GlobusWORLD 2012
Innovating Research Data Management

Ian Foster – April 2012
The data deluge
Run a research project from a coffee shop?
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

1. Collect Data
2. Move to Storage Store
3. Ingest Processing
4. Move to Community Store
5. Publish in Registry
6. Validate
7. Backup
8. Mirror
9. Search, Browse, Analyze, Update, Annotate
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
• **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
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).

Globus Toolkit
- Globus Transfer
- Globus Storage
- Globus Integrate
  - Globus Connect Multi User
  - Globus Connect
  - Globus Nexus
Towards “research IT as a service”

Research Data Management-as-a-Service

- Globus Transfer
- Globus Storage
- Globus Collaborate
- Globus Catalog

Globus Integrate

...SaaS

...PaaS

1. Collect Data
2. Move to Storage Store
3. Ingest Processing
4. Move to Community Store
5. Publish in Registry
6. Validate
7. Backup
8. Mirror
9. Search, Browse, Analyze, Update, Annotate

www.globustoolkit.org
www.globusonline.org
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
Join with a few or many people to:

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

With:

- Common groups
- Delegated management
Write programs that access/manage user identities, profiles, groups, resources—and data …

Globus Integrate: For when you want to …

… via REST APIs and command line programs
Earth System Grid – Portal integration

- 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
New Feature Requests (1)

💡 Allow users to define an initial endpoint directory
Let’s take a look at what’s coming…

Traumatic Brain Injury (TBI)

Coup injury

Contrecoup injury

Fixed object

Diffusion Tensor Imaging (DTI)
Let's take a look at what's coming...

Kyle

- DTI Group
  - Kyle
  - Bryce

Globus Nexus
Add Bryce to TBI collaboration

Globus Connect
Move MRI files to TBI shared volume

Globus Storage
Create volume and share with TBI group

Globus Collaborate
Publish DTI data to TBI web site

Globus Connect
Move DTI results to Bryce's laptop

Globus Transfer
Copy TBI data to compute cluster

Globus Transfer
Move DTI results to shared volume

Globus Transfer
Move DTI results to Bryce's laptop

Globus Storage
Create snapshot to share with group

UCheicago Object Store

Amazon S3
SDSC Cloud
Cornell Red Cloud

PADS Compute Cluster

www.globustoolkit.org
www.globusonline.org
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
• **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 Connect Multi User (GA)**

**Globus Connect (GA)**

**Globus Toolkit (GA V5.2)**

**Globus Nexus (Alpha)**
• **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
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)
Motivated by sustainability: resource cost recovery, user support, future development

- **Globus Transfer plans (resource providers)**
  - [www.globusonline.org/premium-plans/](http://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)
Thank you!