Best Practices for Data Sharing

Rachana Ananthakrishnan - **rachana@globus.org**

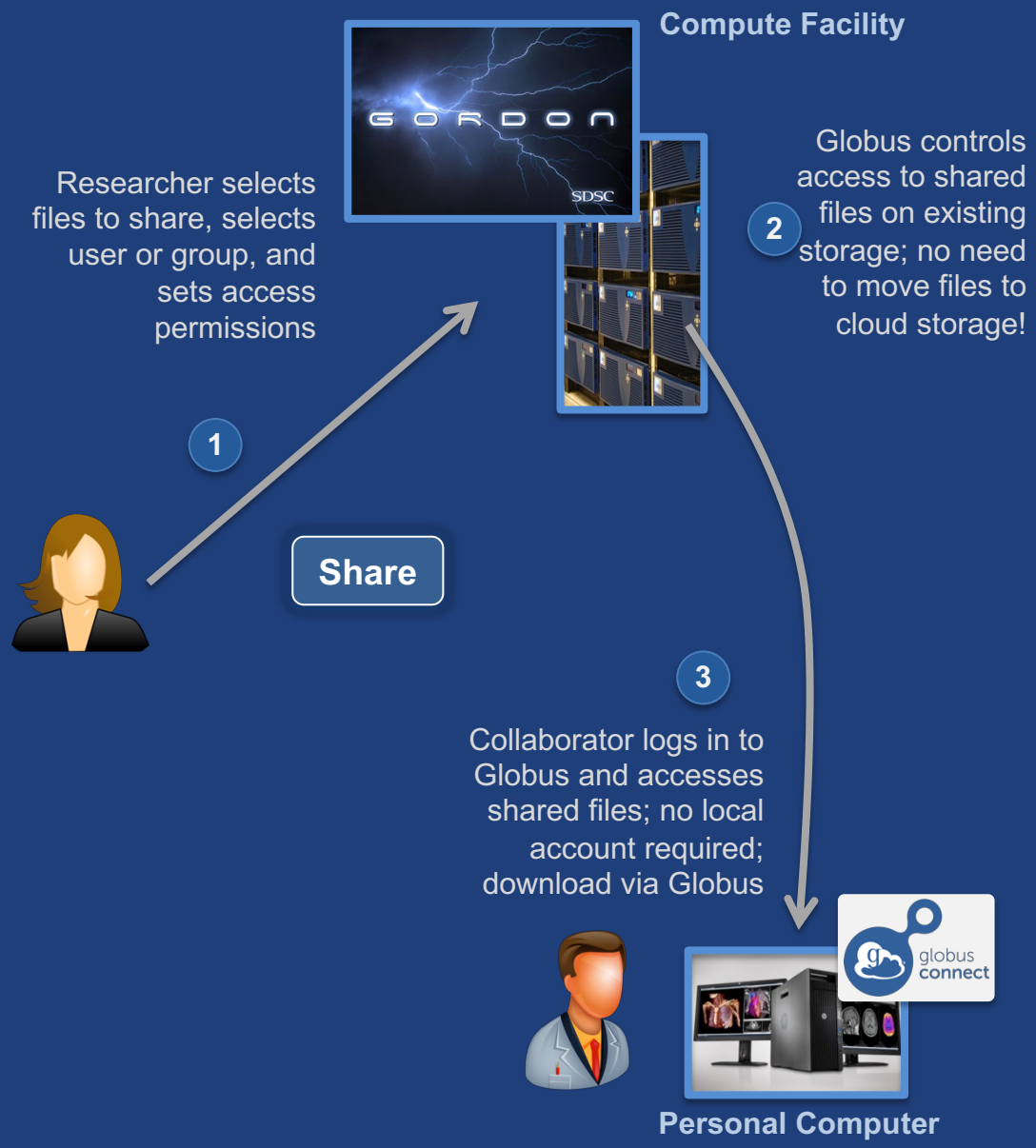
Greg Nawrocki - **greg@globus.org**

Johns Hopkins University
April 11, 2019

globus

Ad hoc data sharing

- **Individual users share data with collaborators**
- **Using a known email or identity for user/group**
- **Make data publicly – at least to any logged in Globus user - available**



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

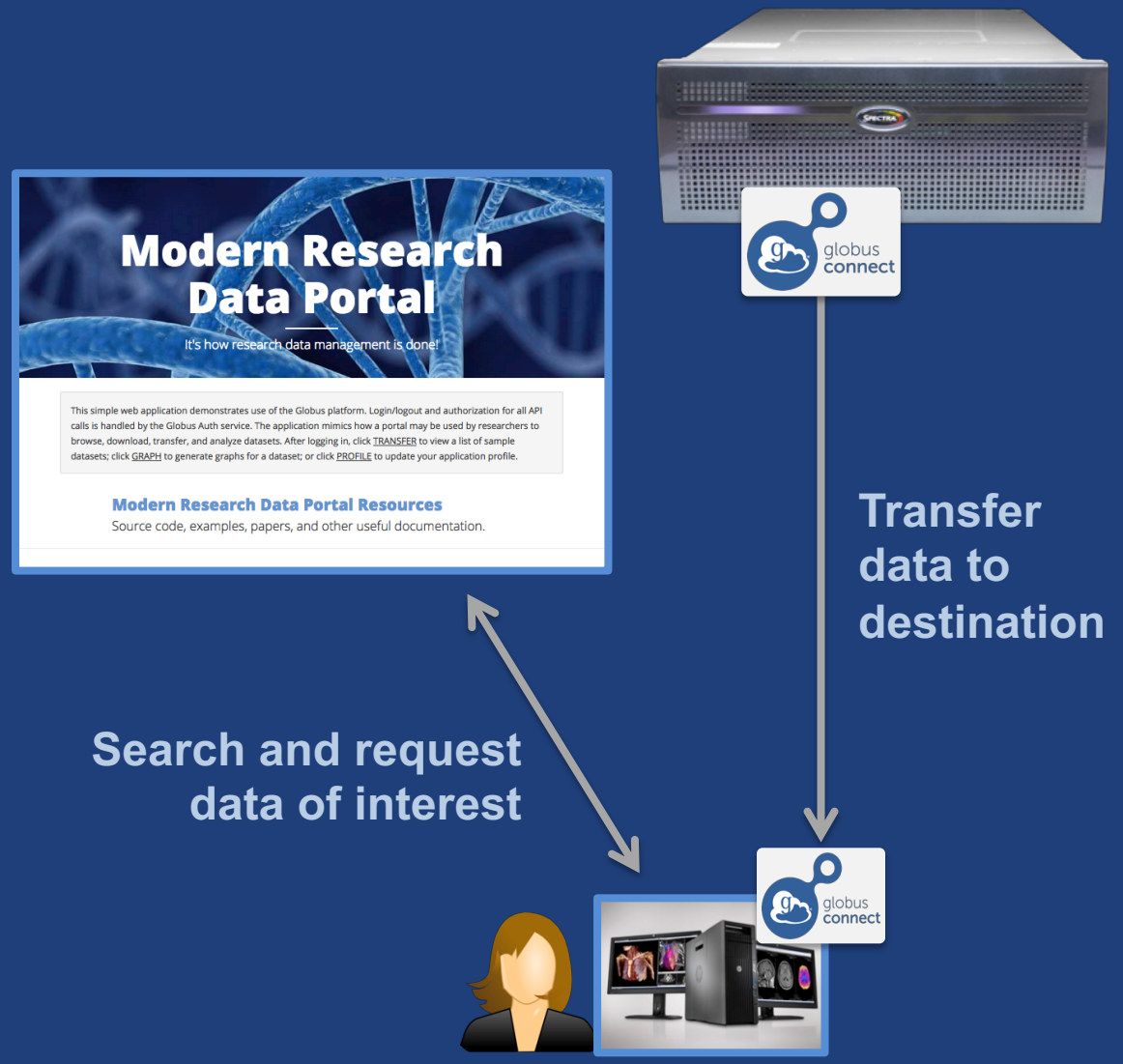
Local policy store

Personal Computer

Remote visualization/analysis

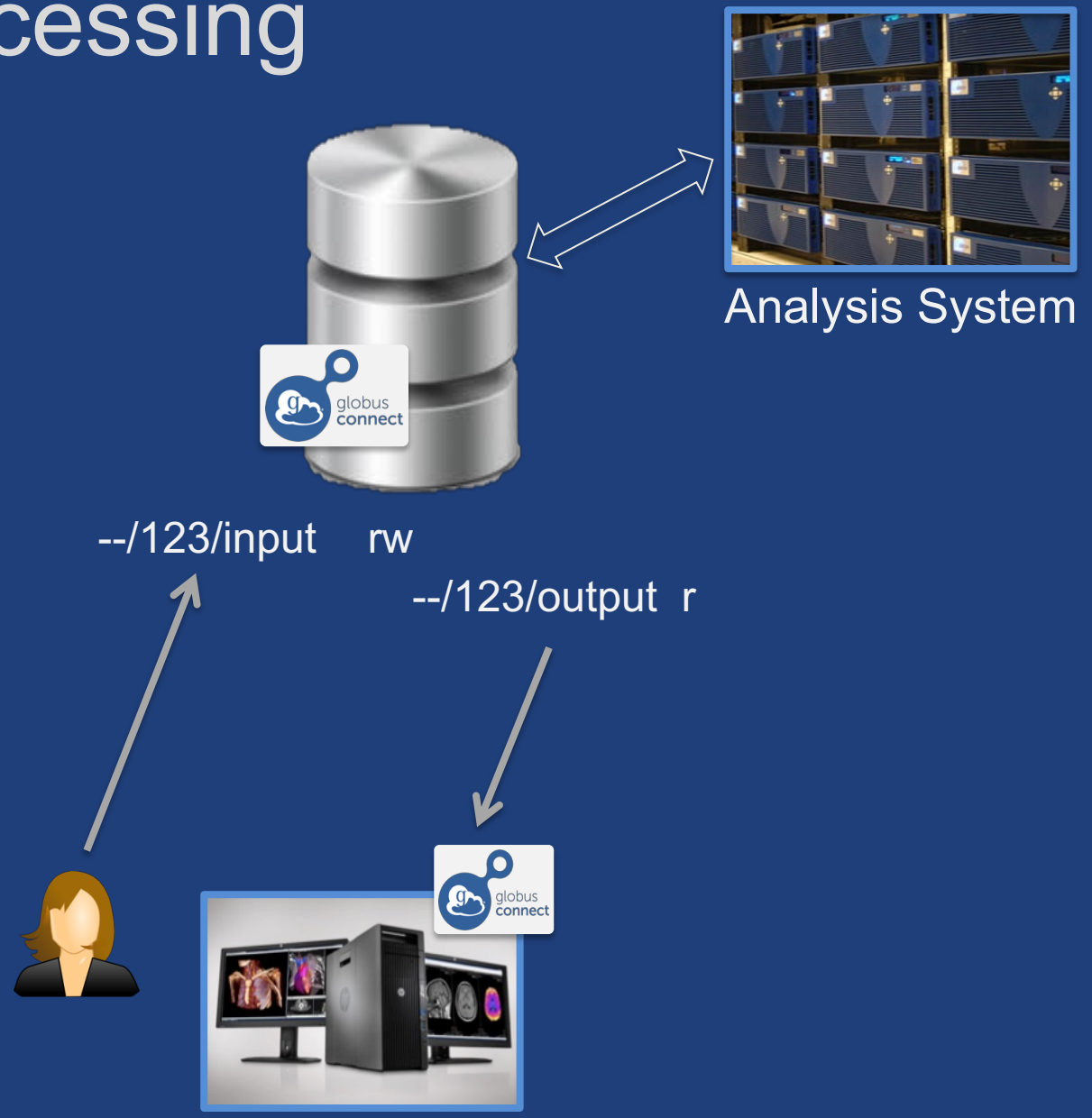
Data from provider/archive

- **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**



Core center data processing

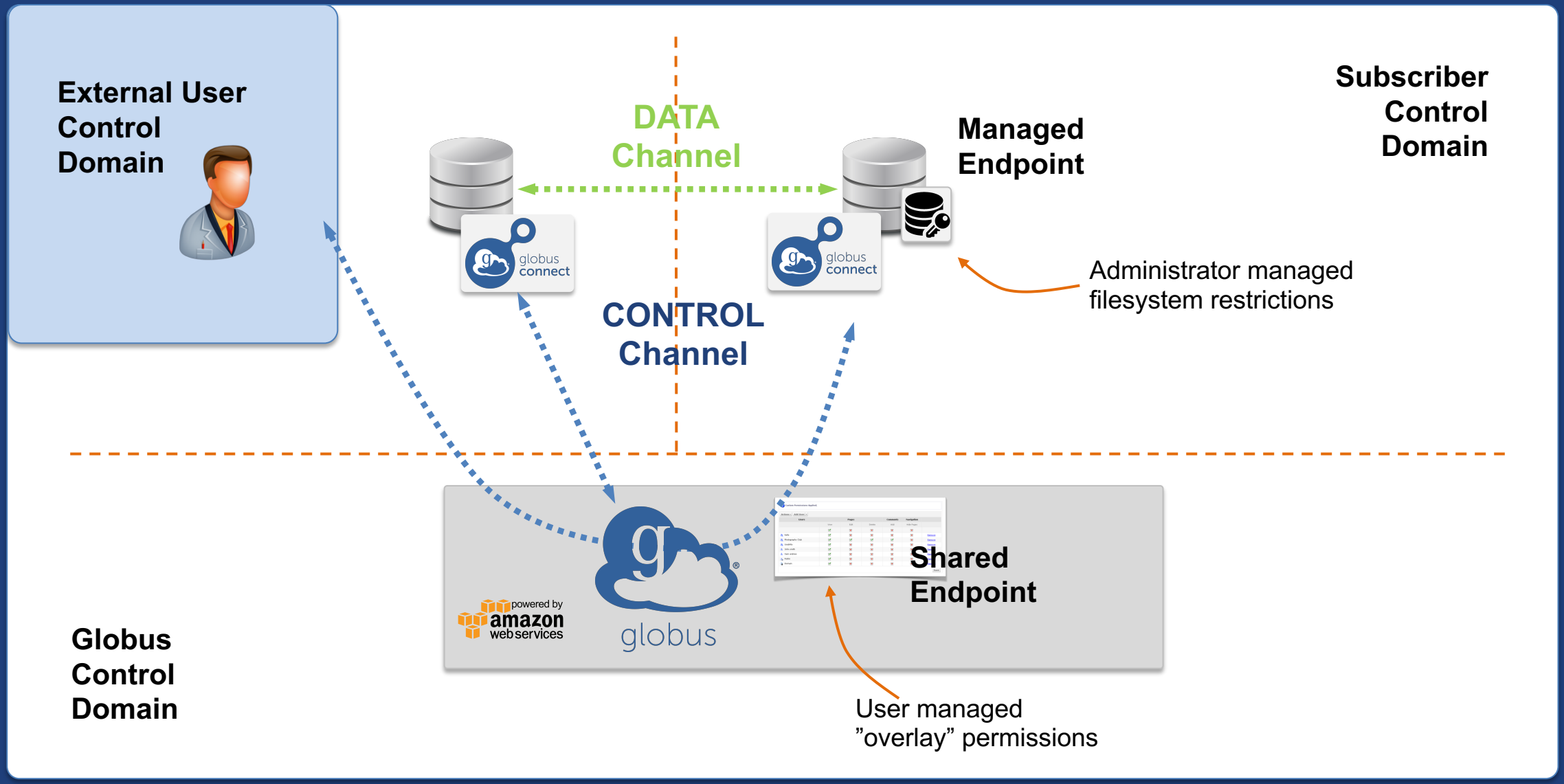
- **Allow user to securely upload data for analysis**
- **Make analysis results available to user**
- **Automate setup and tear down of folders and permissions**



Common solution components

- **Shared endpoint for “staging” data**
  - Globus Connect
- **Application that manages permissions**
  - The notification comes for free!
- **Data transfer, to and from shared endpoint**

Conceptual architecture: Sharing



Data sharing features

- **Shared endpoint creation requires authentication**
  - Cannot be completely automated – must ”log in”
  - 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

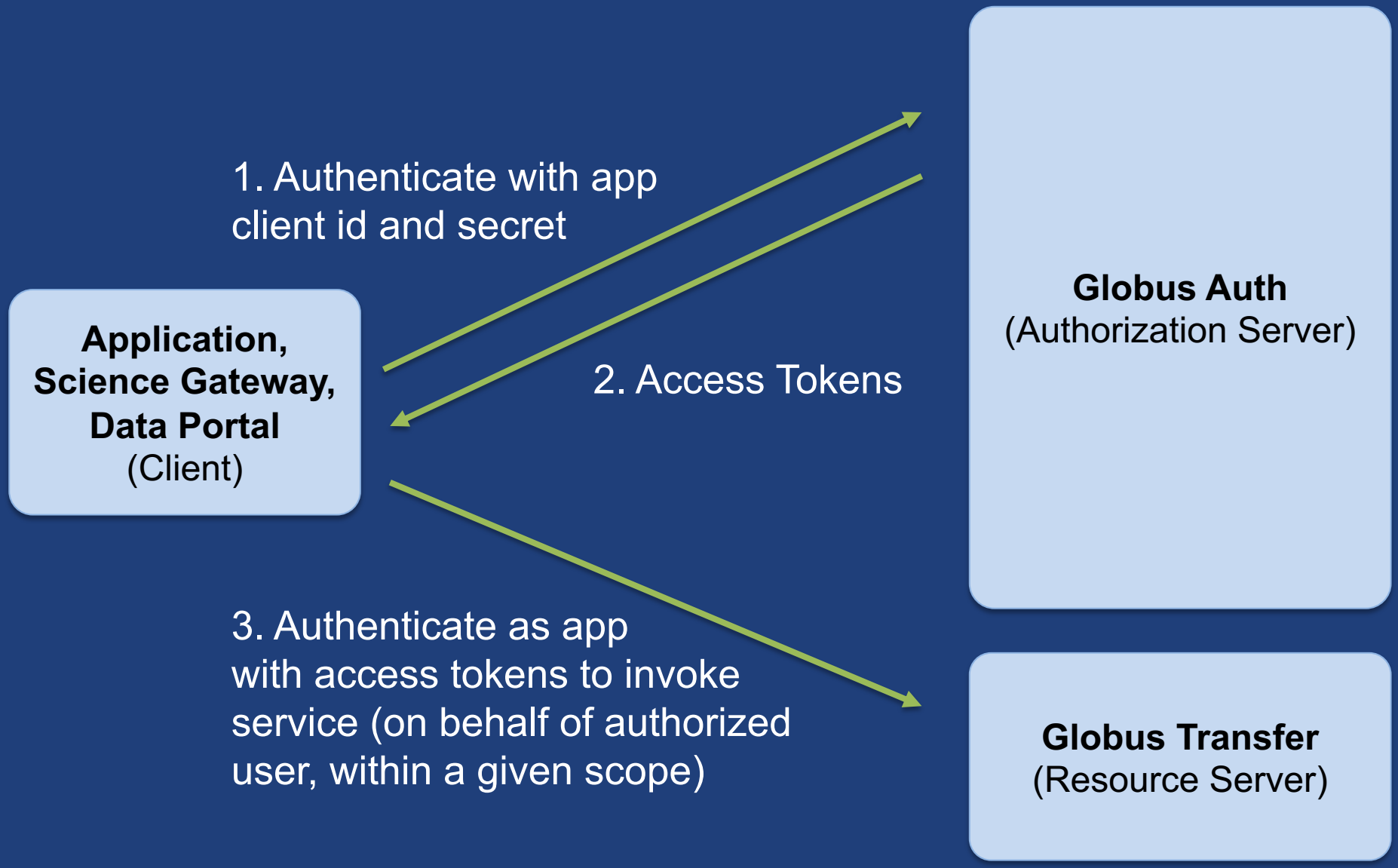
Data sharing permissions management

- **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

Application concepts

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

Client credential grant



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
  - 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 (yes the app is a 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

- **To make the confidential client grant work**
- **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**
- **Helpdesk and issue escalation: support@globus.org**
- **Mailing lists**
  - https://www.globus.org/mailing-lists
  - developer-discuss@globus.org
- **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**