The globusworld TOUR

Best Practices for Data Sharing

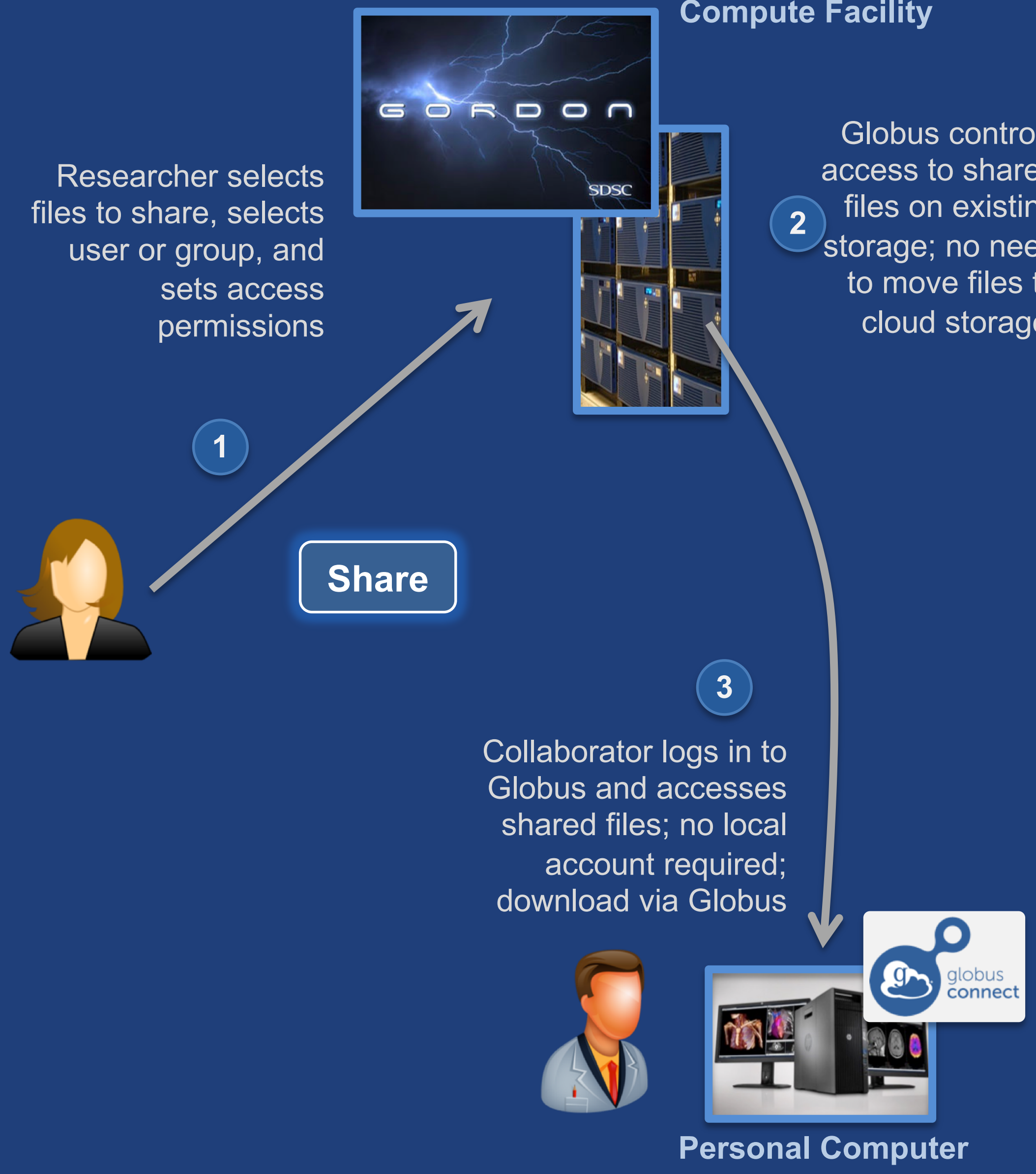
Vas Vasiliadis
vas@uchicago.edu

NCAR – September 5, 2018

globus®

Ad hoc data sharing

- **Individual users share data with collaborators**
- **Using a known email or identity for user/group**
- **Make data publicly available**



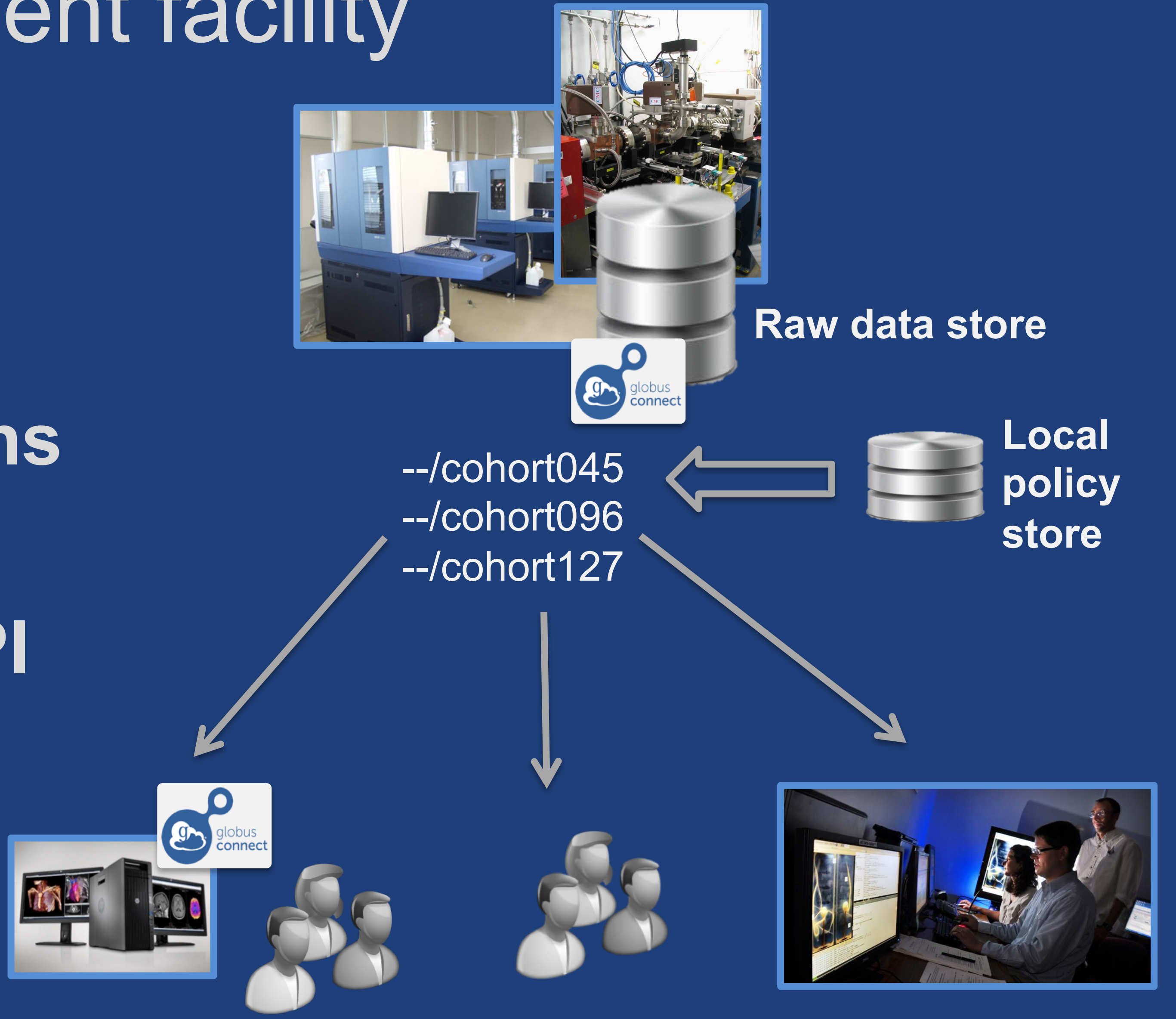
1 Researcher selects files to share, selects user or group, and sets access permissions

2 Globus controls access to shared files on existing storage; no need to move files to cloud storage!

3 Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus

Data from instrument facility

- Provide near-real time access to data
- Automated permissions based on site policy
- Self managed by the PI
- Federated login to access data



Raw data store

--/cohort045
--/cohort096
--/cohort127

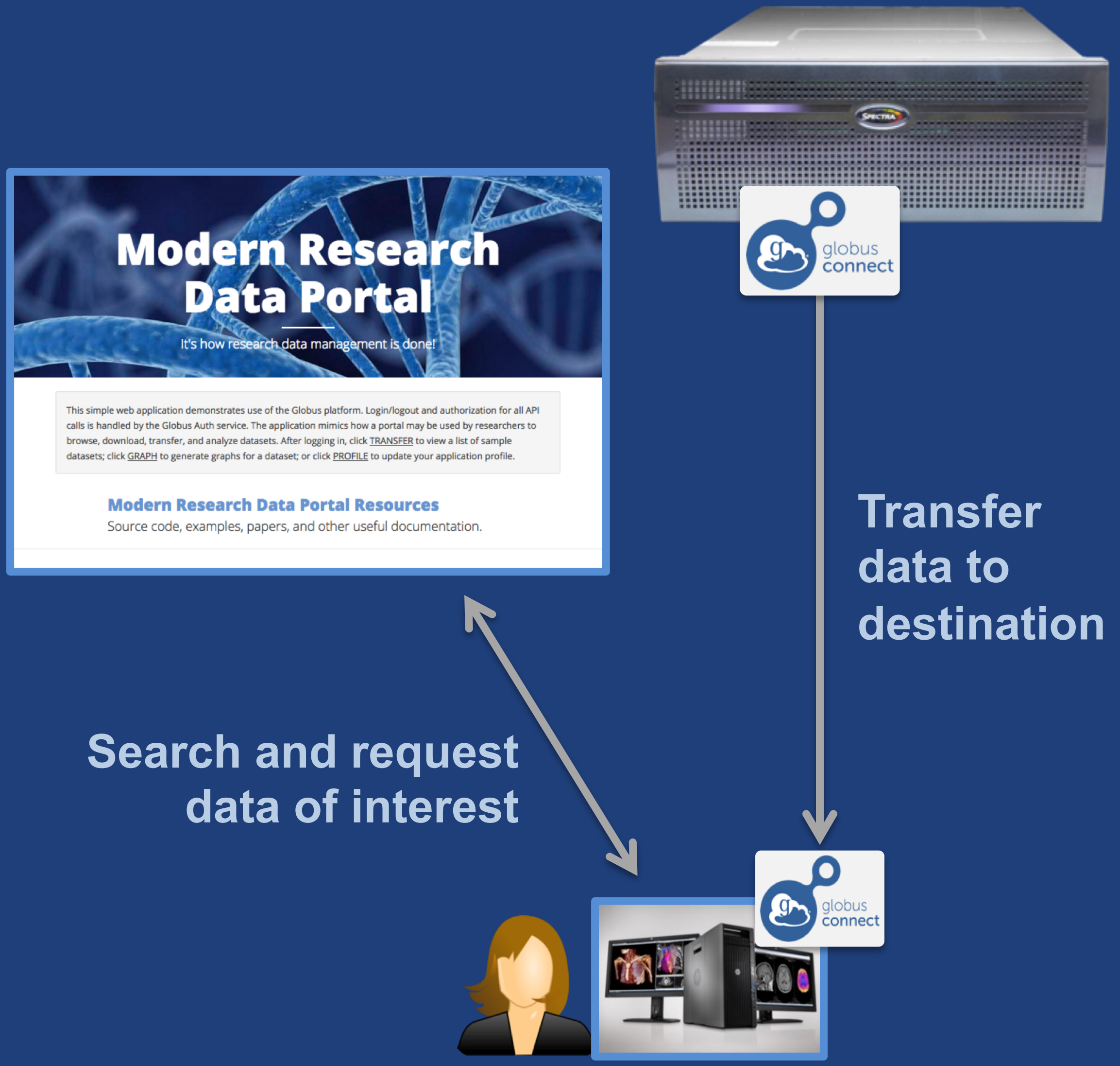
Local policy store

Personal Computer

Remote visualization/analysis

Data from provider/archive

- Portal/science gateway to distribute data
- Interface to search and gather data of interest
- Asynchronous transfer to user’s system or via HTTPS to “staged” data
- Fine-grained authorization enforced



Modern Research Data Portal

It's how research data management is done!

This simple web application demonstrates use of the Globus platform. Login/logout and authorization for all API calls is handled by the Globus Auth service. The application mimics how a portal may be used by researchers to browse, download, transfer, and analyze datasets. After logging in, click TRANSFER to view a list of sample datasets; click GRAPH to generate graphs for a dataset; or click PROFILE to update your application profile.

Modern Research Data Portal Resources

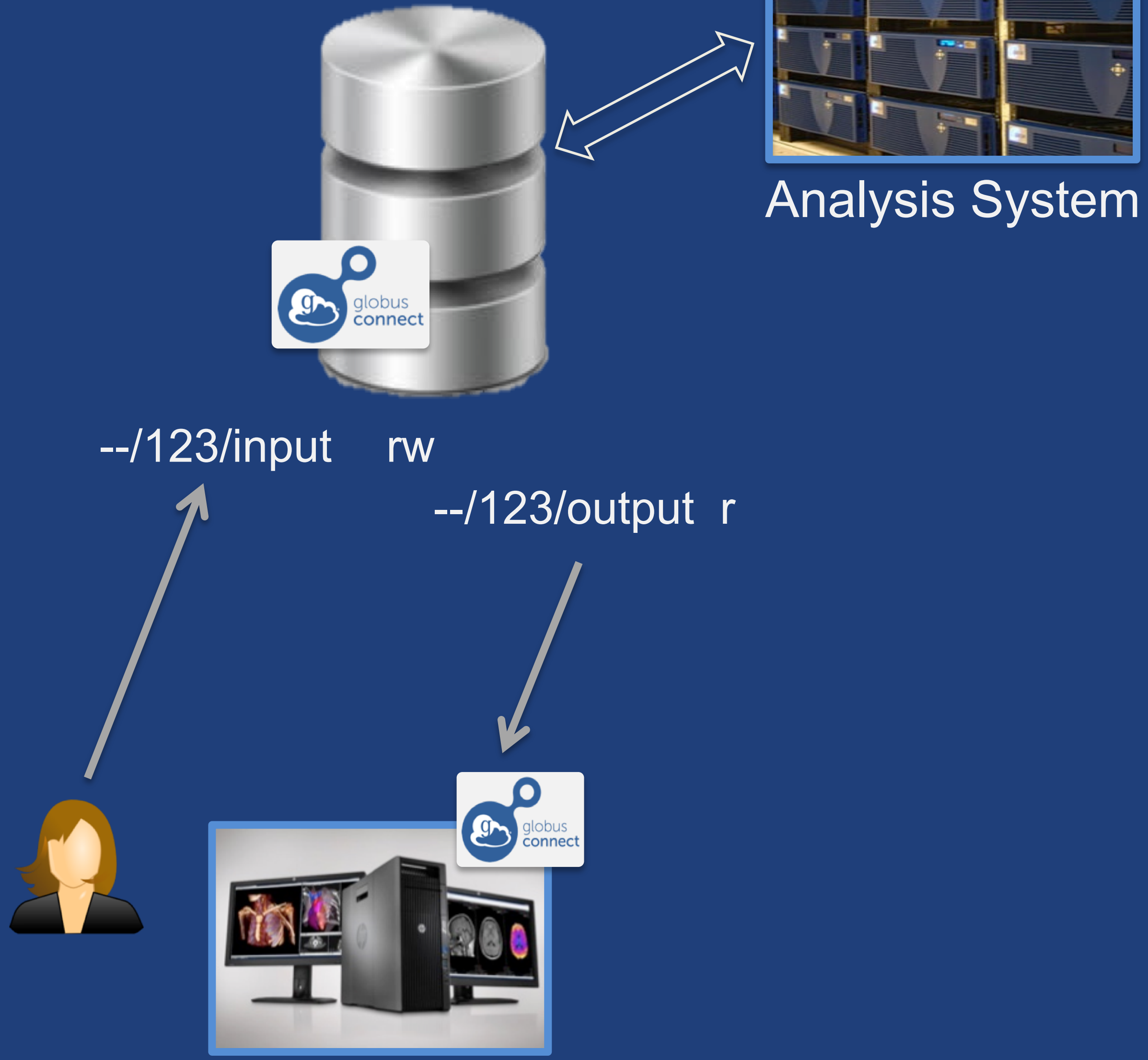
Source code, examples, papers, and other useful documentation.

Transfer data to destination

Search and request data of interest

Core center data processing

- Allow user to securely upload data for analysis
- Make analysis results available to user
- Automate setup and tear down of folders and permissions



Analysis System

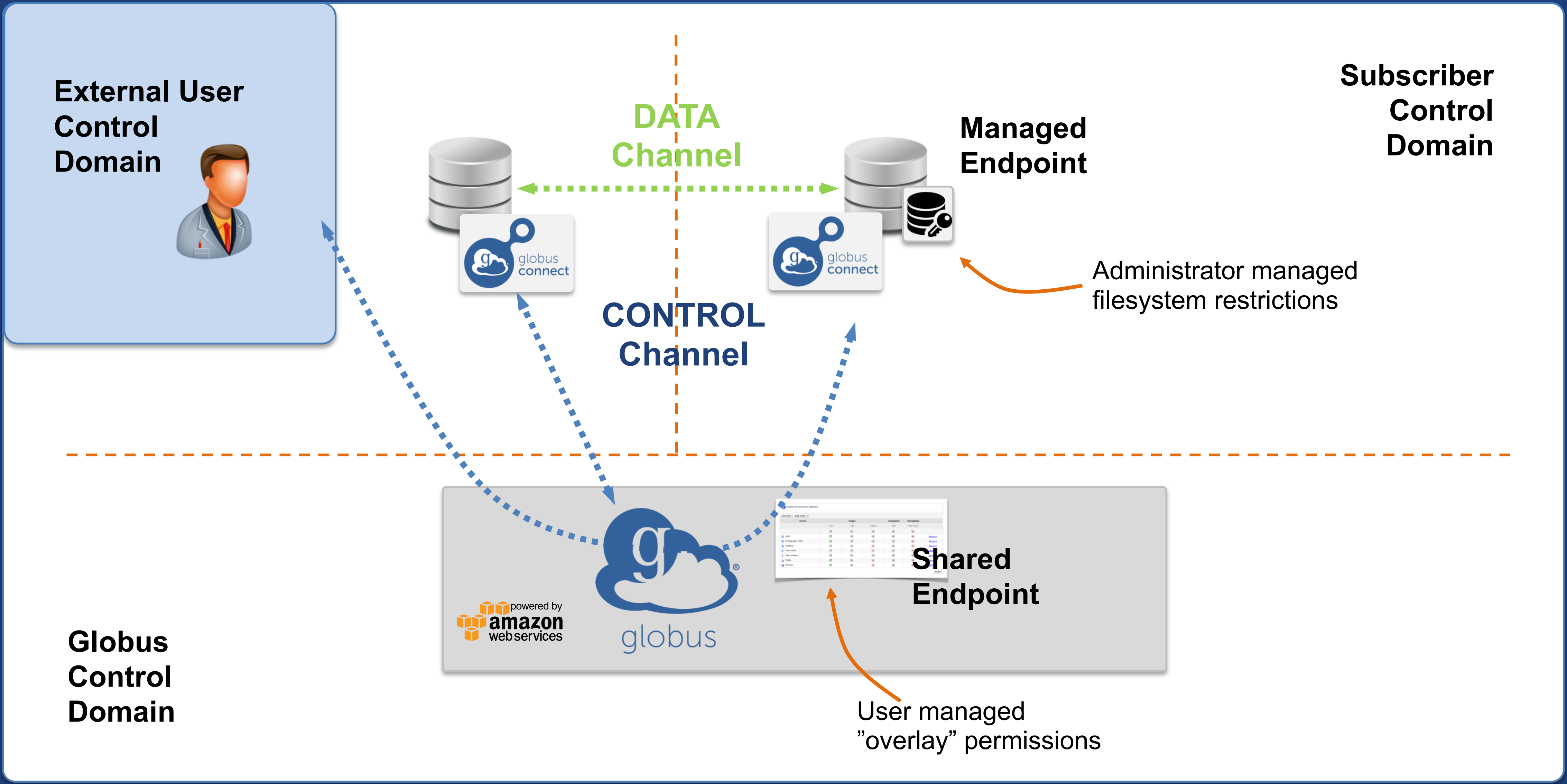
--/123/input rw

--/123/output r

Common solution components

- Shared endpoint for “staging” data
- Application that manages permissions
- Data transfer, to/from shared endpoint

Conceptual architecture: Sharing



External User Control Domain

DATA Channel

Managed Endpoint

Subscriber Control Domain

Administrator managed filesystem restrictions

CONTROL Channel

Shared Endpoint

User managed ”overlay” permissions

Globus Control Domain

Data sharing features

- **Shared endpoint creation requires user authentication**
  - Cannot be completely automated
  - Must be a managed endpoint
- **Roles for management of endpoint and tasks**
  - Grant rights to other users, groups or applications
- **Access manager role grants others the rights to manage permissions**
  - Grant to users, groups, applications

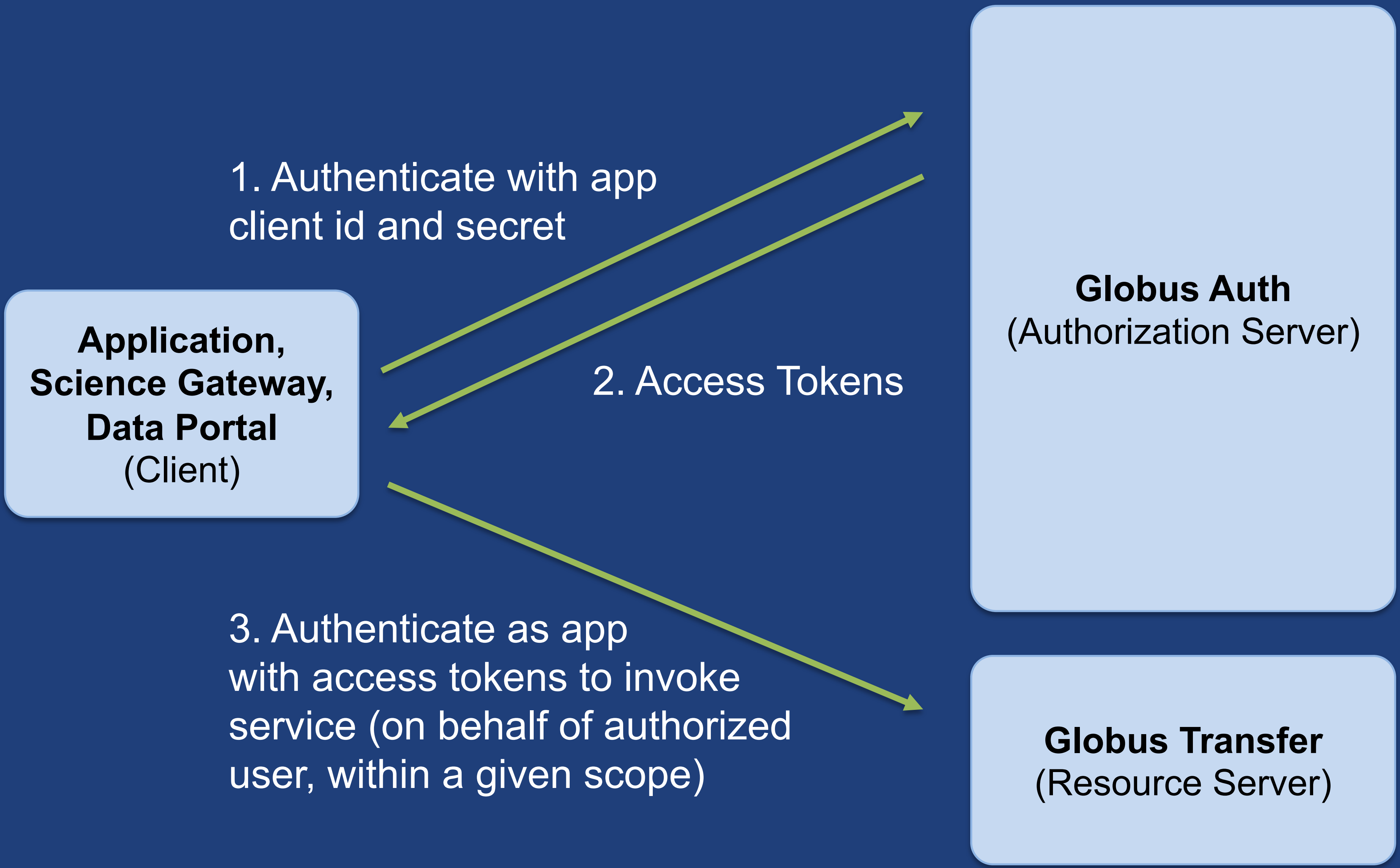
Data sharing permissions management

- **Permissions are set per folder, on a shared endpoint**
- **Permissions management can be automated**
- **For a user**
  - Identity: user must log in with this
  - Email: user gets a code via email; link to their Globus Account
- **For a group**
  - Group UUID: search for group to get UUID
  - Access governed by membership in the group
- **For an application**
  - Application identity: appclientid@clients.auth.globus.org (apps are people too!)

Application concepts

- **Custom app (automatically) manages permissions**
  - Can use Globus CLI (more in next presentation)
- **Confidential apps: use client id and secret**
  - Ensure application is on a secure device
  - Set up policy for rotation of secret
  - Identity: appclientid@clients.auth.globus.org

Client credential grant



1. Authenticate with app client id and secret

Application, Science Gateway, Data Portal
(Client)

Globus Auth
(Authorization Server)

2. Access Tokens

3. Authenticate as app with access tokens to invoke service (on behalf of authorized user, within a given scope)

Globus Transfer
(Resource Server)

Data transfer scenarios

- **Application moving data of its own accord**
  - App has access to source data and can write to destination
  - Requires shared endpoints on both sides (why?)
  - Client credential grant
- **Application moving data as user**
  - Only user has access to data on source/destination
  - Authorization code grant
  - Similar to the data portal example presented earlier

Walkthrough

What: Make select data available to authorized user(s)

Who: Data distribution application

How:
1. Creates folder on shared endpoint
2. Moves data to folder
3. Sets permissions on folder for user/group

See example code at:
github.com/globus/automation-examples/blob/master/share_data.py

On your EC2 instance in ~/automation-examples

Application registration

- **Register the application at** developers.globus.org
  - Redirects: `https://auth.globus.org/v2/web/auth-code`
  - Scopes: `globus:auth:scope:transfer.api.globus.org:all`
- **Get client id and secret**
- **Add client id to the app**

Shared endpoint configuration

- **Create at top level folder**
- **Set access manager role for app to manage access permissions**
- **Optionally…**
  - Set endpoint administrator role (can change endpoint definition)
  - Set endpoint manager role (can monitor and manage tasks)
  - Set endpoint monitor role (can monitor tasks)

Application access

- **Use client credential grant to authenticate as app**
  - Client id and secret used for obtaining tokens
  - Identity username is appclientid@clients.auth.globus.org
- **Create a folder for user (or project)**
- **Set permissions on folder (user/group)**
- **Create transfer task to move data to folder**
- **(Optionally) notify user(s) that data is available**

Support resources

- **Globus documentation: docs.globus.org**
- **Sample code: github.com/globus**
- **Helpdesk and issue escalation: support@globus.org**
- **Customer engagement team**
- **Globus professional services team**
  - Assist with portal/gateway/app architecture and design
  - Develop custom applications that leverage the Globus platform
  - Advise on customized deployment and integration scenarios

Join the Globus community

- Access the service: **globus.org/login**
- Create a personal endpoint: **globus.org/app/endpoints/create-gcp**
- Documentation: **docs.globus.org**
- Engage: **globus.org/mailing-lists**
- Subscribe: **globus.org/subscriptions**
- Need help? **support@globus.org**
- Follow us: **@globusonline**