to

- provide millions of researchers with unprecedented access to powerful tools;
- enable a massive shortening of cycle times in time-consuming research processes; and
- reduce research IT costs dramatically via economies of scale



Time-consuming Tasks in Research

- Run experiments
- Collect data
- Manage data
- Communicate with colleagues
- Publish papers

• *“Civilization advances by extending the number of important operations which we can perform without thinking of them”*

—Alfred North Whitehead , 1911

- Run simulations
- Compare experiment with simulation
- Search the literature
- Order supplies
- Write proposals
- Write reports
- ...



Addressing the big data challenge

Simulation



Telescope



In millions of labs worldwide, researchers struggle with massive data, advanced software, complex protocols, burdensome reporting



Next-gen genome sequencer

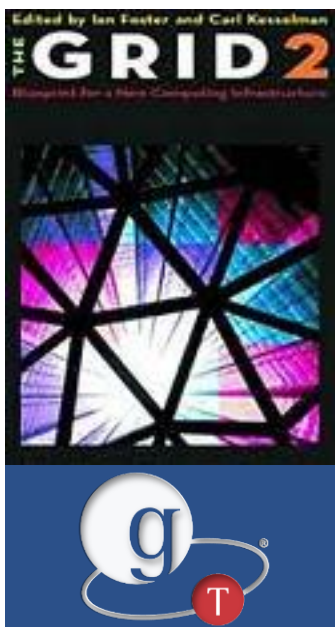


Accelerate discovery and innovation by outsourcing difficult tasks



Grid meets Cloud

Millions of researchers worldwide need advanced IT to tackle important and urgent problems



Reliable, high-performance, secure file transfer.
Move files fast. No IT required.

+ WATCH A VIDEO

Globus Online in a nutshell



> GET STARTED

Sign up and get moving

6,836,755,450 MB
TRANSFERRED



Why Use Globus Online?
See how easy file transfer can be



For HPC Resource Owners
Enable Globus Online for your users



For Developers
Integrate with Globus Online



What is Globus Online?

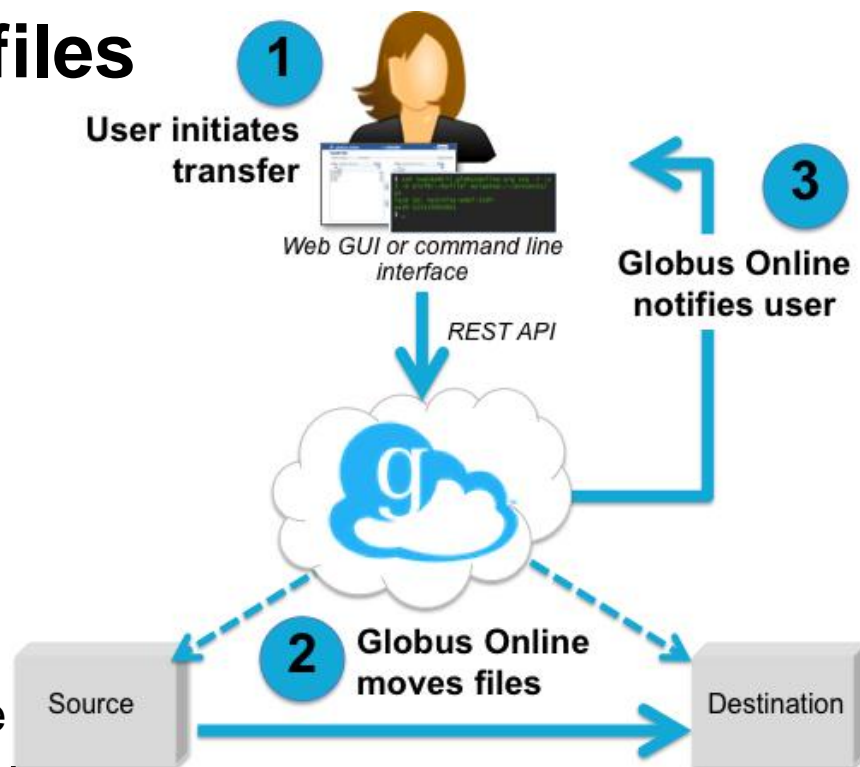
- **Transfer and synchronize files**

- Easy “fire-and-forget” transfers
- Automatic fault recovery
- High performance
- Across multiple security domains

- **Minimize IT costs**

- Software as a Service (SaaS)
 - No client software installation
 - New features automatically available
- Consolidated support & troubleshooting
- Simple endpoint installation with Globus Connect and GridFTP

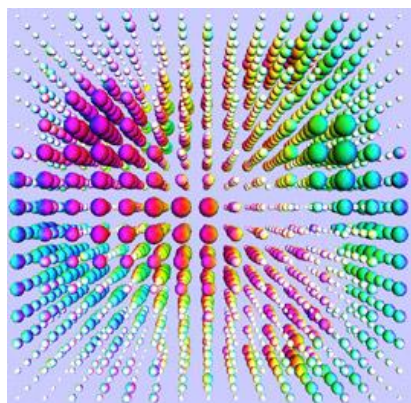
- **Recommended by XSEDE, Blue Water, NERSC, ALCF, ESnet, many Universities**





Case Study: Lattice QCD

NICS



- **Fast:** Reduced transfer times
- **Easy:** Fire-and-forget transfers
 - Automated retry
 - No file pre-staging
 - No complex infrastructure
 - Convenient CLI and GUI interfaces

Indiana University
researcher moved
~6 TB from Tennessee
to Texas in 2 days

“Globus Online frees up my time to do more creative work than typing scp commands or devising scripts to initiate and monitor progress to move many files.”

“I moved 100 7.3 GB files tonight in about 1.5 hours. I am very impressed. I also like the new commands and help system.”

TACC





Case Study: Enabling Users @ NERSC

- **Challenge**
 - “We need to provide web-based ways to accomplish computing tasks – it’s what our scientists expect. And it will make them more productive.”
- **Solution**
 - Globus Online endpoints maintained by NERSC
 - Globus = recommended transfer method
- **Benefits for NERSC users**
 - Drag and drop archiving
 - Easy to use
 - Users can focus on their research (not on IT)
- **Benefits for NERSC**
 - Operations and support outsourced to Globus
 - Fast and easy to make endpoints available
 - Automated authentication
 - Reliable performance and support



“Fantastic! I have already started using Globus Connect to transfer data, and it only took me 5 minutes to set up. Thank you!” – NERSC user



Hopper, Franklin and HPSS are among the NERSC resources leveraged by Globus Online.



[SIGN IN](#) [SIGN UP](#)

Reliable, high-performance, secure file transfer by Globus Online.

Blue Waters has partnered with the Globus Online file transfer service.

You may access this service by entering your Blue Waters username and password.

NOTE - If you are accessing this file transfer service for the first time, you will be asked to link your Blue Waters account to a Globus Online account (if you don't have a Globus Online account you'll be able to create one).

Sign In

[Use Your NCSA Blue Waters login](#)

[alternate login](#)


You will now jump to NCSA Blue Waters's authentication page.

[Proceed](#)



Partners



globus online  Manage Data | News & Events | About | Support | Log In | Sign Up

Sign In [Sign Up with Globus Online](#)

Use Your GlobusOnline login [alternate login](#)

Username

Password

I have read the [Privacy Policy](#) and by logging in I agree to the terms therein. Specifically, I understand and agree that this service may store data on servers worldwide, including the University of Chicago through Globus Online's central servers in the United States.

[Sign In](#) [Forgot Password?](#)

- EU-friendly identity providers
- EU relevant endpoints
- Integrated support with EGI & IGE
- EU Cookie Law compliant
- Consent to store log data in US



Cookies are used on this site

This website uses cookies so that we can improve our website and enhance your user experience. To find out more about cookies and their use by us, please see the "Cookies" section in our [Privacy Policy](#).

By clicking the Enter Site button, I accept cookies from this website.

[Enter Site](#)



Why SaaS?

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

Infrastructure-as-a-Service (IaaS)

- **Deliver advanced functionality that:**
 - Requires no user software installation or operation
 - Minimal IT proficiency required
 - Can be cheaply and incrementally adopted
 - Usage-based subscription pricing; no big up-front costs
 - Consolidates troubleshooting and support
 - An expert group can proactively detect and correct problems
 - Utilizes an efficient software delivery lifecycle
 - Updates developed, tested and deployed quickly
- **Dominates commercial & consumer markets**
 - What about the research market?



Software as a Service (SaaS) vs. traditional software delivery

- **SaaS changes assumptions and approach throughout the software lifecycle**
 - Architecture and design
 - Designed for specific environment (e.g., AWS)
 - Software development
 - No porting or cross platform testing. Focus on functionality.
 - Operations
 - Continuous update cycle. Nobody else will operate.
 - Focus on availability, automation, monitoring
 - Support
 - Tightly integrated with operations
- **We are delivering a *service*, not *software***



Potential economies of scale

Small laboratories

- PI, postdoc, technician, grad students
- Estimate 10,000 across US research community
- Average ill-spent/unmet need of 0.5 FTE/lab?

+ Medium-scale projects

- Multiple PIs, a few software engineers
- Estimate 1,000 across US research community
- Average ill-spent/unmet need of 3 FTE/project?

= Total 8,000 FTE: at ~\$100K/FTE => \$800M/yr

(If we could even find 8,000 skilled people)

Plus computers, storage, opportunity costs, ... www.globusonline.org



SaaS and reproducible publications

- **Hosted services can contribute to relieving author burden**
- **Store data, code, and workflows**
- **Maintain bindings among data, code, executions, workflows, provenance**
- **Etc.**



Other innovative science SaaS projects

my experiment beta

myExperiment makes it really easy to find, use and share scientific workflows and other files, and to build communities.

Use myExperiment to...

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

Explore



- About myExperiment
- Join the Mailing List
- Give us Feedback For Developers
- The myGrid Project
- Taverna Workflow Workbench
- The BioCatalogue Project
- myExperiment Publications

Register

or Login:

Username or Email

Password

Remember me:

Or use OpenID:

(eg. name.myopenid.com)

Login

Forgot Password?

myExperiment has over 1250 users, 100 groups, 490 workflows, 130 files and 40 packs



Other innovative science SaaS projects

my experiment

Use myExperiment

- Find Workflows
- Find Files
- Share Your Work Files
- Create and Find Items
- Create and Join
- Find People and Friends
- Send Messages
- Get Feedback
- Tag and Rate
- Write Reviews and Comments
- Build your

nanoHUB.org
an ncn project

ONLINE SIMULATION AND MORE FOR NANOTECHNOLOGY

Home My HUB Resources Members Explore About Support

NANO HUB

FUNDAMENTALS OF NANO ELECTRONICS
Online Course Spring 2012

Learn More >

1 2 3 > || >

SIMULATE with over 160 tools for nanoelectronics, nanophotonics and more >

RESEARCH & COLLABORATE via groups, question board and more >

TEACH & LEARN with tool-powered curricula, courses, seminars and more >

SHARE & PUBLISH tools and research through our easy upload process

RESOURCES

Search

Popular Tags: nanoelectronics course lecture material science Illinois nano/bio nanotransistors research seminar devices nanophotonics quantum transport tutorial transistors molecular electronics nano electro-mechanical systems NEGF carbon nanotubes nanomedicine education/outreach UIUC band structure ABACUS atomic force microscopy quantum dots MOSFET nanowires More tags >

FEATURED

MIT Atomic Scale Modeling Toolkit : Tools for Atomic Scale Modeling - in Tools

Nanotechnology: Considerations for Facility Design - in Online Presentations

Greg Lush, University of Texas at El Paso - Contributions: 30

Topics For Introductory Materials Classes - in Topics

Leap. Drag here to tag and file.



Other innovative science SaaS projects

myExperiment

myExperiment makes it really easy to find, use and share scientific workflows and other files, and to build communities.

nanoHUB.org
an ncn project

ONLINE SIMULATION AND MORE FOR NANOTECHNOLOGY

Home My HUB Resources Members Explore About Support

MG-RAST

metagenomics analysis server



LOGIN REGISTER PASSWORD FORGOT?

login



Browse Metagenomes

search for metagenomes



Register



Contact



Help



Upload



News

About

MG-RAST (the Metagenomics RAST) server is an automated analysis platform for metagenomes providing quantitative insights into microbial populations based on sequence data.

# of metagenomes	35,586
# base pairs	9.24 Tbp
# of sequences	85.21 billion
# of public metagenomes	7,167

The server provides web based upload, quality control, automated annotation and analysis for samples up to 10GBp. Comparison between large numbers of samples is enabled via pre-computed abundance profiles. MG-RAST was launched in 2007 and has over 5000 registered users and 35,586 data sets. The current server version is 3.1.2.

Updates

MG-RAST Version 3.1.2 released



Other innovative science SaaS projects

myExperiment makes it really easy to find, use and share workflows and other files, and to build communities.

nanoHUB.org
an ncn project

ONLINE SIMULATION AND MORE
FOR NANOTECHNOLOGY

Home My HUB Resources Members Explore About Support

my exp SQLShare - View Query

https://sqlshare.cac.washington.edu/sqlshare/#s=query/charlon/United%20States%20Wind%20Energy%20Potential.csv

SQLSHARE

United States Wind E... x

Logged in: charlon@washington.edu

United States Wind Energy Potential Test.csv Last modified: Apr 19, 2011 3:15 PM chartest

[Click here to add a description](#)

```
SELECT * FROM [chartest].[table_United States Wind Energy Potential.csv]
```

Edit query Download Query dataset More actions

DATASET PREVIEW (Rows 1 - 48 of 48)

<< first < prev 1 2 3 next > last >>

State	Total (km)	Excluded (km)	Available (km)	Available % of State	% of Total Windy Land Excluded	Potential Installed Ca
Alabama	80.36	56.72	23.64	0.017686450652558%	70.5823792931807%	118.2
Arizona	4544.96	2364.14	2180.82	0.73866427637171%	52.016739421249%	10904.1
Arkansas	4663.24	2823.18	1840.06	1.33593638826854%	60.5411688010911%	9200.3
California	26901.28	20079.24	6822.04	1.66671846119056%	74.6404632047248%	34110.2
Colorado	95830.36	18386.46	77443.9	28.7253372905834%	19.1864665853285%	387219.5
Connecticut	31.36	26.06	5.3	0.041365143388757%	83.0994897959184%	26.5
Delaware	36.56	34.66	1.9	0.03739686338729%	94.8030634573304%	9.5
Florida	9.56	9.48	0.08	0.000054815635831%	99.163179916318%	0.4
Georgia	281.28	255.26	26.02	0.01708279443529%	90.7494311717861%	130.1
Idaho	13420.4	9805.28	3615.12	1.67025347276445%	73.0625018628357%	18075.6
Illinois	70763.56	20787.14	49976.42	34.2484872158604%	29.3754864791992%	249882.1
Indiana	46255.24	16800.74	29454.5	31.6228028072802%	35.008820710617%	148227.5




Other innovative science SaaS projects

myExperiment makes it really easy to find, use and share other files, and to build communities.

nanoHUB.org
an ncn project

ONLINE SIMULATION AND MORE FOR NANOTECHNOLOGY

mye iPlant Collaborative Web Port...
www.iplantcollaborative.org

 **iPlant Collaborative™** Empowering A New Plant Biology

[About](#) [Contact Us](#) [Login or Register](#) [Feedback](#)

Challenge

iPlant develops a *vision* for plant science cyberinfrastructure

CHALLENGE

iPlant Genotype to Phenotype (iPG2P)
Mapping the links between genotypes and phenotypes



iPlant Tree of Life (IPToL)
Understanding the



DISCOVER

Discovery Environment
Access iPlant tools through a single user-friendly interface
[MORE...](#)



DNA Subway
An educator-tailored interface for bringing



LEARN

Upcoming Events

- iPlant Tools and Services Workshop @ University of Missouri, Columbia (FULL WORKSHOP)**
September 24 2012 - September 25 2012
- iPlant Genomics in Education Barcoding Workshop @ SACNAS, Seattle**
October 10 2012
- iPlant Genomics in Education Workshop @ Schoolcraft College - Detroit, MI**

CONNECT

People at iPlant

Community driven science



Idaho	13420.4	9805.28	3615.12	1.67025347276445%	73.0625018628357%	18075.6
Illinois	70763.56	20787.14	49976.42	34.2484872158604%	29.3754864791992%	249882.1
Indiana	46265.24	16600.74	20645.5	21.6228026072802%	25.0088820740847%	148227.5



Other innovative science SaaS projects

The concept

As simple as 1,2,3

1. A researcher has an **idea**.
2. The researcher writes a **paper** based on this idea.
3. Using RunMyCode, the researcher creates a **companion website** associated with this paper. The companion website allows people to implement the methodology presented in the paper.

[Learn more >>](#)

[About](#) [Concept](#) [Purpose](#)

[Create your own companion website >>](#)

nanoHUB.org
an ncn project
ONLINE SIMULATION AND MORE FOR NANOTECHNOLOGY

Home My HUB Resources Members

Login or Register Feedback

mye

Home
First visit?
Our offering
Submit your code

Search by themes
Advanced search

Help/FAQ
Our partners
The team
Contact us

Your data
All data
Favorite
Recent
Share

iPlant Tree of Life (iPToL)
Understanding the

Idaho	13420.4	9805.28	3615.12	1.67025347276445%	73.062501000
Illinois	70763.56	20787.14	49976.42	34.2484872158604%	29.3754864791992%
Indiana	46255.24	16600.74	29654.5	24.6238006702803%	25.0088820740847%



Towards “research IT as a service”

Research Data Management-as-a-Service



Globus
Transfer

Globus
Storage

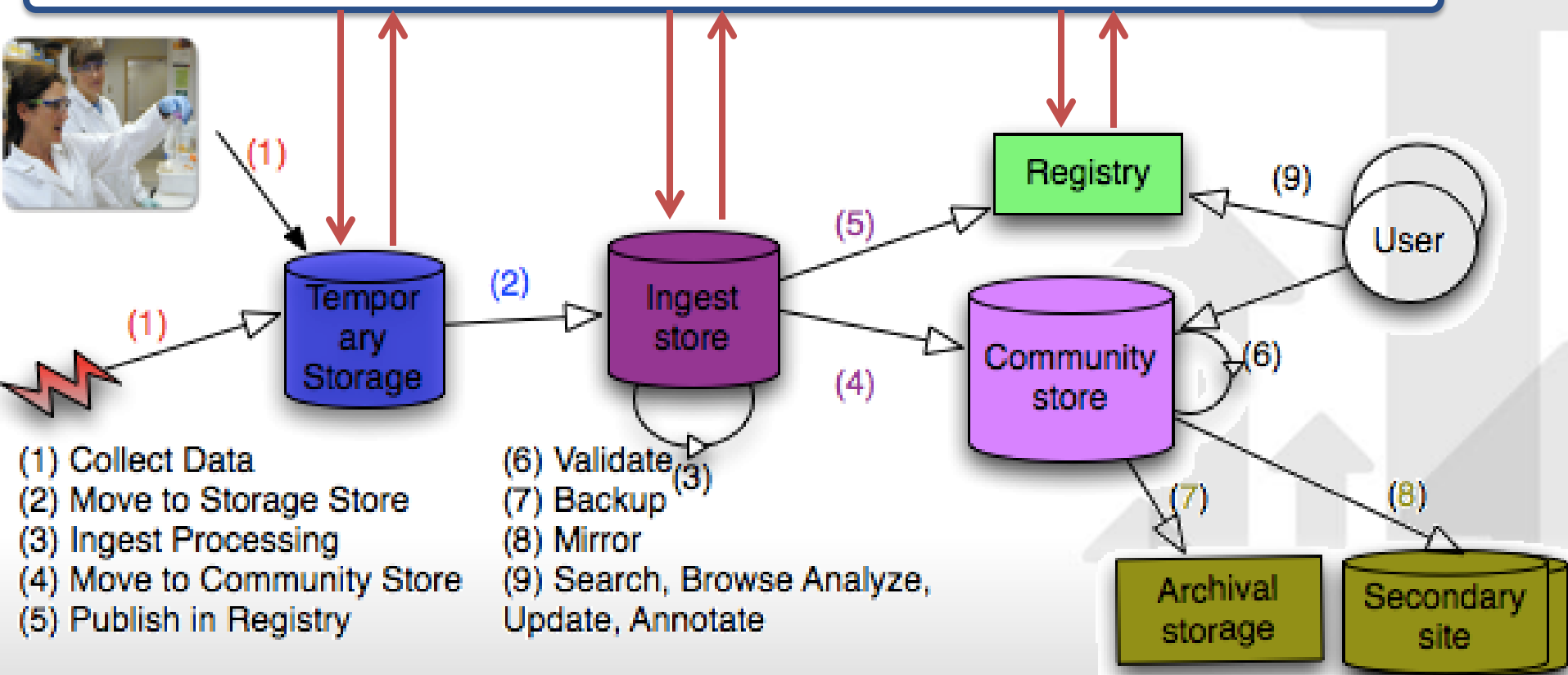
Globus
Catalog

Globus
Collaborate

...SaaS

Globus Integrate

...PaaS





PaaS for Research

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

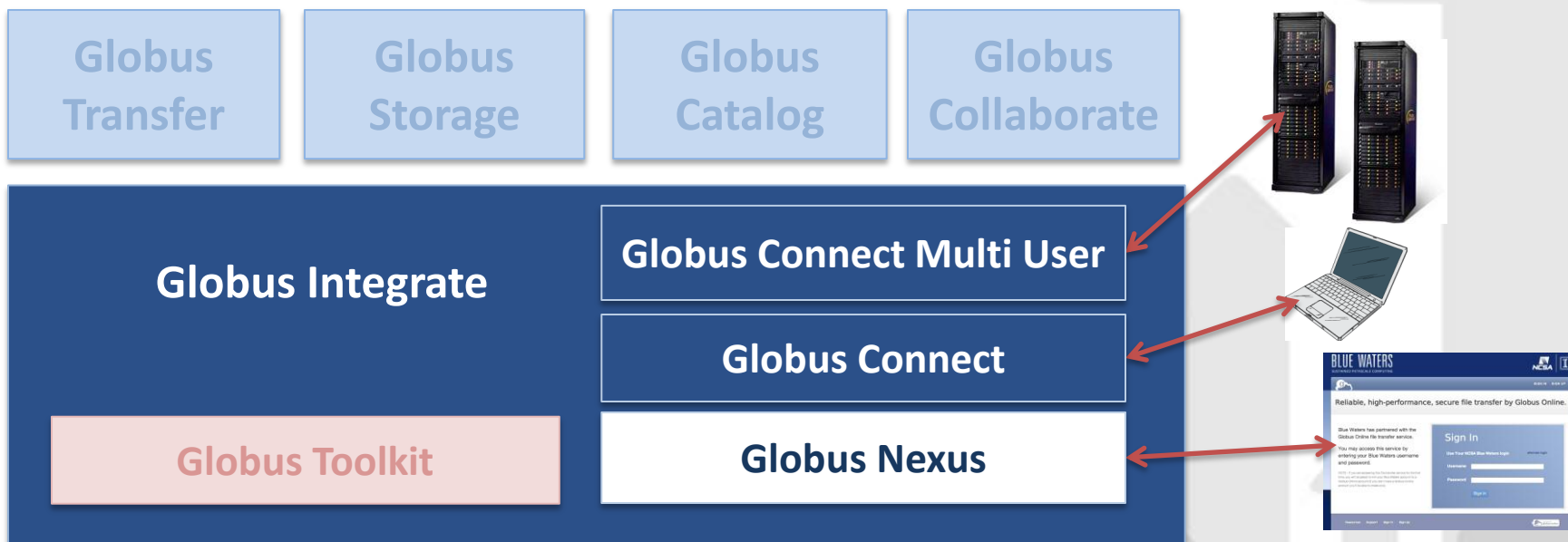
Infrastructure-as-a-Service (IaaS)

- **No one SaaS provider can deliver it all**
- **Must create ecosystem that:**
 - Allows any SaaS provider to easily participate
 - Dramatically reduces the cost of creating and operating services within the ecosystem
 - Provides seamless user experience across services
 - Agnostic to / works across any cloud IaaS provider
- **Ecosystem requires Platform as a Service**
 - Target the unique needs of the research community



Globus Integrate: For when you want to...

- Integrate with the **Globus research cloud ecosystem**
- **Write programs that leverage:**
 - (federated) user identities, profiles, groups
 - data, compute and collaboration



... via **REST APIs** and **command line programs**



Dark Energy Survey – Leveraging transfer

- **Every night, they receive 100,000 files in Illinois**
- **They transmit files to Texas for analysis ... then move results back to Illinois ... and make them available to users**
- **Process must be reliable, routine, and efficient**
- **The cyberinfrastructure team is not large!**

Blanco 4m on Cerro Tololo



Image credit: Roger Smith/NOAO/AURA/NSF



Earth System Grid – Portal integration

ESGF Earth System Grid Federation

OAK RIDGE National Laboratory

Home Search Browse Account Login Admin

Current Selections

- (x) Institute: GFDL

Temporal Search
Geospatial Search
Turn on Distributed Search

Search

(press ESC to close suggestions) < 1 > displaying 1 to 1 of 1 search results

Search Categories

- project (1)
- Institute (1)
- model (1)

Results Data Cart

pcmdi.jpcc4.GFDL.gfdl_cm2_0.picntrl.mon.land.run1 Expand WGET | Globus Online | Remove

ESGF Earth System Grid Federation

Argonne NATIONAL LABORATORY

Home Search Account Logout

Globus Online Data Transfer

Globus Online Transfer: Step 2 of 3

Please choose where you'd like to move the data:

Destination Endpoint

Target Endpoint Path Location

next

neillm#ci-vm
neillm#ci-vm
neillm#go-esg
neillm#laptop
neillm#pcmdi
neillm#testep
alcf#dtn
alcf#dtn_test
aps#clutch
arm#bds2-vm1
ballen#oathabe
ballen#oathathena
ballen#oathbigred
ballen#oathblacklight
ballen#oathcondor
ballen#oathdash
ballen#oathdcwan
ballen#oathfrost
ballen#oathgpfswan
ballen#oathhps
ballen#oathkraken

Model Params sent

DatasetName: anl.cssef.homme

GO Username: neillm

Source Myproxy Username: neillm

Source Myproxy Password: *****

User Certificate: /tmp/x509up_5108938099718227095.pem

Endpoints: neillm#ci-vm neillm#go-esg neillm#laptop neillm#pcmdi neillm#testep alcf#dtn alcf#dtn_test aps#clutch arm#bds2-vm1 ballen#oathabe ballen#oathathena ballen#oathbigred ballen#oathblacklight

- Outsource data transfer to Globus
 - Data download to user machine from search
 - Data transfer to another server by user
 - Replication of data between sites by administrator
- No ESGF client software needed

ESGF Earth System Grid Federation

Argonne NATIONAL LABORATORY

Home Search Account Logout

Globus Online Data Transfer

GlobusOnline Transfer Status

Your Globus Online Transfer has been started!

The transfer has been accepted and a task has been created and queued for execution.

Globus Online TaskID: f7e37822-fa70-11e0-bc85-1231381a212f

[View Globus Online Transfer](#)

Model Params sent

Error:



KBase – Leveraging the full platform

← → ↻ kbase.science.energy.gov



Search this site



[Home](#)

[About](#) ▾

[News](#) ▾

[Developer Zone](#) ▾

[KBase Labs](#)

[Contact Us](#) ▾

The new Systems Biology Knowledgebase (KBase) is a collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants. It will be a community-driven, extensible and scalable open-source software framework and application system. KBase will offer free and open access to data, models and simulations, enabling scientists and researchers to build new knowledge and share their findings.

For scientists

[Collaborate with us](#)

For users

[Get Started](#)

For developers

[Develop with us](#)

What can KBase do?

- ✓ Combine heterogenous data types
- ✓ Offer standardized access to bioinformatic and modeling analyses
- ✓ Use evidence-supported annotations of genome structure and genetic function



Latest News

[August Build](#)

Posted by salazar August 09th, 2012 at 17:13pm

[Nomi Harris Joins KBase](#)

Posted by salazar July 19th, 2012 at 15:15pm

[KBase in June HPC Source](#)

Posted by salazar June 30th, 2012 at 15:38pm

[view news](#)

Upcoming Events

2012-09-27

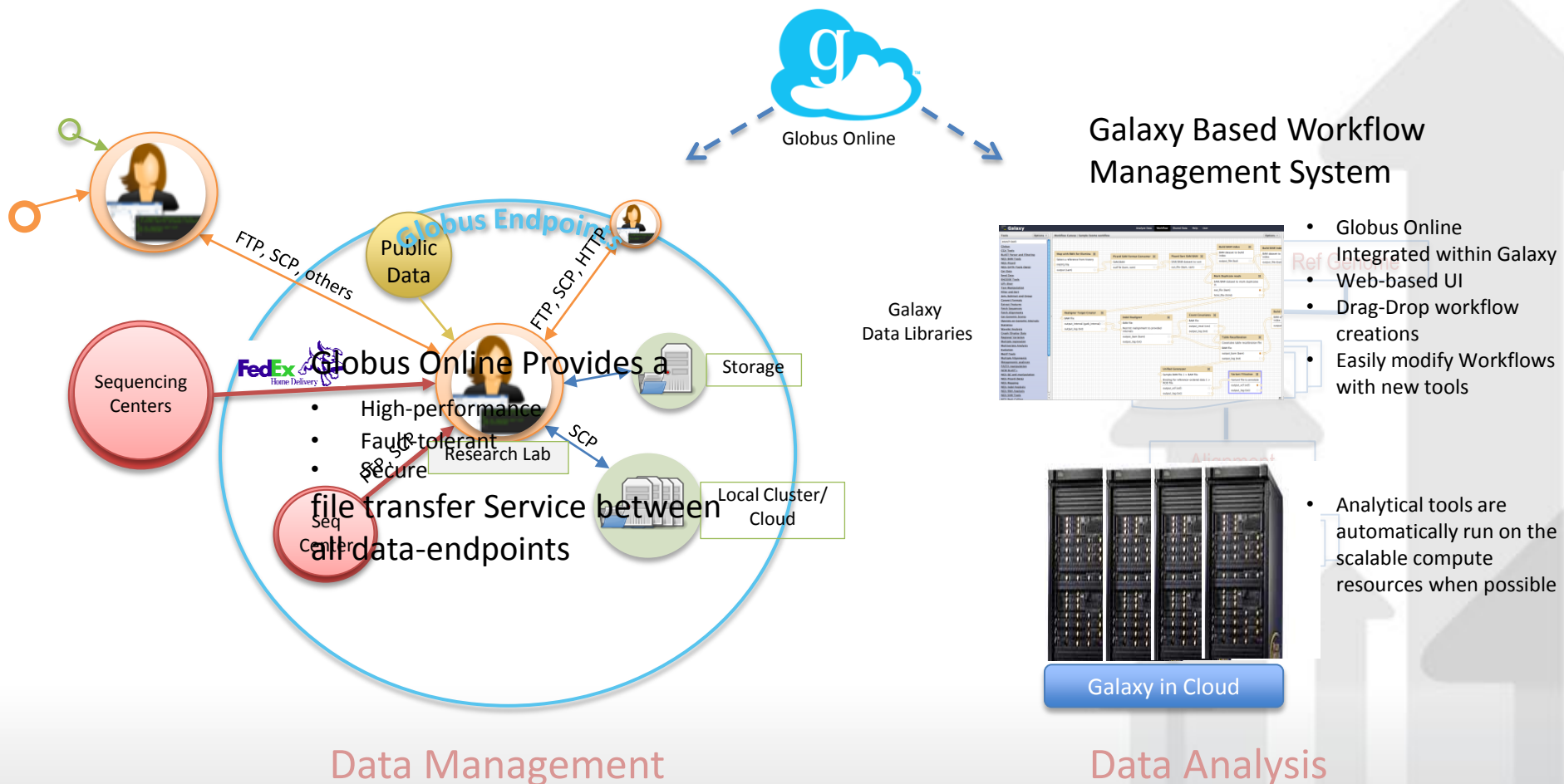
[Mike Schatz at Beyond the Genome](#)

2012-10-21

[ASA-CSSA-SSA Conference](#)



Genomics Analysis using Globus Online and Galaxy





Thank you!

