Globus Online
Toward a hosted solution for Research Data Management

Rachana Ananthakrishnan
Computation Institute
University of Chicago and Argonne National Laboratory
Time-consuming tasks in science

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature

- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- ...
Towards “research IT as a service”

Scientific data management as a service

GO-Store  GO-Collaborate  GO-Galaxy  GO-Transfer
GO-Compute  GO-Catalog  GO-Team  GO-User

(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

Registry

User

Archival storage
Secondary site
• **Goal:** Accelerate discovery and innovation worldwide by providing research IT as a service

• **Leverage software-as-a-service to:**
  – provide millions of researchers with unprecedented access to powerful tools
  – reduce research IT costs dramatically via economies of scale
  – Simplify mundane, but required research tasks, so researchers can focus on science
Exemplar problem space: Typical multi-site study
## Exemplar problem space: Data movement

<table>
<thead>
<tr>
<th>User</th>
<th>Data location</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomedical Researcher</td>
<td>FBIRN sites (8 different sites)</td>
<td>Multiple security domains, repetitive transfer, deadline bound, metadata based search</td>
</tr>
<tr>
<td>Nuclear Scientist</td>
<td>Oak Ridge to NERSC</td>
<td>Two security domains, blocked by transfer, repetitive task</td>
</tr>
<tr>
<td>System Administrator</td>
<td>To GFDL</td>
<td>Many security domains, administrative task, deadline bound</td>
</tr>
<tr>
<td>System Builder</td>
<td>To and from NERSC</td>
<td>Many security domains, support adhoc users, legacy code integration, multiple science domains</td>
</tr>
</tbody>
</table>
Globus online

Globus Transfer
Globus Transfer

• Easy “fire and forget” file transfers
• Automatic fault recovery
• High performance
• Simplify use of multiple security domains
• No client software installation
• New features automatically available
• Consolidated support and troubleshooting
SignUp... Already a member? Sign In.

Full Name: Rachana Ananthakrishnan
Email: ranantha@mcs.anl.gov
Username: ranantha
Password: **********
Confirm Password: **********
Terms of Service: Last Updated: 10-Nov-2010

This is a legal agreement between you and Globus Consortium, Inc. ("Globus Online"); please read it carefully. By using or accessing this website or any website hosted at globusonline.org (the "Sites"), or by accessing or using any part of the Globus Online file transfer service (together with the Sites, the "Service"), you agree to be bound by these Terms of Service and by the Globus Online Privacy Policy, which are collectively referred to as the "Agreement." If you do not agree to these terms then you have no right to access or use the Service. The terms of this

I agree to the Globus Online Terms of use
Please email me updates about Globus Online

Register
### Globus Online

- Why Use Globus Online?
- How It Works
- Features
- Quickstart Guides

### Resources

- Tutorials
- Downloads
- FAQs
- Pro Tips
- Documentation
- Release Notes
- Troubleshooting

### What's Happening

- What Others Are Saying
- News & Events
- Blog
- Case Studies

### Community
Welcome to the XSEDE User Portal Authorization Page

Science Gateway Access
The XSEDE Science Gateway or Service below is requesting access to your XSEDE account. If you approve, please sign in with your XSEDE username and password.

<table>
<thead>
<tr>
<th>SCIENCE GATEWAY INFORMATION</th>
<th>SIGN IN</th>
</tr>
</thead>
</table>
| The XSEDE Science Gateway listed below is requesting access to your XSEDE account. If you approve, please sign in. | Username: raaganta
Password: ********** |

Name: Globus Online
URL: https://www.globusonline.org/

Please send any questions or comments about this site to help@xsede.org.
Please select an endpoint above.
## Transfer Activity

<table>
<thead>
<tr>
<th>Status</th>
<th>Label</th>
<th>Task Progress</th>
<th>Completion Time</th>
<th>Request Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:f5444d56-617e-11e1-81e6-1231381bd061</td>
<td>4 / 4</td>
<td>02/27/2012 02:14 PM</td>
<td>02/27/2012 02:14 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:9ba3a3a4-1c33-11e1-bc86-1231381a21f</td>
<td>44 / 47</td>
<td>12/01/2011 09:47 AM</td>
<td>12/01/2011 09:46 AM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:7be092de-11e1-bc86-1231381a21f</td>
<td>33 / 33</td>
<td>11/16/2011 07:33 PM</td>
<td>11/16/2011 07:33 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:360d7b2e-11e1-bc86-1231381a21f</td>
<td>12 / 12</td>
<td>11/16/2011 01:12 PM</td>
<td>11/16/2011 01:12 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:d777c82-11e1-bc86-1231381a21f</td>
<td>12 / 12</td>
<td>11/16/2011 01:12 PM</td>
<td>11/16/2011 01:12 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:b43675e6-11e1-bc86-1231381a21f</td>
<td>12 / 12</td>
<td>11/16/2011 01:12 PM</td>
<td>11/16/2011 01:12 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:ca9e1510-11e1-bc86-1231381a21f</td>
<td>12 / 12</td>
<td>11/16/2011 01:12 PM</td>
<td>11/16/2011 01:12 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:70c0b944-11e1-bc86-1231381a21f</td>
<td>12 / 12</td>
<td>11/16/2011 01:12 PM</td>
<td>11/16/2011 01:12 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:1c778346-11e1-bc86-1231381a21f</td>
<td>12 / 12</td>
<td>11/16/2011 01:12 PM</td>
<td>11/16/2011 01:12 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:b45c1866-0fd3-11e1-bc86-1231381a21f</td>
<td>1 / 1</td>
<td>11/15/2011 05:04 PM</td>
<td>11/15/2011 05:04 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:de52cd2c-0fd3-11e1-bc86-1231381a21f</td>
<td>1 / 1</td>
<td>11/14/2011 05:46 PM</td>
<td>11/14/2011 05:46 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:76369b34-0fd3-11e1-bc86-1231381a21f</td>
<td>1 / 1</td>
<td>11/14/2011 04:53 PM</td>
<td>11/14/2011 04:53 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:2f08b44b-0fd3-11e1-bc86-1231381a21f</td>
<td>1 / 1</td>
<td>11/14/2011 04:28 PM</td>
<td>11/14/2011 04:28 PM</td>
</tr>
<tr>
<td>![Green Checkmark]</td>
<td>Task Id:22e60280-0ef6-11e1-bc86-1231381a21f</td>
<td>1 / 1</td>
<td>11/14/2011 01:16 PM</td>
<td>11/14/2011 01:16 PM</td>
</tr>
</tbody>
</table>

### Transfer Details

- **Task ID:** f5444d56-617e-11e1-81e6-1231381bd061
- **Status:** SUCCEEDED
- **Request Time:** 02/27/2012 02:14 PM
- **Deadline:** 02/28/2012 03:19 PM
- **Completion Time:** 02/27/2012 02:14 PM
- **Origin:** ci#pads
- **Destination:** xsede#ranch
- **User:** rantha
- **Directories:** 1
- **Bytes Transferred:** 87
- **Files:** 3
- **Task Statistics:**
  - Pending: 0
  - Succeeded: 4
  - Expired: 0
  - Failed: 0
  - Retrying: 0

[View Event Log]
Globus Connect Installation

Globus Connect allows you to use Globus Online to transfer files to and from your computer.

Need Help? Click Here

Step One: Choose Your Download

- Globus Connect For Mac OS X
- Coming Soon For Linux
- Coming Soon For Windows

Step Two: Get Your Globus Connect Setup Key

- Endpoint Name: jan-laptop
- Description: (optional) Enter a description for this new endpoint

- Setup Key: 432e8ba5-45cf-442b-a374-5a8d1cfa75cb

Step Three: Finish Globus Connect Setup

Copy the setup key displayed above. Run Globus Connect and paste the key into the Initial Setup window when prompted. This setup key can only be used once.

Ready to use your endpoint? Click here to start a transfer.
Outsource their transfer
Ability to use logical names for servers
Prompted to correct credentials
  - Ability to use local credentials
Can secure the data transfer
Performance tuning is automatically done
Notified of success, failure, credential renewal
Globus Connect Multi User
  – Enabling your resource as Globus Transfers endpoint
  – Pre-packaged software for administrators to install

Create a transfer endpoint in minutes
  – Sets up a shared GridFTP server
  – Links to the local authentication system
  – Registers with Globus Transfer
Evolving Features

- Custom branded website with transfer support
- Enhanced endpoint administrator interfaces
Globus Integrate
• Platform that allows other applications to integrate into the Globus Online ecosystem

• Horizontal platform services
  – Make it easier to create new services in ecosystem
  – Currently Identity management and Group management

• APIs to integrate with Globus Online vertical offerings
  – REST APIs to Globus services: File Transfers
  – Others coming soon
Identity Management Use Cases

- **Users need to manage multiple identities & formats**
  - E.g. XSEDE, Campus cluster, science portal, SSH logins
- **Need multiple identities to accomplish a single task**
  - E.g. Data transfer across above domains
- **Provision multiple applications with different identity formats**
  - E.g. X.509 Certificates, local usernames, federated identity
- **Simple interfaces for associating identity requirements with resource**
  - E.g. I need a credential from XSEDE MyProxy server for using XSEDE resource
Globus Online as Identity Provider
- User account creation
- Username and password

Globus Online as Identity Hub
- Link external identities
  - MyProxy, MyProxy OAuth, Google OpenID
  - SSH Public Key, PKI/X.509 Certificate
- Provide federated identity provider and service provider capabilities
Globus Online Profile Management

- **Ability to associate multiple identities with single account**
  - E.g. My Google OpenID, my XSEDE username and DOE Grids DN
  - No passwords or private key information is stored
- **Allows mapping from one to the other, and provision applications**
  - E.g. For same user, InCommon credential to a NIH portal and XSEDE username to a XSEDE portal
- **Additional profile information**
  - Email address
  - Potentially other user attributes from IdP
### My Account

#### Update Profile

- Change Password

#### Manage Identities

<table>
<thead>
<tr>
<th>Alias</th>
<th>Type</th>
<th>Authentication Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>system-generated openid alias</td>
<td>openid</td>
<td>google.com</td>
</tr>
<tr>
<td>Default MyProxy Account</td>
<td>myproxy</td>
<td>ec2-184-73-254-246.compute-1.amazonaws.com</td>
</tr>
<tr>
<td>tg11-tutorial</td>
<td>ssh2</td>
<td>SSH Public Key</td>
</tr>
<tr>
<td>GCMU Test</td>
<td>ssh2</td>
<td>SSH Public Key</td>
</tr>
<tr>
<td>GCMU SC11</td>
<td>ssh2</td>
<td>SSH Public Key</td>
</tr>
<tr>
<td>system-generated ssh alias 1</td>
<td>ssh2</td>
<td>SSH Public Key</td>
</tr>
<tr>
<td>DOE Certificate</td>
<td>x509</td>
<td>CN=DOEGrids CA 1...</td>
</tr>
<tr>
<td>ESG Certificate</td>
<td>x509</td>
<td>CN=ESG-NCAR CA, ...</td>
</tr>
<tr>
<td>system-generated x509 alias 1</td>
<td>x509</td>
<td>/C=US/O=Gridus A...</td>
</tr>
</tbody>
</table>
• **Other external identity providers**
  – Federated Service Provider: SAML/InCommon, OpenID/OpenID LOA-1, OAuth
  – Federated Identity Provider: SAML, OpenID, LDAP
  – Simpler external provider sign-up and sign-in

• **Profile and attribute Management**
  – Consume attributes from other providers
  – Assert attributes to other consumers
• **Manage Collaborations**
  – Create and manage groups of users
  – Provision them in various applications

• **Key Features**
  – Autonomous group management
  – Admission and vetting workflows
  – Visibility, inheritance policies
  – Customized presence for groups
BIRN use of Globus Integrate
The main goal of the BIRN is to enable collaborative bioscience
- Infrastructure and services to support the above goal

User Requirements:
- Share data (internal and/or external)
- Share tools
- Develop, add to, merge ontologies
- Query across disparate data sources
- Utilize high-performance computing resources
BIRN Customized Web Presence

Sign Up has partnered with Globus Online to provide us with our community platform and tools. If you already have a Globus Online account, click here.

- **Full Name**: Rachana Ananthakrishnan
- **Email**: rananba@mcs.anl.gov
- **Username**: rananbirn4
- **Password**: ******** (Strength: Very Weak)
- **Confirm Password**: ********
- **Terms of Service**: Last Updated: 10-Nov-2010

This is a legal agreement between you and Globus Consortium, Inc. ("Globus Online"); please read it carefully. By using or accessing this website or any website hosted at globusonline.org (the "Sites"), or by accessing or using any part of the Globus Online file transfer service (together with the Sites, the "Service"), you agree to be bound by these Terms of Service and by the Globus Online Privacy Policy, which are collectively referred to as the "Agreement". If you do not agree to these terms, you have no right to access or use the Service. The terms of this Agreement apply to all users of the Service.

I agree to the Globus Online Terms of use.

Register
<table>
<thead>
<tr>
<th><strong>Full Name</strong></th>
<th>Rachana Ananthakrishnan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E-Mail Address</strong></td>
<td><a href="mailto:ranantha@mcs.anl.gov">ranantha@mcs.anl.gov</a></td>
</tr>
<tr>
<td><strong>Current Project</strong></td>
<td>FBIRN</td>
</tr>
<tr>
<td><strong>Contact Phone Number</strong></td>
<td>123456789</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>University of Chicago</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td>Computational Institute</td>
</tr>
</tbody>
</table>

Submit  cancel
Email validation is required before your request can be accepted.

Check your email inbox and proceed to the location provided, and we'll have you on your way.

Can't find your validation email?

Confirm your email address: ranantha@mcs.anl.gov

Resend
Create New Group

Name Group

Group Belongs To

- BIRNCommunity
- FBIRN

Group Name

University of Chicago

Description

Group of FBIRN users at University of Chicago

Create Group
Configure New Group

Set Policies

Group Visibility
- Public (visible to all)
- BIRN Community Members only (visible to all)
- Visible within parent group (FBIRN)
- Private (by invitation only)

Member Name Visibility
- Public (visible to all who can view group)
- Visible within parent group (FBIRN)
- Visible to this group's members only
- Visible to this group's admins only

Required Fields to Join Group
- Pre-selected checkboxes indicate that this field is required by the parent group.
  - User Name
  - Full Name
  - E-Mail Address
  - Current Project Name
  - Institution
  - Organization
  - Address
  - City
  - State
  - Zip Code
  - Phone

Who Can Invite New Members
- Any member of BIRN Community
- This group's members and members of FBIRN
- This group's members only
- Admin only

Who can request membership?
- Anybody can request membership to this group.
- Any member of the BIRN Community can request membership.
- Only members of FBIRN can request membership.
- Requests to join are disabled, membership to this group is by invitation only.

Approvals
- Auto approve member once all policies are met
- Auto approve admin invited member once all policies are met, others require admin approval
- Admin approval required

Continue
Finish Later
Provisioning of groups

- Secure access to documents
- Secure access to data
- Manage accounts and groups

Wiki (Confluence)

BIRN Community Web Presence

Data Transfer (Globus GridFTP)

Provision accounts and groups

Globus Integrate

Provision accounts and groups

Globus Integrate
Evolving Features

• Custom vetting policies
• Pull and notification interfaces for group information
• Standard attribute service interfaces
Globus Collaborate (Beta)
• **Simplify creation of community web presence**
  – Hosted for the community
  – Shared document space, forums, mailing lists, blogs
  – Customized presence for the community
  – Autonomous administration by community

• **Easy integration with Globus ecosystem**
Case Studies
Lattice QCD – Ad hoc user

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

“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.”

Indiana University researcher moved ~6 TB from Oak Ridge to TACC in 2 days

“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.”
ESGF – Portal Integration

- **Outsource data transfer to GO**
  - 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**
• **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 Transfer = 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 GO
  – 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.
CMU – Cluster administrator solution

• Challenge:
  – Cluster at Department of Physics at CMU
  – Need to move data to and from XSEDE resources and cluster
  – Data sets of varying size

• Solution
  – Globus Connect Multi User on cluster
  – Globus Connect for personal machine

• Benefits for user
  – Auto tuning for various file sizes
  – High performance transfer
  – Move seamlessly across multiple security domains

• Benefits for cluster administrator
  – Enable Globus Transfer at cluster easily using GCMU
  – Transfer feature outsourced to Globus Online

CMU researcher moved data from XSEDE to his cluster at ~2Gbps
GCMU made moving data in and out of this cluster trivial
Time-consuming tasks in science

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature
- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- ...

www.globusonline.org
For More Information

• GlobusWorld: April 10-12 @ Chicago
• Visit https://www.globusonline.org/signup to:
  – Get a free account and start moving files
• Visit www.globusonline.org for:
  – Tutorials, FAQs, Pro Tips, Troubleshooting
  – Papers, Case Studies
• Contact support@globusonline.org for:
  – Help getting started and using the service
• Follow us at @globusonline on Twitter
  and Globus Online on Facebook