



GlobusWORLD 2012

Innovating Research Data Management

Ian Foster – April 2012

The data deluge





Run a research project from a coffee shop?

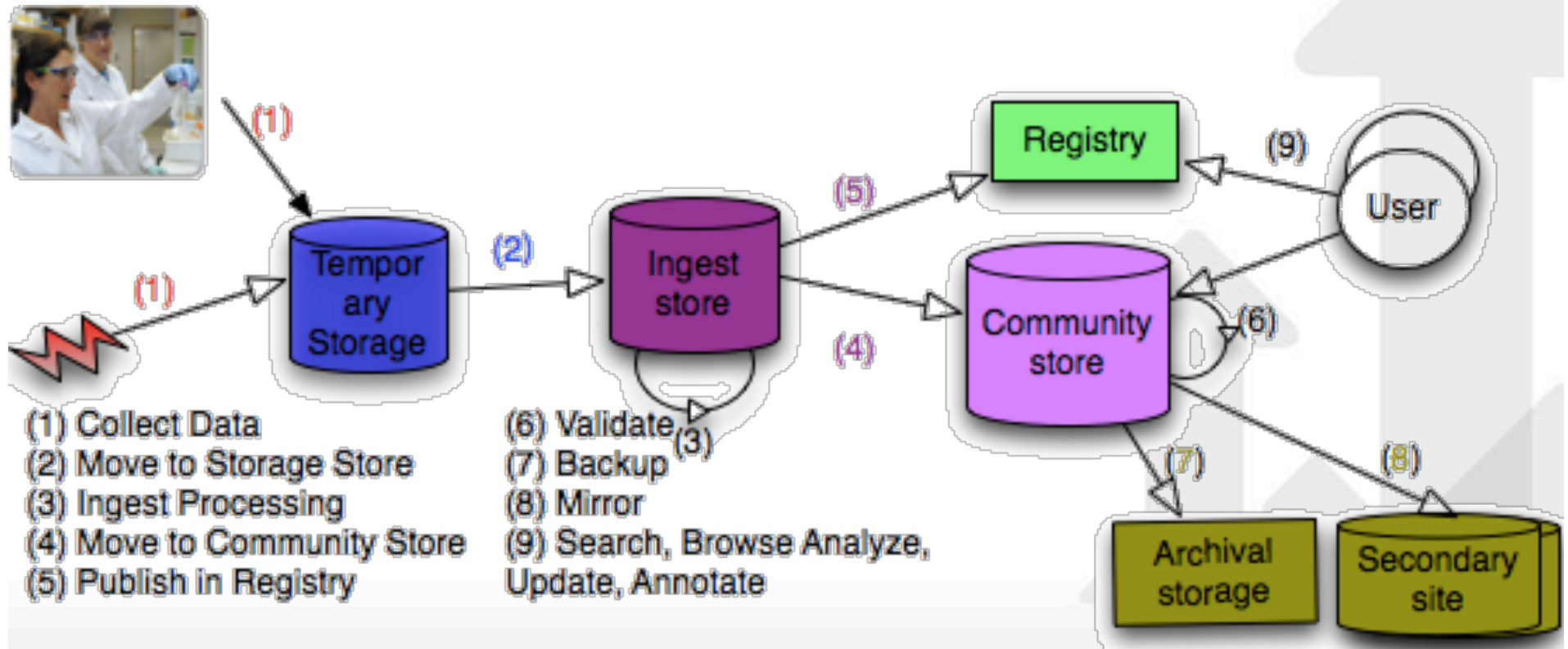
STARBUCKS COFFEE





Towards “research IT as a service”

- Dark Energy Survey
- Galaxy genomics
- LIGO observatory
- SBGrid structural biology consortium
- NCAR climate data applications
- Land use change; economics





Let's rethink 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**



2011: Grid meets Cloud

Globus Toolkit

Build the Grid



Components for building
custom grid solutions

globustoolkit.org

Globus Online

Use the Grid



Reliable file transfer
Software-as-a-Service

globusonline.org



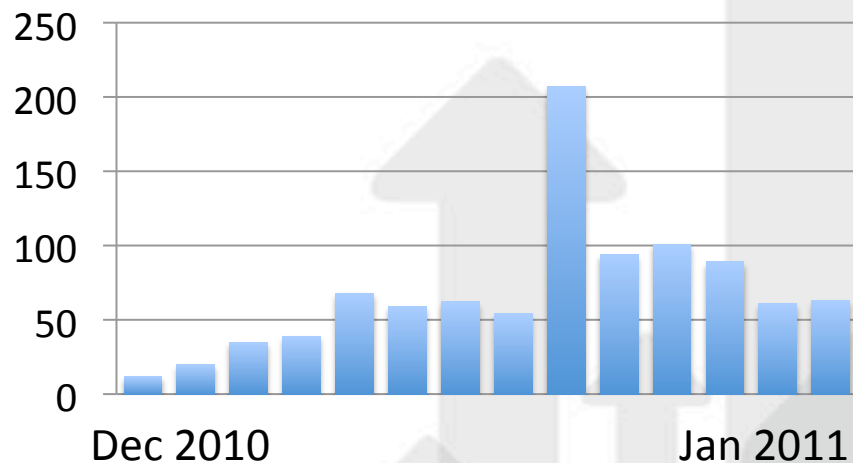
Globus Online – The first 18 months

- **>4,500** registered users; adding **~300/month**
- **>3.5 PB** moved; averaging **~50 TB/week**

Data Transferred (TB/week)



Support Requests Received





Globus Online – The first 18 months

- **514 active endpoints** (year-to-date) including **62 at Tier 1** research institutions
- **>390,000,000 files** moved
- **>100,000 transfer requests**
- **99.9% uptime** in 2012
- **18 identity providers** supported, using multiple protocols: X.509, MyProxy OAuth, OpenID

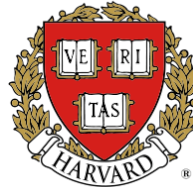


Our user communities

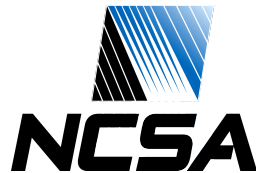
XSEDE

Extreme Science and Engineering
Discovery Environment

NERSC



APS
physics



**Carnegie
Mellon
University**

KSU

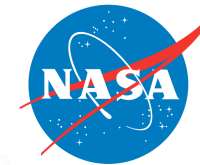


**THE UNIVERSITY
OF AUCKLAND**
NEW ZEALAND

Te Whare Wānanga o Tāmaki Makaurau



Information Sciences Institute



EMORY



INDIANA
UNIVERSITY

Fermilab

Los Alamos
NATIONAL LABORATORY
EST. 1943

CORNELL
UNIVERSITY



**THE UNIVERSITY OF
CHICAGO**



NEW YORK UNIVERSITY

Ole Miss

JÜLICH
FORSCHUNGSZENTRUM

Argonne
NATIONAL LABORATORY



[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](#)

Username

Password

Sign In



2012: Move. Store. Collaborate.



the globus® toolkit
www.globustoolkit.org



globus online



**Globus
Transfer**

Globus Storage

**Globus
Collaborate**

Globus Integrate

Globus Connect Multi User

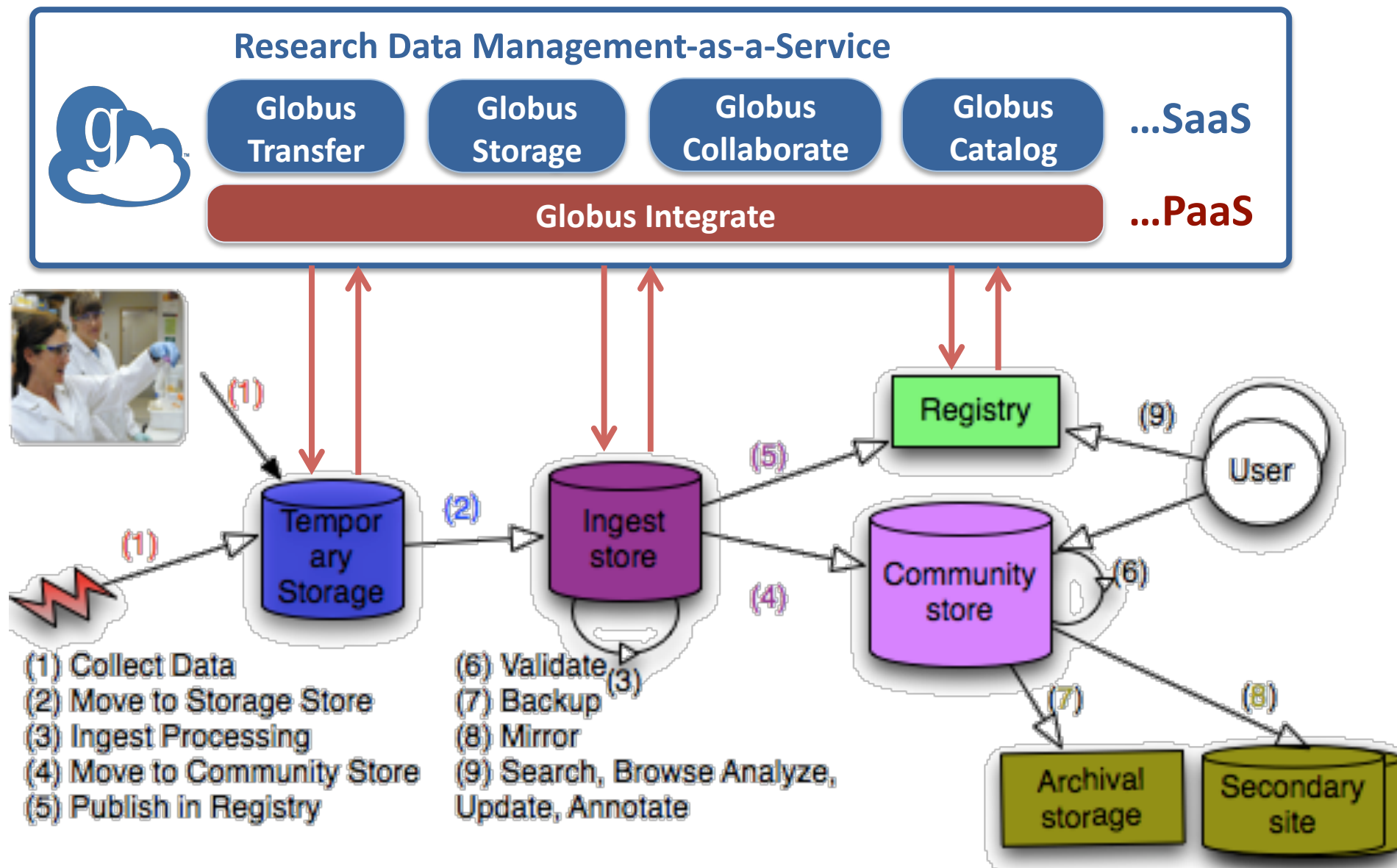
Globus Connect

Globus Nexus

Globus Toolkit



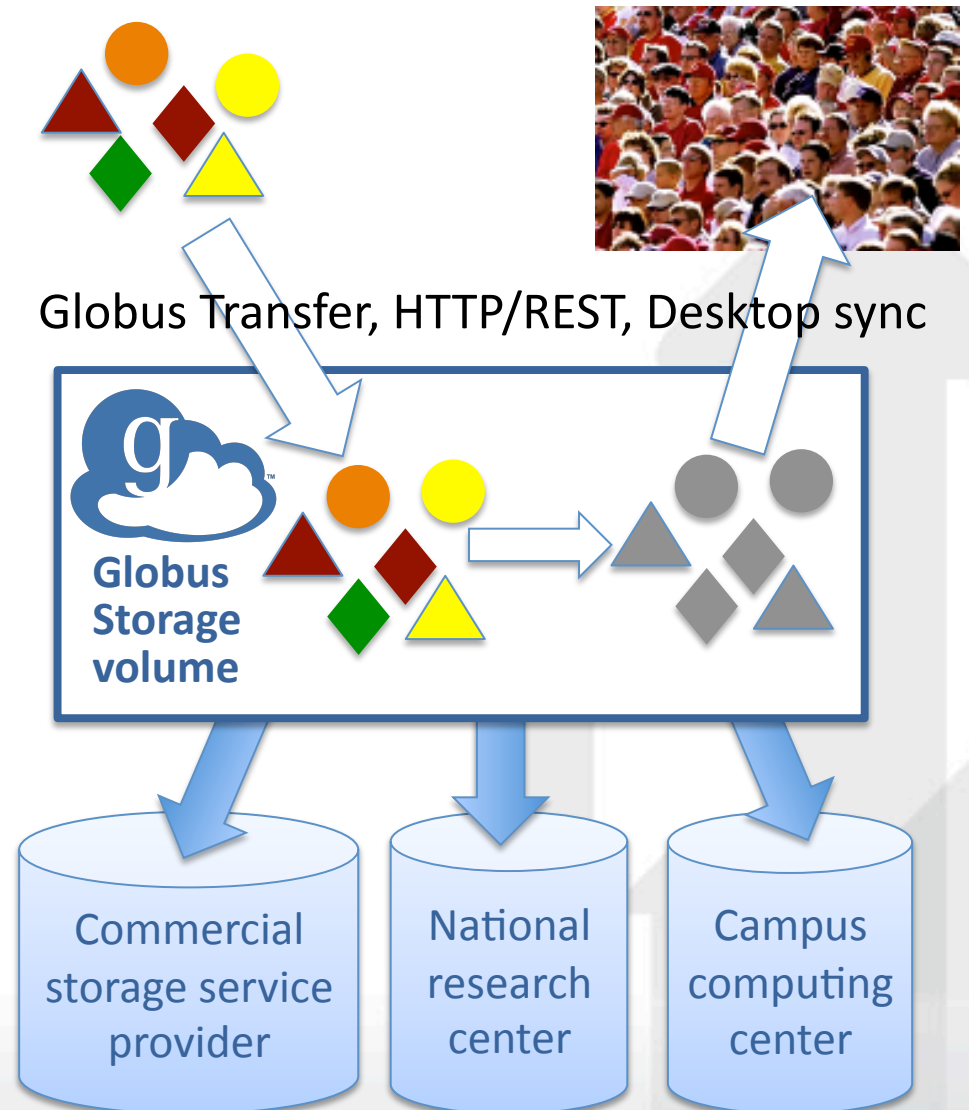
Towards “research IT as a service”





Globus Storage: For when you want to ...

- **Place** your data where you want
- **Access** it from anywhere via different protocols
- **Update** it, **version** it, and take **snapshots**
- **Share** versions with who you want
- **Synchronize** among locations





Globus Collaborate: For when you want to

Join with a few or many people to:

- **Share documents**
- **Track tasks**
- **Send email**
- **Share data**
- **Do whatever**

With:

- **Common groups**
- **Delegated management**

CENTER for MULTISCALE THEORY and SIMULATION
NSF CENTER for CHEMICAL INNOVATION

home | who we are | latest research | collaborators | industrial collaborators | our environment | opportunities

Providing transformative theoretical and computational methods relating the molecular scale to cellular processes

Research Sponsors
NSF National Science Foundation
THE UNIVERSITY OF CHICAGO

This NSF Center for Chemical Innovation (CCI) project is focused on developing a novel, systematic, and transformative scientific capability for the scientific community. The project will combine conceptual advances in statistical mechanics and condensed phase dynamics with computer simulation methodology and cyberinfrastructure.

Scientific Goals and Impacts
o Develop a rigorous theoretical and computational methodology to describe biomolecular systems at

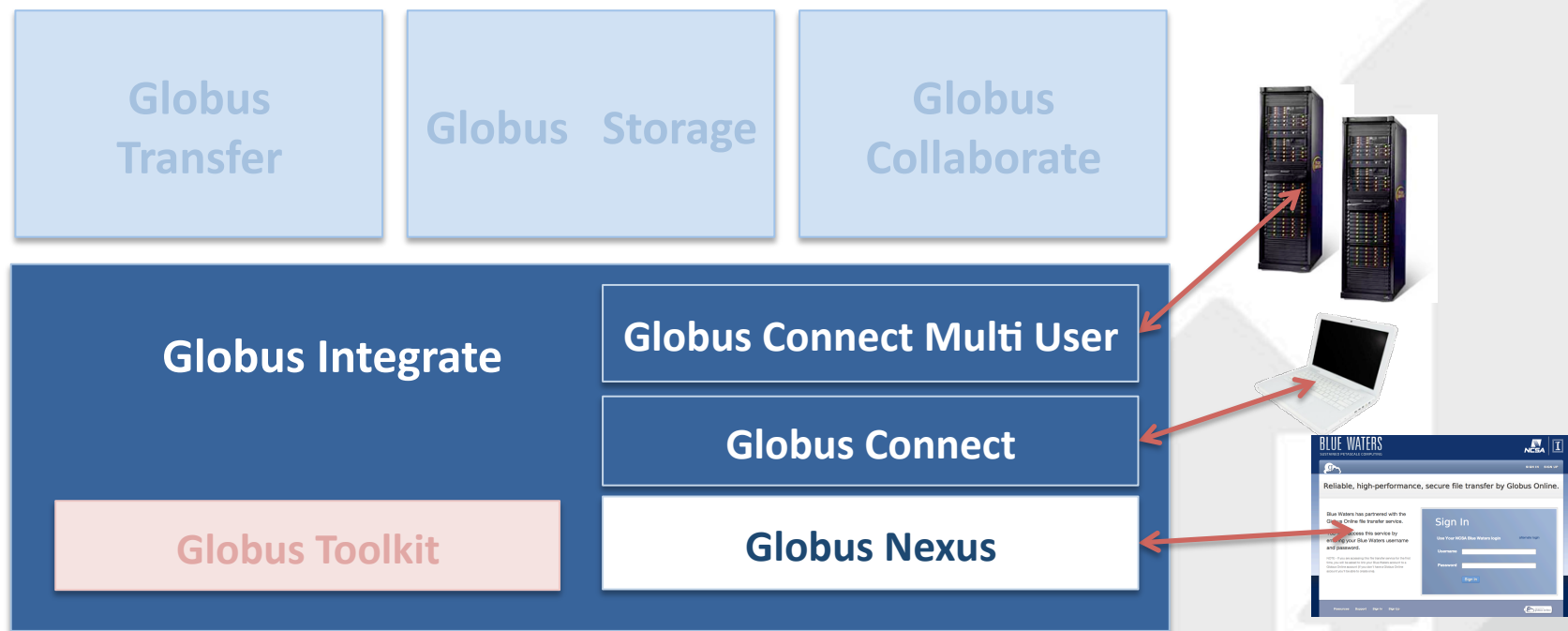
Broader Impacts
o Innovations in computational software will be disseminated publicly and

A CMTS Case Study
Actin, a cytoskeletal protein, is an ideal example



Globus Integrate: For when you want to...

Write programs that access/manage user identities, profiles, groups, resources—and data ...



... via REST APIs and command line programs



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_piontri.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: neillm#ci-vm

Target Endpoint Path Location: neillm#go-esg

next

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 ballen#oathcondor ballen#oathdash ballen#oathdcwan ballen#oathfrost ballen#oathgpfswan ballen#oathhpss ballen#oathkraken

- 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

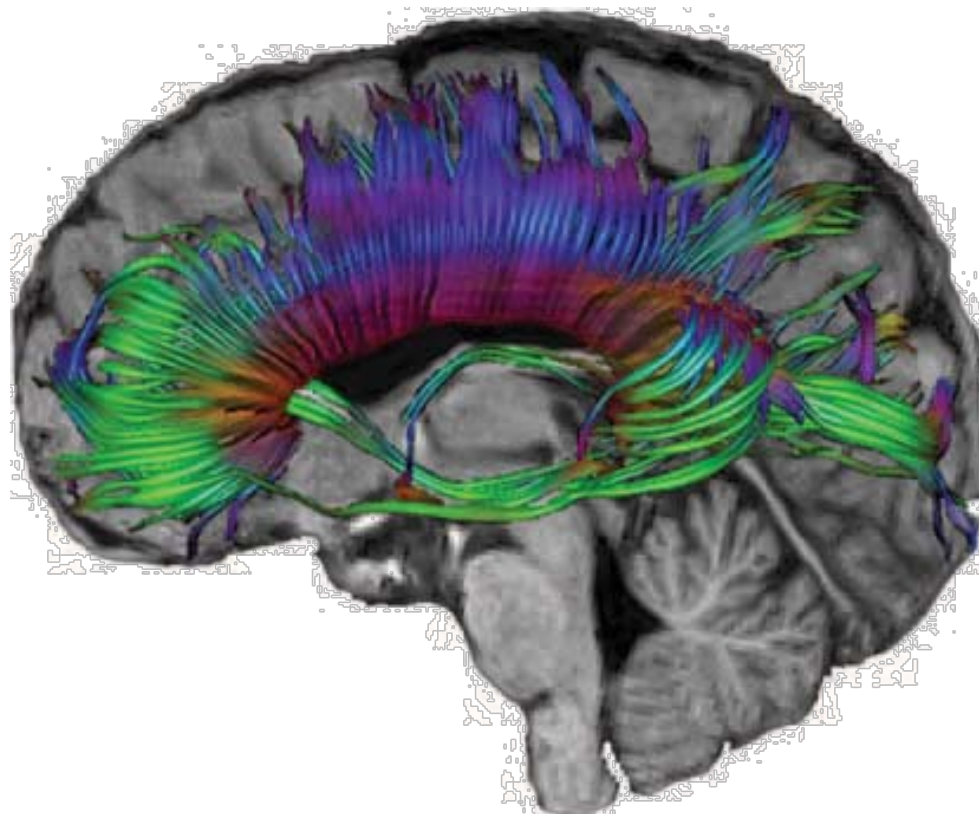
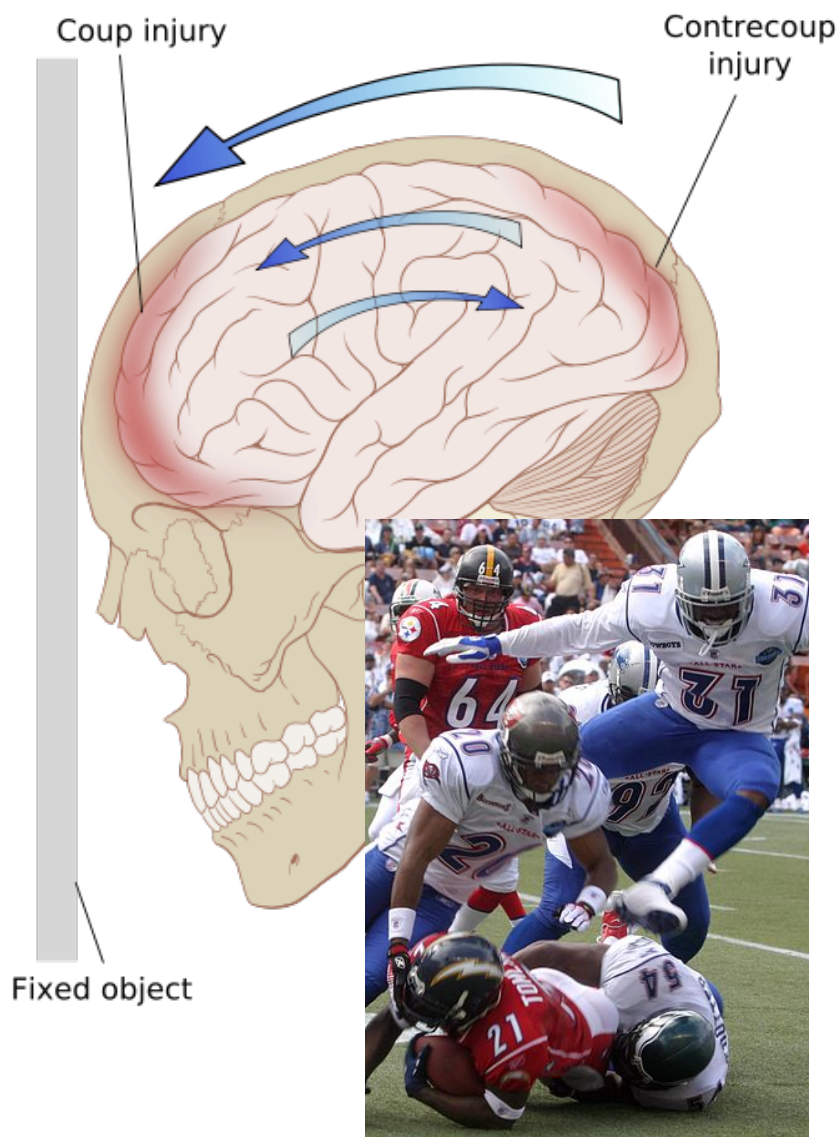
[View Globus Online Transfer](#)

Model Params sent

Error:



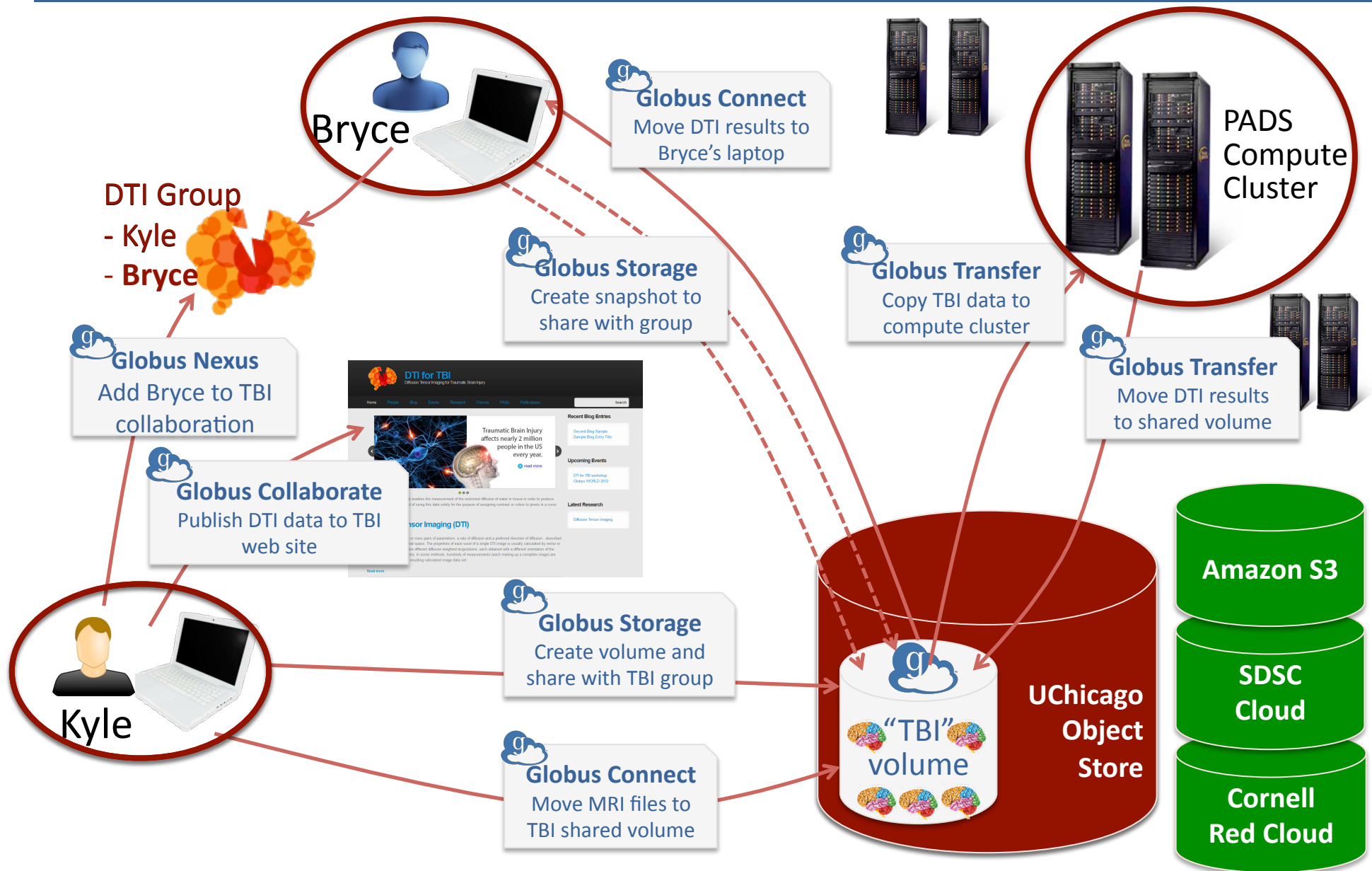
Traumatic Brain Injury (TBI)



Diffusion Tensor Imaging (DTI)



Let's take a look at what's coming...





Globus Toolkit update

GT Releases 2011-2012

5/4/11 GT 5.1.0

5/18/11 GT 5.0.4

8/2/11 GT 5.1.1

11/16/11 GT 5.1.3 (5.2 beta)

12/15/11 GT 5.2.0

3/8/12 GT 5.0.5

Likely last 5.0.x release

4/5/12 GT 5.2.1rc1

5.2.1 final planned for April

Highlights

- Native packaging
- GridFTP: DCSC support
- New GRAM5 version; in use on OSG



Globus Toolkit update – GridFTP

- **GT 5.2.0**

- Support for DCSC command
 - Allows client to specify credentials used to secure data channel connection
 - Utilized by Globus Transfer for seamless data movement across multiple security domains
- Server administrators may now restrict client access to a set of paths

- **Coming in 5.2.1**

- MLSC – stream directory listings over control channel
 - Enables simple (control channel-only) implementation of Globus Storage ftp server
 - Can help directory listings for gridftp servers behind restrictive firewalls



Delivery plans

Globus Transfer

- Generally available
- Service and Web UI enhancements continue

Globus Storage

- Early release
- Generally available in Fall 2012

Globus Collaborate

- Initial projects at UChicago
- Early release sometime before year-end 2012

Globus Integrate

- Transfer API available
- APIs for Storage, Collaborate planned after app release

Globus Toolkit (GA V5.2)

Globus Connect Multi User (GA)

Globus Connect (GA)

Globus Nexus (Alpha)



Delivery plans

- **Globus Storage**

- Current: Early access – contact us if interested
- Open beta planned in late Spring; GA in late Summer
- V1.0: UChicago and Amazon S3 object stores

- **Globus Collaborate**

- Current: Pre-release with two UChicago/CI groups
- First release planned around year-end



Premium offerings

Motivated by sustainability: resource cost recovery, user support, future development

- **Globus Transfer plans (resource providers)**
 - www.globusonline.org/premium-plans/
- **Globus Storage premium plans**
 - Plans variables include: size, object store type, durability, support level, personalization
- **Globus Collaborate premium plans**
 - Plans variables include: branding capabilities, support level, integration (e.g. identity)



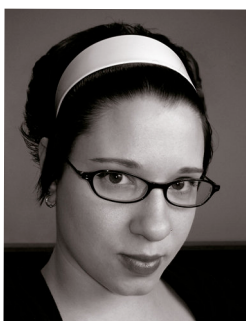
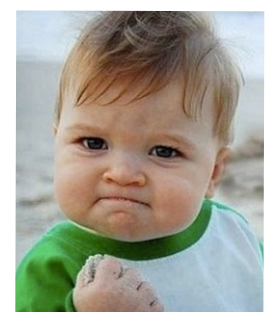
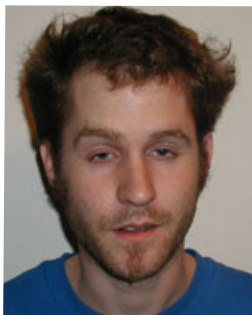
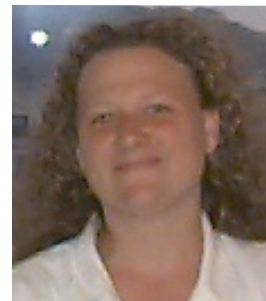
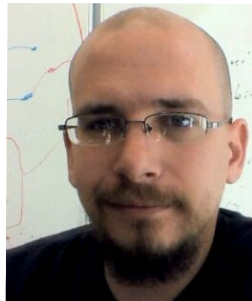
Premium offerings

Motivated by sustainability: resource cost recovery, user support, future development

- **Globus Transfer plans (resource providers)**
 - www.globusonline.org/premium-plans/
- **Globus Storage premium plans**
 - Plans variables include: size, object store type, durability, support level, personalization
- **Globus Collaborate premium plans**
 - Plans variables include: branding capabilities, support level, integration (e.g. identity)



The Globus yeam





Thank you!

