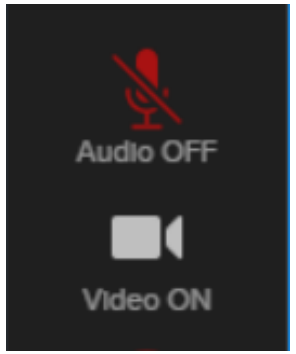
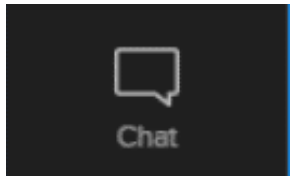
Welcome to the SGCI Webinar!

- We will be starting shortly.
- Your audio has been muted, and you are encouraged to turn off your video during the presentation.
  - Controls for these are near the bottom of the right-side control panel for BlueJeans.



- You may submit questions at any time using Chat, and the moderator will share them with the presenter when appropriate.



- This presentation will be recorded and slides will be posted.

Building the Modern Research Data Portal

Introduction to the Globus Platform

Steve Tuecke
tuecke@globus.org

Thank you to our sponsors!

U.S. DEPARTMENT OF
ENERGY

THE UNIVERSITY OF
CHICAGO

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Argonne
NATIONAL LABORATORY

Thank you, our users...

5 major services	**245 PB** transferred	**30 Bn** files processed	**45,000** registered users
13 national labs use Globus	**10,000** active endpoints	**10,000** active users/year	**99.9%** uptime
60+ institutional subscribers	**1 PB** largest single transfer to date	**3 months** longest continuously managed transfer	**130** federated campus identities

...and thank YOU, our subscribers!

NERSC

NEW YORK UNIVERSITY

MICHIGAN

Berkeley
UNIVERSITY OF CALIFORNIA

wellcome trust sanger institute

NIH

JOHNS HOPKINS
UNIVERSITY

CORNELL
UNIVERSITY

NCAR

UF | UNIVERSITY of FLORIDA

INDIANA UNIVERSITY

MICHIGAN STATE
UNIVERSITY

Yale

THE UNIVERSITY OF CHICAGO

Los Alamos
NATIONAL LABORATORY
EST. 1943

NCSA

syngenta

NIST

VirginiaTech
Invent the Future

Argonne
NATIONAL LABORATORY

https://www.globusworld.org/

docs.globus.org

https://docs.globus.org

globus docs

APIs ▾ How To Guides ▾ Support ▾

Harness the power of the Globus research data management cloud.

Transfer API

The Transfer API provides a REST-style interface to the Globus reliable file transfer service. The API can be used to monitor the progress of file transfers, manage file transfer endpoints, list remote directories, and submit new transfer and delete tasks.

LEARN MORE

Resource Providers

Globus allows you, as a resource provider, to easily offer reliable, secure, high-performance research data management capabilities to your users and their collaborators, directly from your own storage infrastructure.

LEARN MORE

Toolkit

The open source Globus Toolkit is a fundamental enabling technology for the "Grid," allowing users to access high-performance computing resources securely across corporate, institutional, and geographic boundaries without sacrificing local autonomy.

LEARN MORE

Cloud has transformed how software and platforms are delivered

Software as a service: **SaaS**

(web & mobile apps)

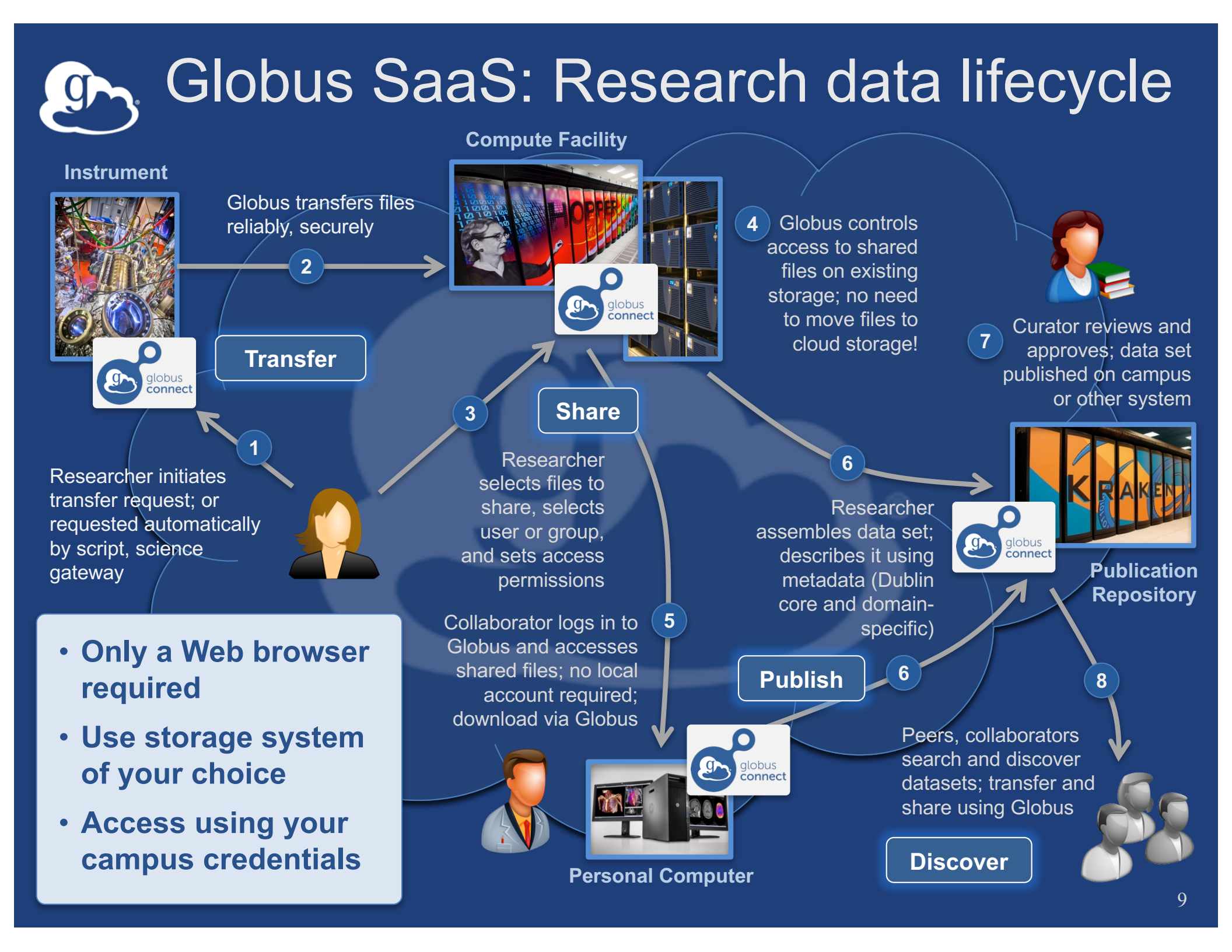
NETFLIX

Platform as a service: **PaaS**

Infrastructure as a service: **IaaS**

PaaS enables more rapid, cheap, and scalable delivery of powerful (SaaS) apps

Globus SaaS: Research data lifecycle

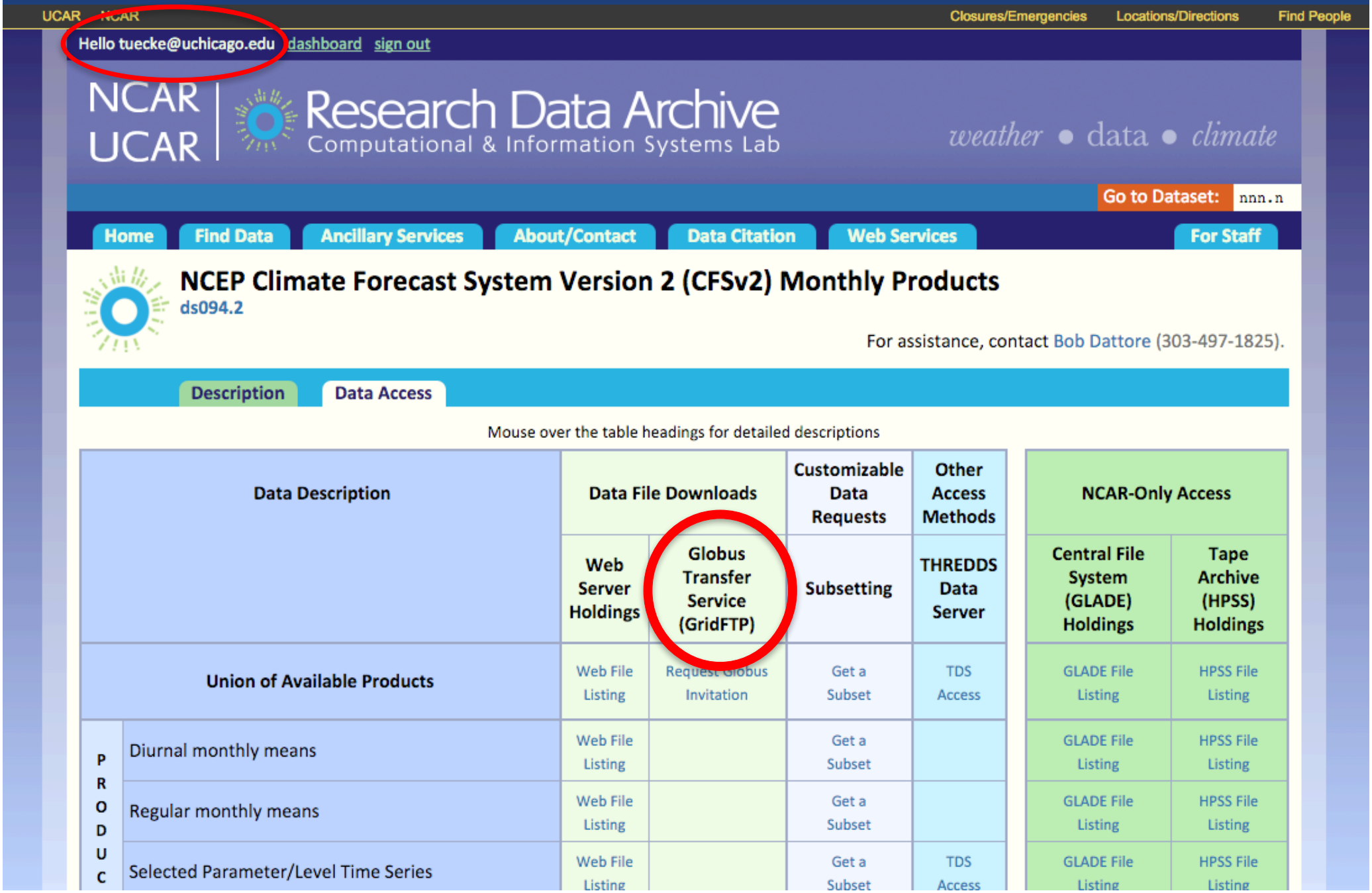


Platform Questions

- **How do you leverage Globus services in your own applications?**
- **How do you extend Globus with your own services?**

- **How do we empower the research community to create an integrated ecosystem of services and applications?**

Example: NCAR RDA



UCAR NCAR Closures/Emergencies Locations/Directions Find People

Hello tuecke@uchicago.edu dashboard sign out

NCAR UCAR | Research Data Archive
Computational & Information Systems Lab

weather • data • climate

Go to Dataset: nnn.n

Home | Find Data | Ancillary Services | About/Contact | Data Citation | Web Services | For Staff

NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products

ds094.2

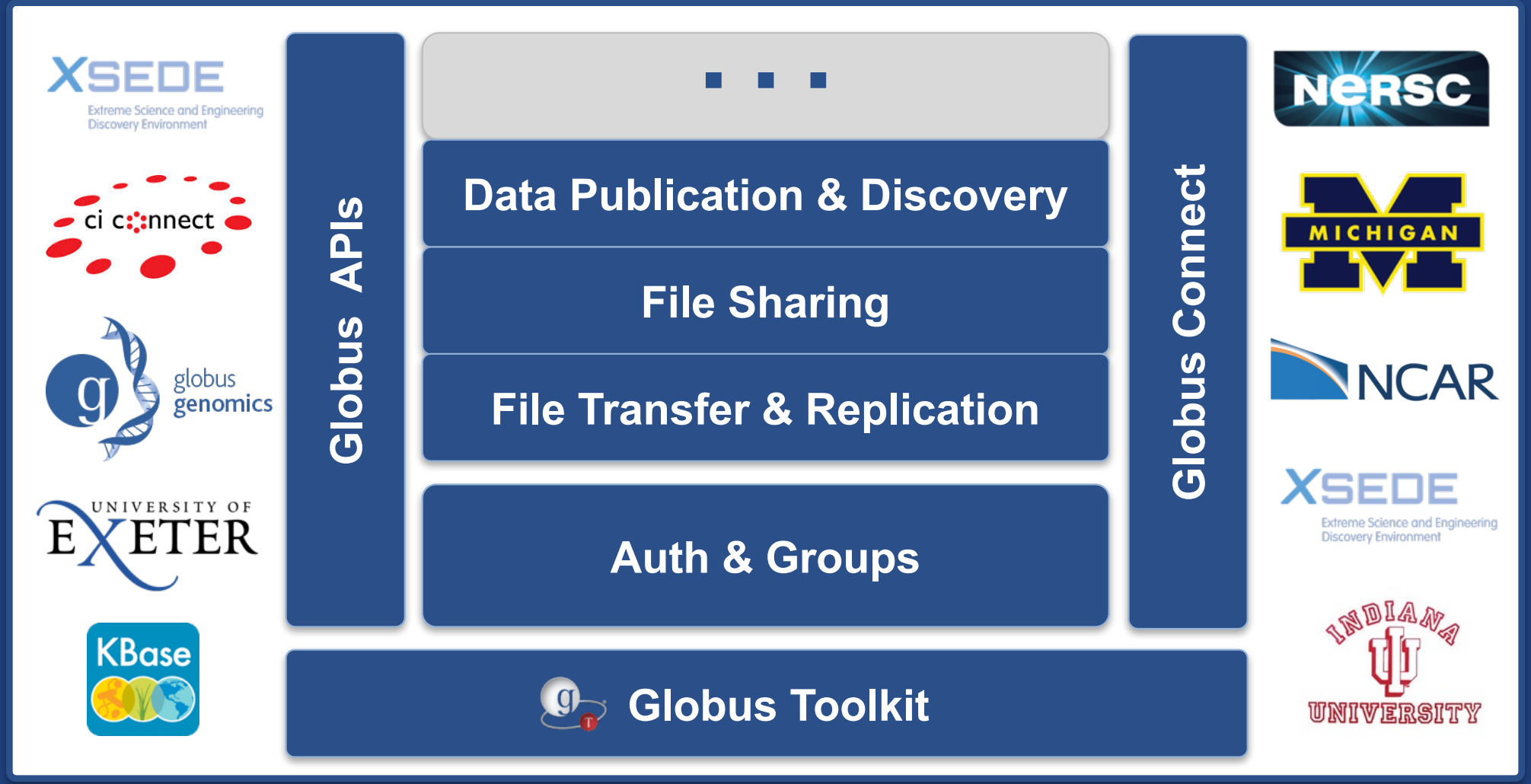
For assistance, contact Bob Dattore (303-497-1825).

Description | Data Access

Mouse over the table headings for detailed descriptions

Data Description		Data File Downloads		Customizable Data Requests	Other Access Methods	NCAR-Only Access	
		Web Server Holdings	Globus Transfer Service (GridFTP)	Subsetting	THREDDS Data Server	Central File System (GLADE) Holdings	Tape Archive (HPSS) Holdings
Union of Available Products		Web File Listing	Request Globus Invitation	Get a Subset	TDS Access	GLADE File Listing	HPSS File Listing
PRODUC	Diurnal monthly means	Web File Listing		Get a Subset		GLADE File Listing	HPSS File Listing
	Regular monthly means	Web File Listing		Get a Subset		GLADE File Listing	HPSS File Listing
	Selected Parameter/Level Time Series	Web File Listing		Get a Subset	TDS Access	GLADE File Listing	HPSS File Listing

Globus PaaS

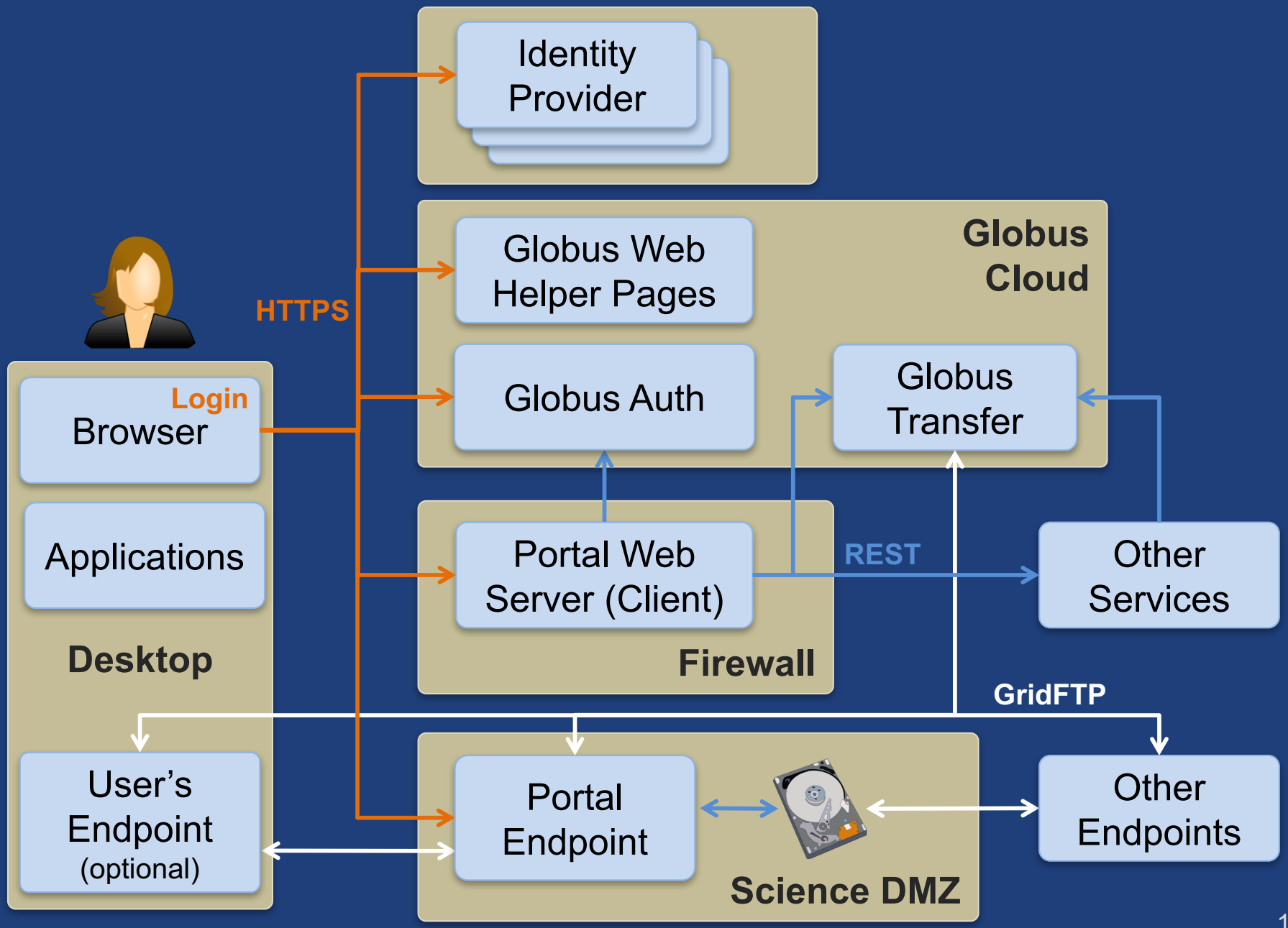

XSEDE
Extreme Science and Engineering Discovery Environment
ci connect
globus genomics
UNIVERSITY OF EXETER
KBase
Globus APIs
. . .
Data Publication & Discovery
File Sharing
File Transfer & Replication
Auth & Groups
Globus Connect
Globus Toolkit
NERSC
MICHIGAN
NCAR
XSEDE
Extreme Science and Engineering Discovery Environment
INDIANA UNIVERSITY

Demo

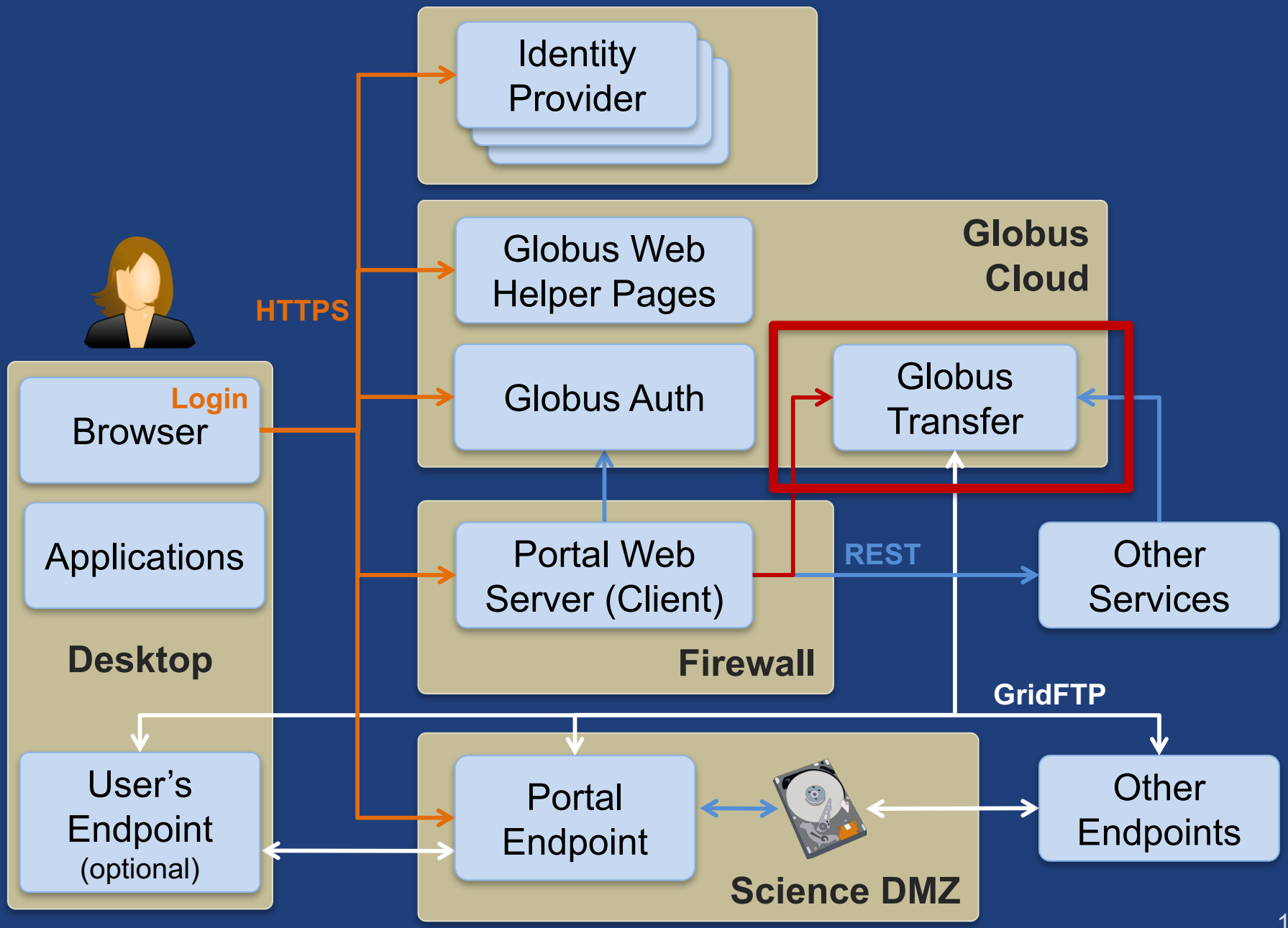
Sample Research Data Portal

github.com/globus/
globus-sample-data-portal

Prototypical research data portal



Prototypical research data portal



Globus Transfer API

- **Nearly all Globus Web App functionality implemented via public Transfer API**

docs.globus.org/api/transfer

Globus Python SDK

- **Python client library for the Globus Auth and Transfer REST APIs**

globus.github.io/globus-sdk-python

- **Open source**

Endpoint Search

- **Plain text search for endpoint**
  - Searches owner, display name, keywords, description, organization, department
  - Full word and prefix match
- **Limit search to pre-defined scopes**
  - `all, my-endpoints, recently-used, in-use, shared-by-me, shared-with-me`
- **Returns: List of endpoint documents**

Endpoint Management

- **Get endpoint (by id)**
- **Update endpoint**
- **Create & delete (shared) endpoints**
- **Manage endpoint servers**

Endpoint Activation

- **Activating endpoint means binding a credential to an endpoint for login**
- **Globus Connect Server endpoint that have MyProxy or MyProxy OAuth identity provider require login via web**
- **Auto-activate**
  - Globus Connect Personal and shared endpoints use Globus-provided credential
  - An endpoint that shares an identity provider with another activated endpoint will use credential
- **Must auto-activate before any API calls to endpoints**

File operations

- **List directory contents (ls)**
- **Make directory (mkdir)**
- **Rename**
- **Note:**
  - Path encoding & UTF gotchas
  - Don’t forget to auto-activate first

Task submission

- **Asynchronous operations**
  - Transfer
    - Sync level option
  - Delete
- **Get `submission_id`, followed by submit**
  - Once and only once submission

Task management

- **Get task by id**
- **Get task_list**
- **Update task by id (label, deadline)**
- **Cancel task by id**
- **Get event list for task**
- **Get task pause info**

Bookmarks

- **Get list of bookmarks**
- **Create bookmark**
- **Get bookmark by id**
- **Update bookmark**
- **Delete bookmark by id**

- **Cannot perform other operations directly on bookmarks**
  - Requires client-side resolution

Shared endpoint access rules (ACLs)

- **Access manager role required to manage permission/ACLs**
- **Operations:**
  - Get list of access rules
  - Get access rule by id
  - Create access rule
  - Update access rule
  - Delete access rule

Management API

- **Allow endpoint administrators to monitor and manage all tasks with endpoint**
  - Task API is essentially the same as for users
  - Information limited to what they could see locally
- **Cancel tasks**
- **Pause rules**

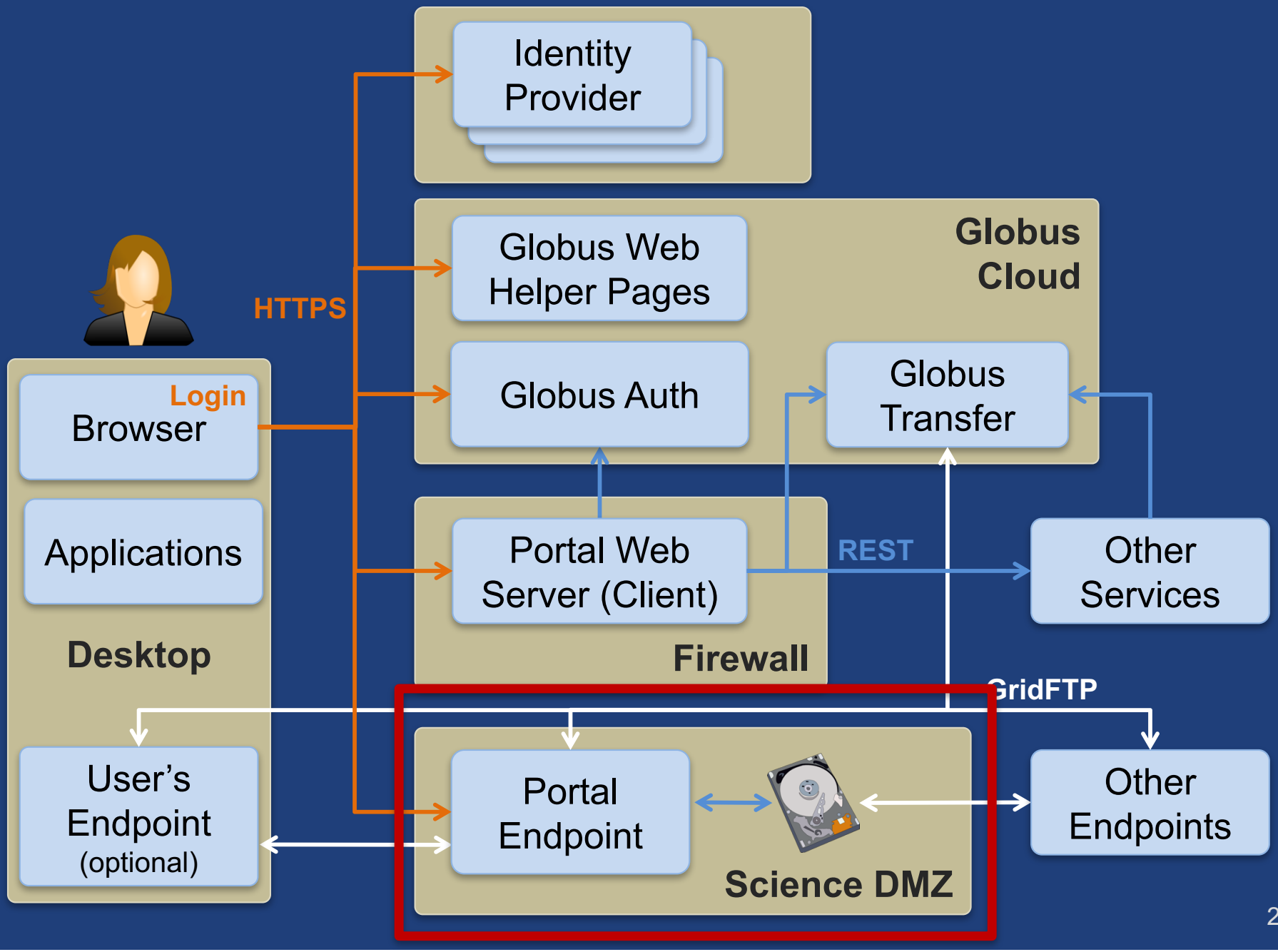
Python SDK Jupyter notebook

- **Jupyter (iPython) notebook demonstrating use of Python SDK**

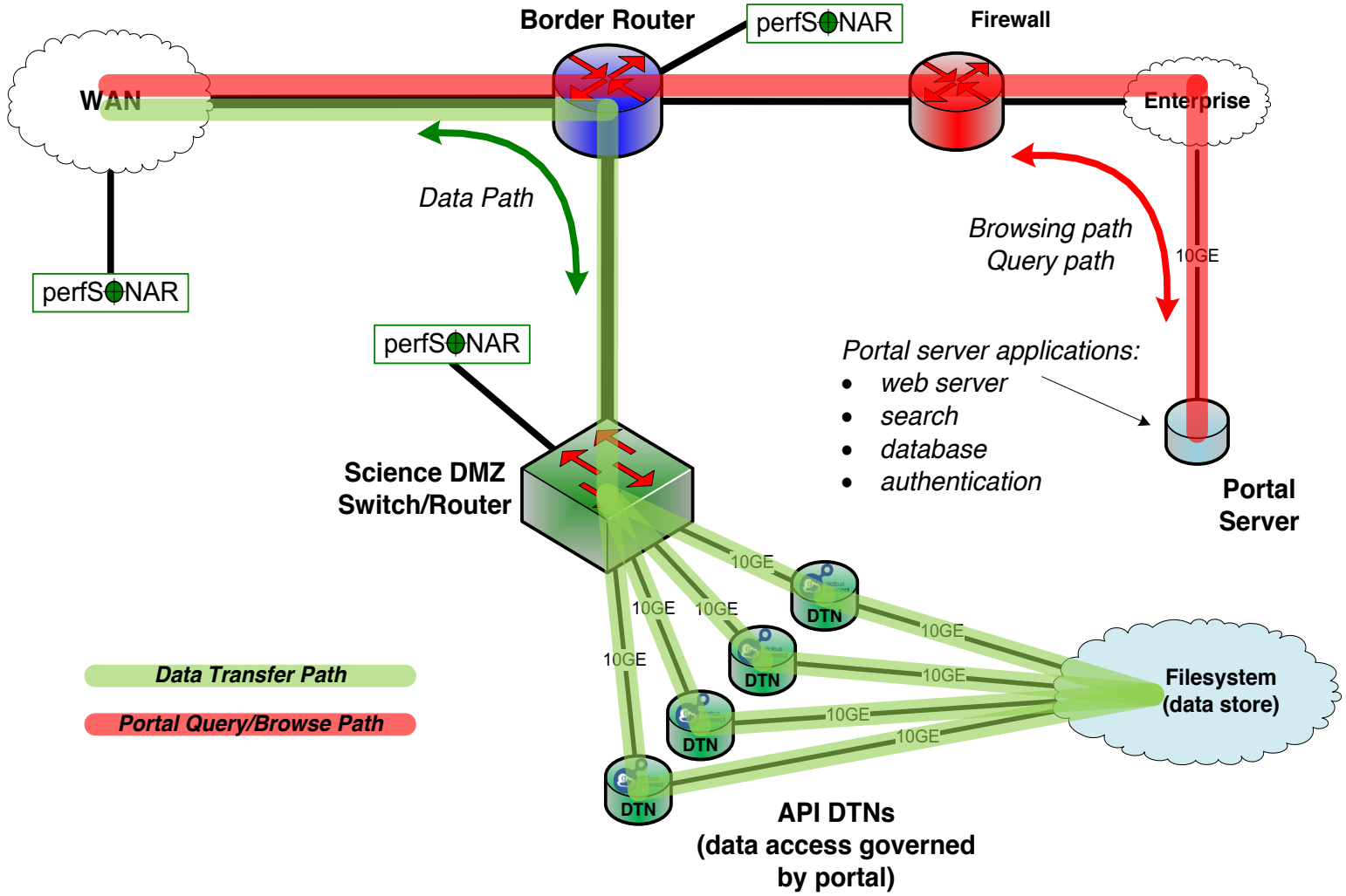
github.com/globus/globus-jupyter-notebooks

- **Open source**

Prototypical research data portal

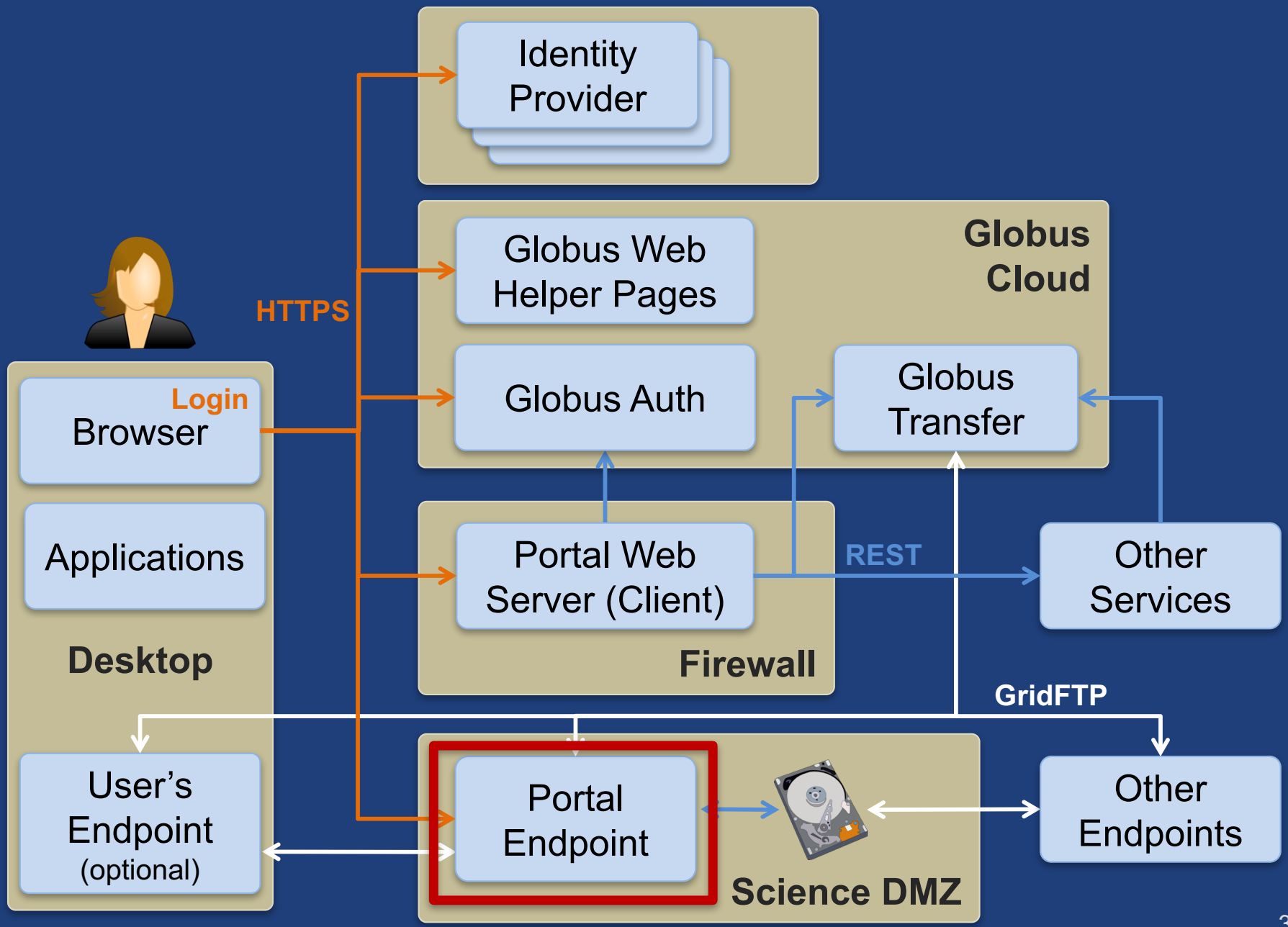


Next-Generation Portal Leverages Science DMZ



https://fasterdata.es.net/

Prototypical research data portal

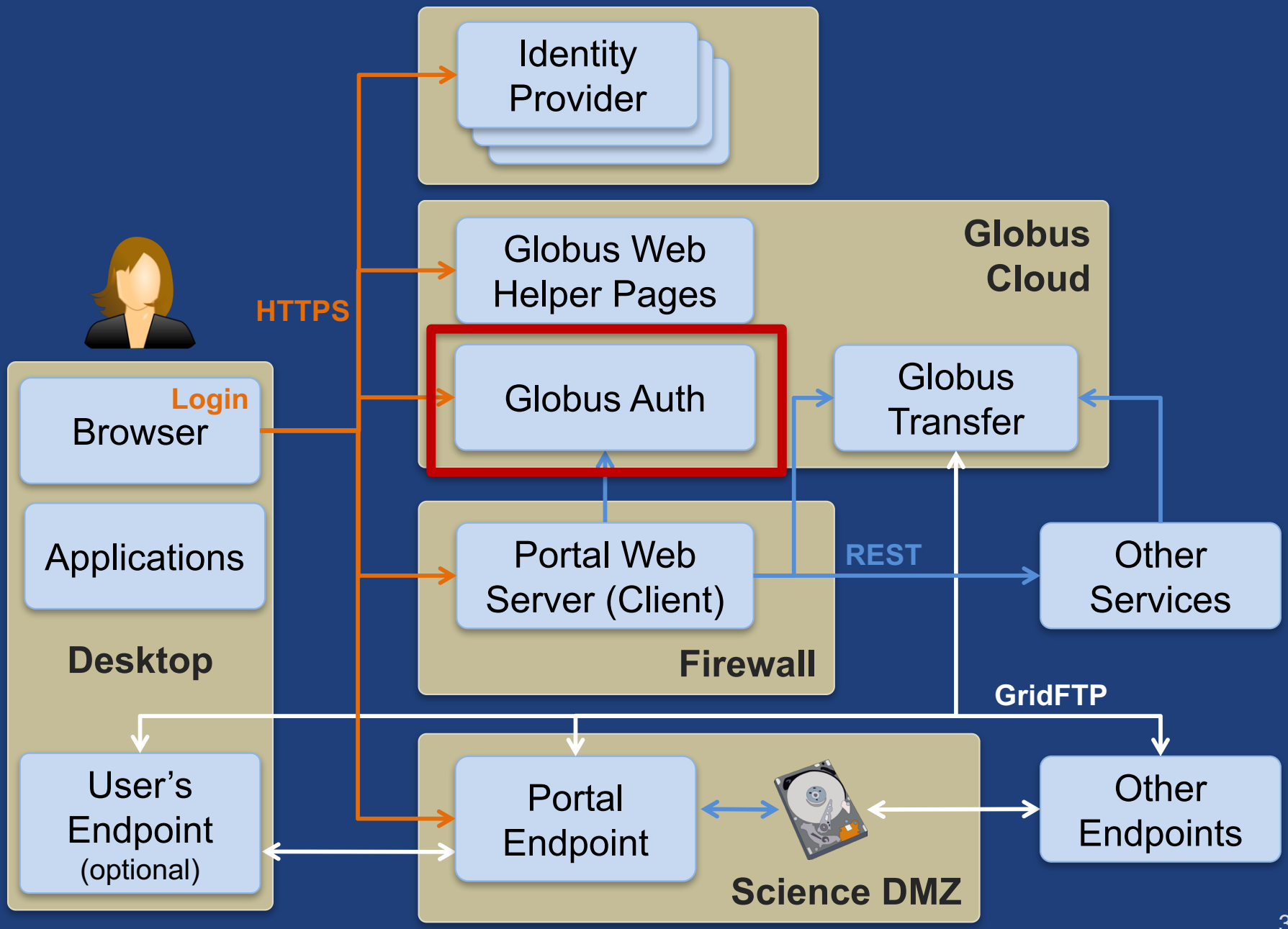


HTTPS to Endpoints

- **Each endpoint HTTPS server is a Globus Auth service (resource server)**
- **Web page can link to file on server**
  - Browser GET will cause HTTPS server to authorize request via Globus Auth (note SSO)
- **Portal (client) can request scope for endpoint resource server**
  - Use access token in requests

“A single global information space”

Prototypical research data portal



Challenge

- **How to provide:**
  - Login to apps
    - Web, mobile, desktop, command line
  - Protect all REST API communications
    - App → Globus service
    - App → non-Globus service
    - Service → service
- **While:**
  - Not introducing even more identities
  - Providing least privileges security model
  - Being agnostic to programming language and framework
  - Being web friendly
  - Making it easy for users and developers

Globus Auth

- **Foundational identity and access management (IAM) platform service**
- **Simplify creation and integration of advanced apps and services**
- **Brokers authentication and authorization interactions between:**
  - end-users
  - identity providers: InCommon, XSEDE, Google, portals
  - services: resource servers with REST APIs
  - apps: web, mobile, desktop, command line clients
  - services acting as clients to other services

Globus Auth

- **Identity and access management PaaS**

docs.globus.org/api/auth

- **Introduction**
- **Developer Guide**
- **Reference**

Based on widely used web standards

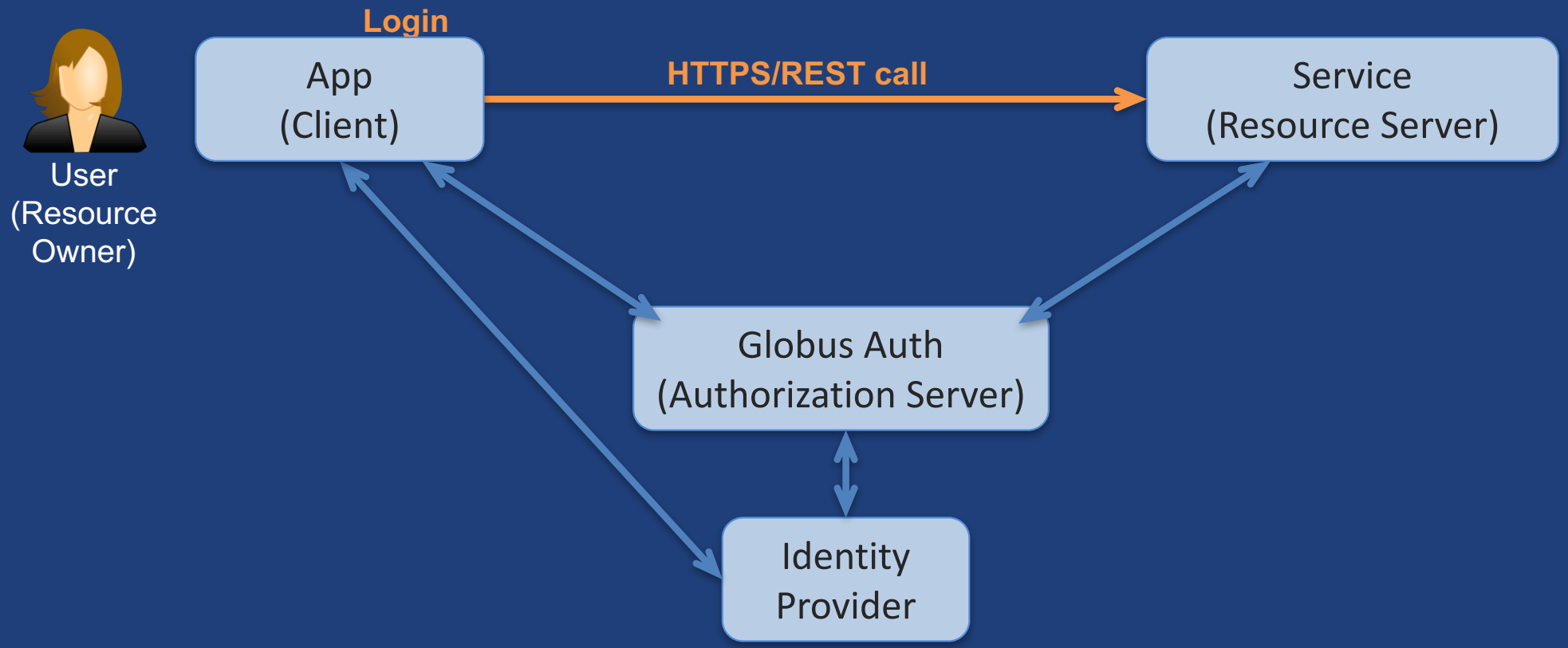
- **OAuth 2.0 Authorization Framework**
  - aka OAuth2
- **OpenID Connect Core 1.0**
  - aka OIDC

- **Use various OAuth2 and OIDC libraries**
  - Google OAuth Client Libraries (Java, Python, etc.), Apache mod_auth_openidc, etc.
  - Globus Python SDK

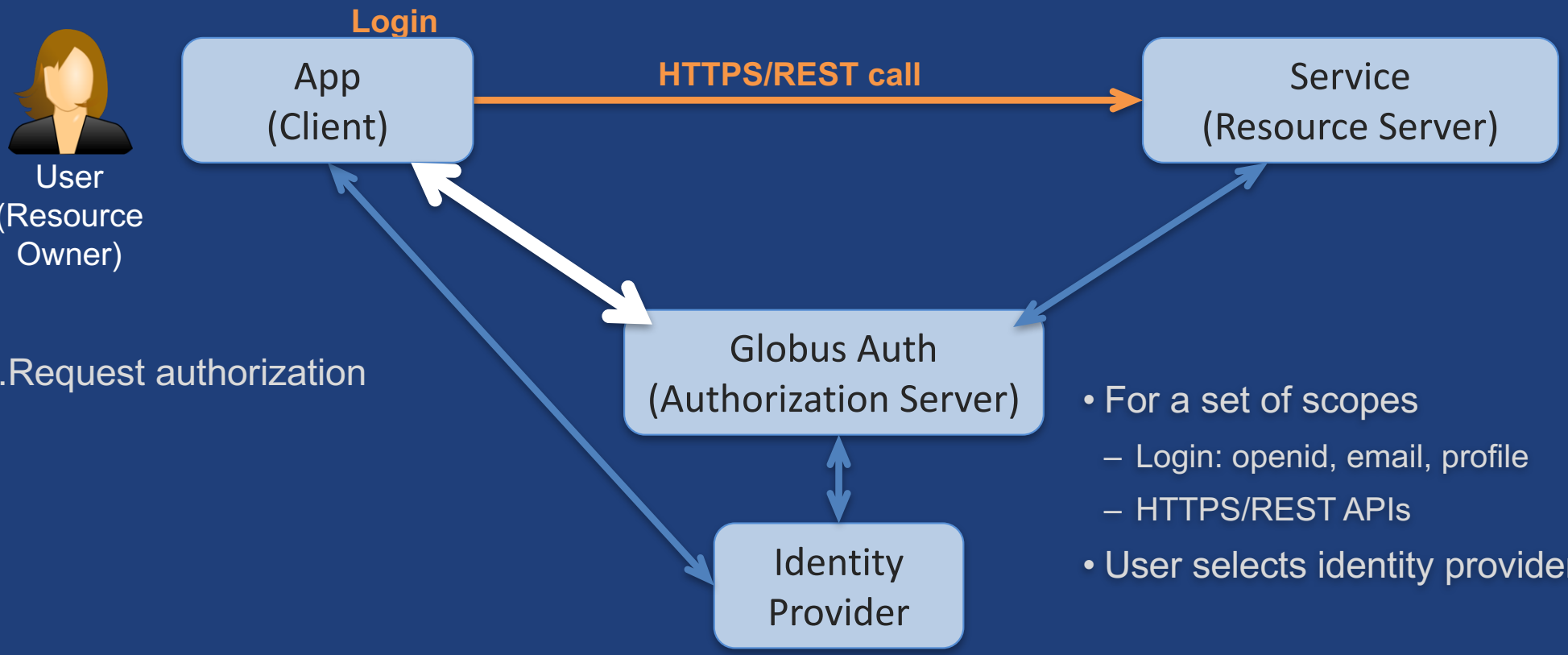
Globus account

- **A Globus account is a set of identities**
  - A *primary identity*
    - Identity can be primary of only one account
  - One or more *linked identities*
    - Identity can (currently) be linked to only one account
- **Account does not have own identifier**
  - An account is uniquely identified using its primary identity

Globus Auth interactions



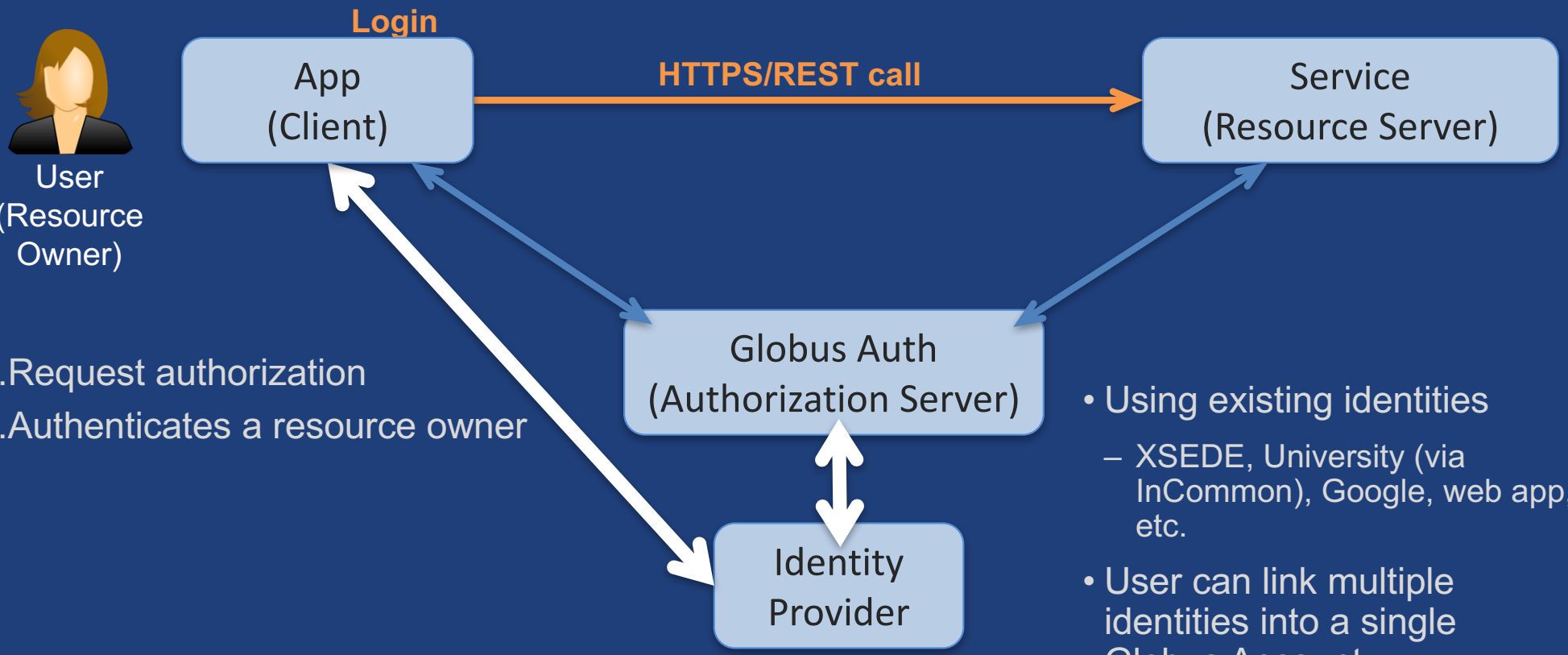
Globus Auth interactions



1.Request authorization

- For a set of scopes
  - Login: openid, email, profile
  - HTTPS/REST APIs
- User selects identity provider

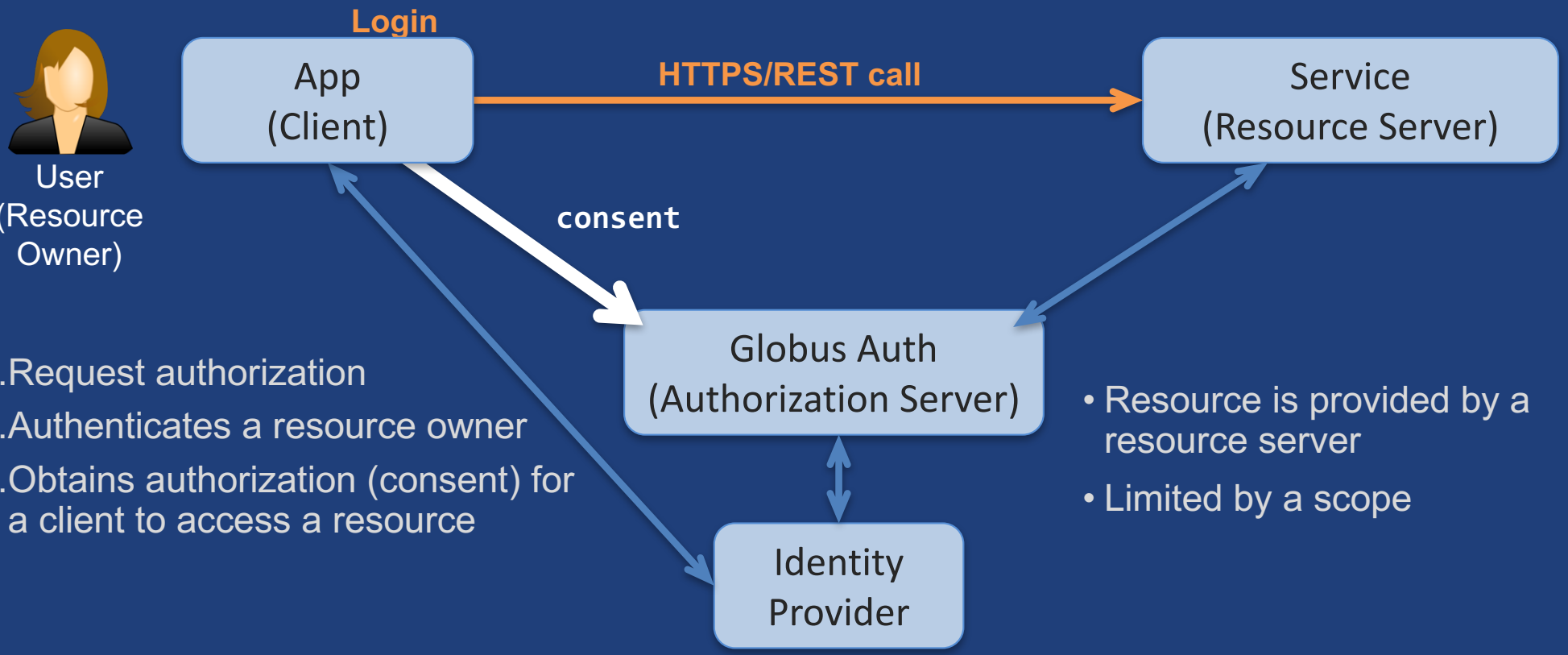
Globus Auth interactions



1. Request authorization
2. Authenticates a resource owner

- Using existing identities
  - XSEDE, University (via InCommon), Google, web app, etc.
- User can link multiple identities into a single Globus Account
- No Globus username (Globus ID) required
- Globus Auth handles naming details, e.g., ePPN vs ePTID

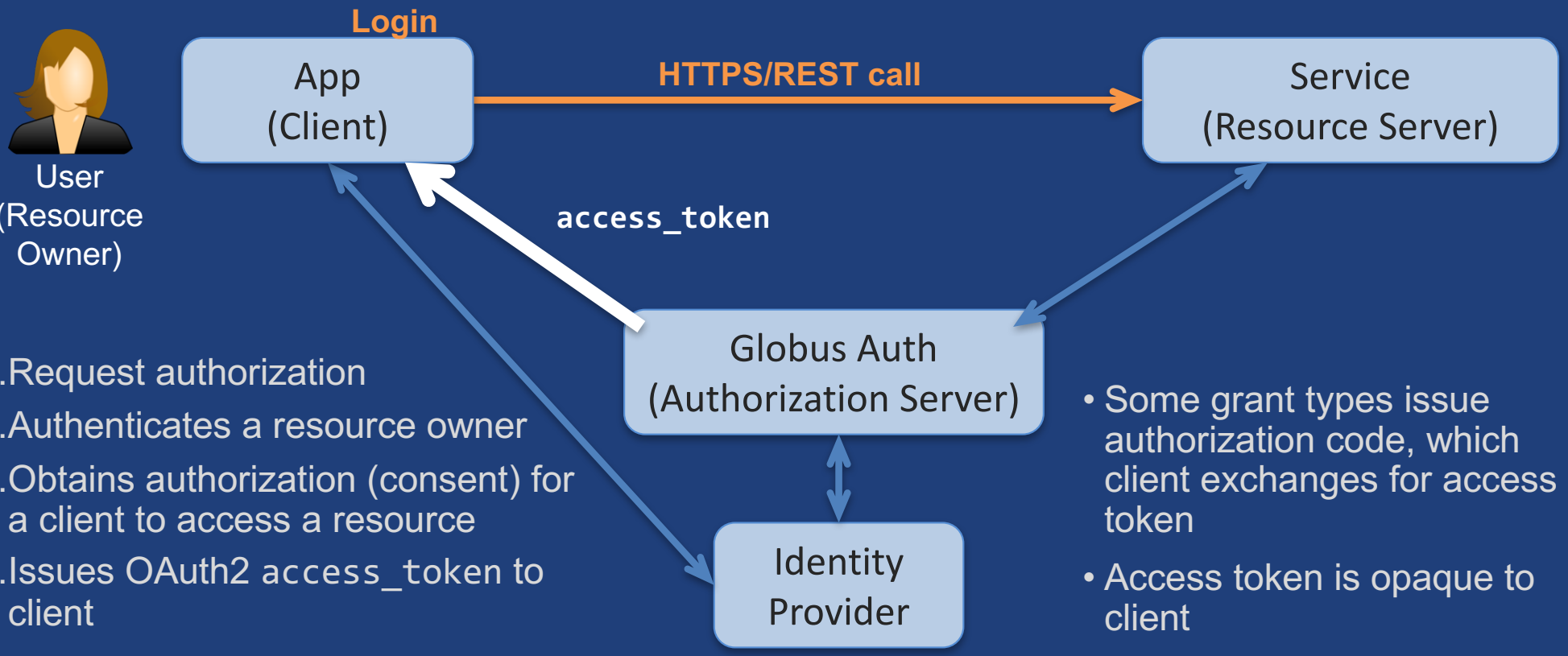
Globus Auth interactions



1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource

- Resource is provided by a resource server
- Limited by a scope

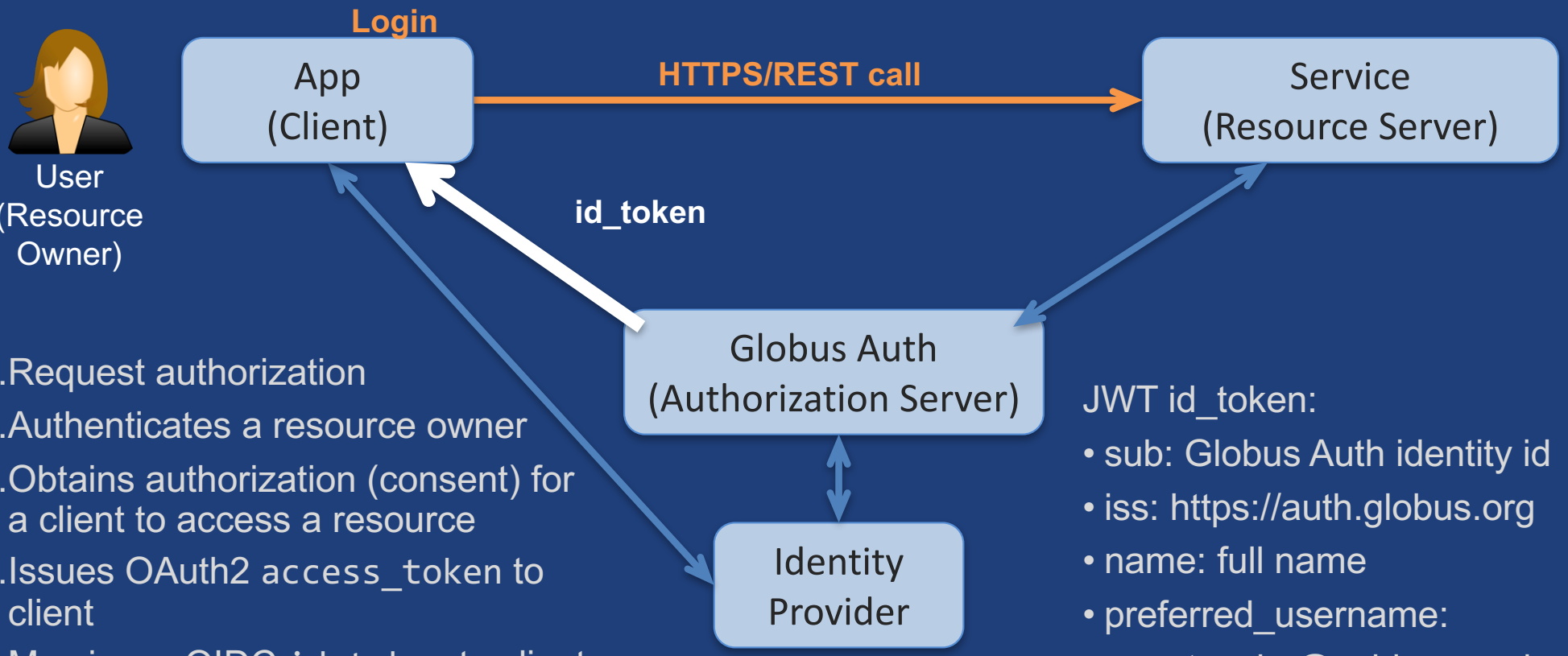
Globus Auth interactions



1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 `access_token` to client

- Some grant types issue authorization code, which client exchanges for access token
- Access token is opaque to client
- May include a refresh token, for offline access

Globus Auth interactions

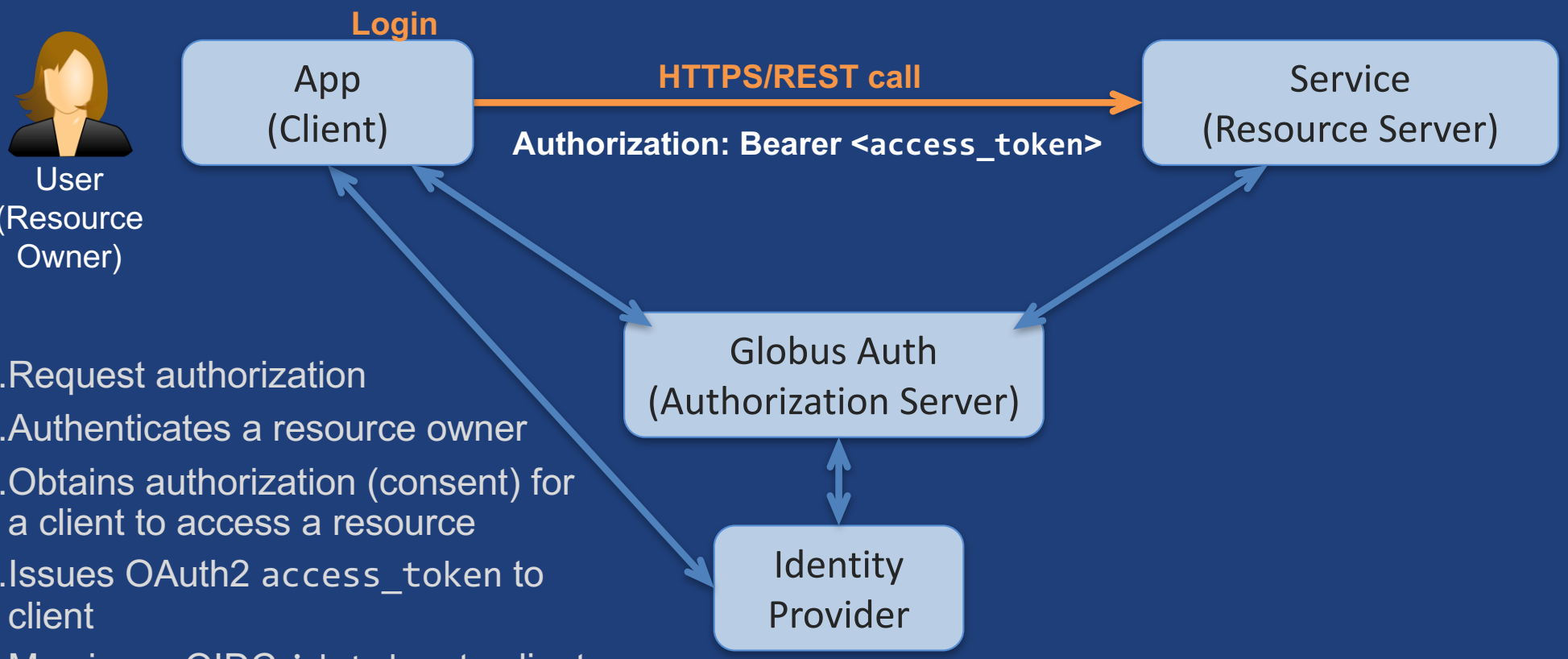


1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 `access_token` to client
5. May issue OIDC `id_token` to client with resource owner identity

JWT id_token:

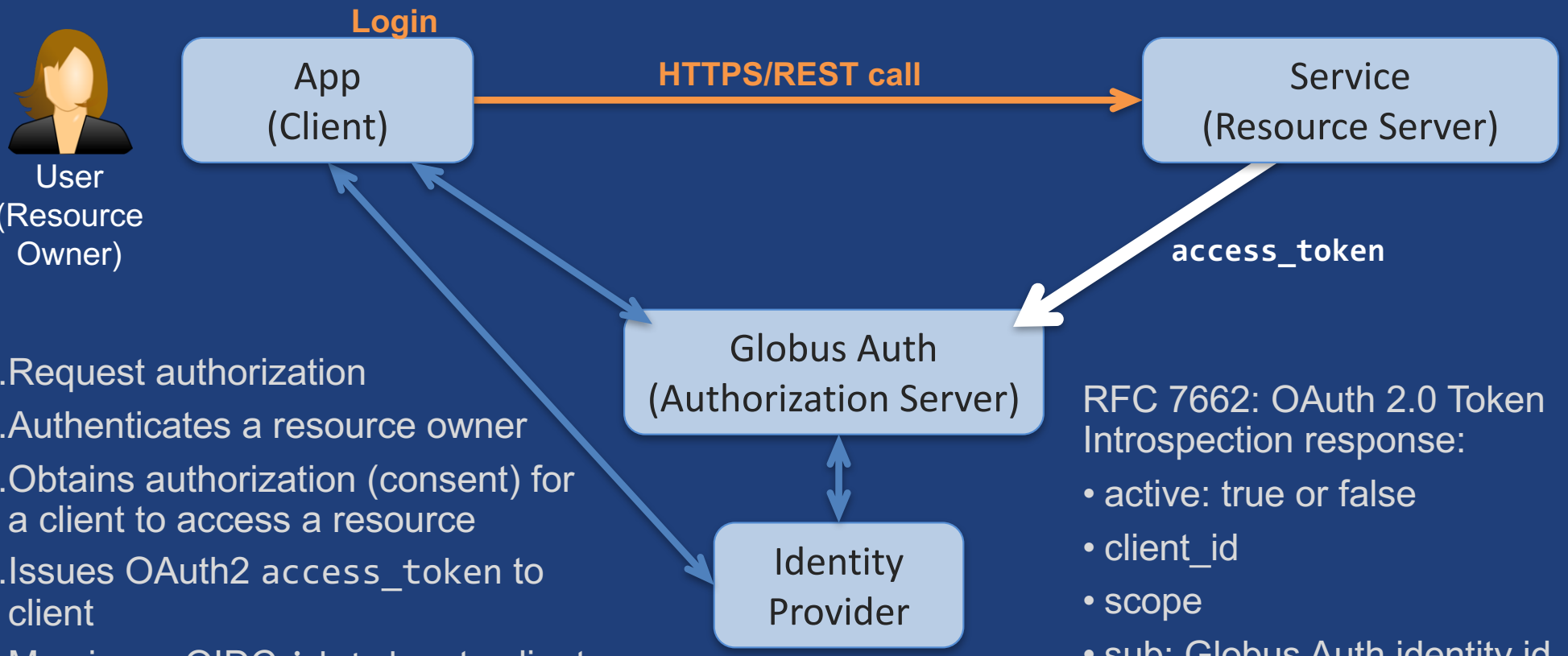
- sub: Globus Auth identity id
- iss: https://auth.globus.org
- name: full name
- preferred_username:
- e.g., tuecke@uchicago.edu
- email: email contact
- other standard OIDC claims

Globus Auth interactions



1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 `access_token` to client
5. May issue OIDC `id_token` to client with resource owner identity
6. HTTPS/REST call with `access_token`

Globus Auth interactions

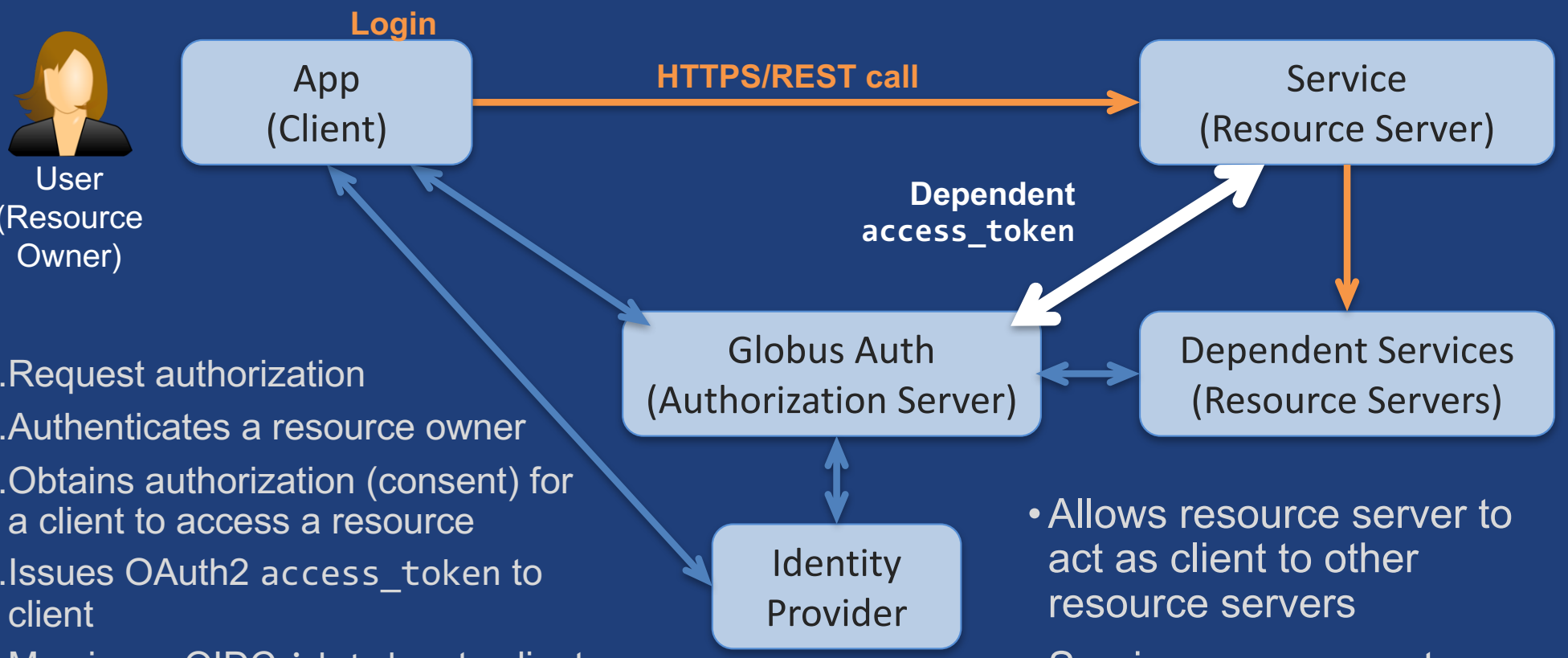


1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 `access_token` to client
5. May issue OIDC `id_token` to client with resource owner identity
6. HTTPS/REST call with `access_token`
7. Validates `access_token` for resource server, gets additional info

RFC 7662: OAuth 2.0 Token Introspection response:

- active: true or false
- client_id
- scope
- sub: Globus Auth identity id
- username: user@myu.edu
- identity_set: linked identities
- email
- name
- other standard claims

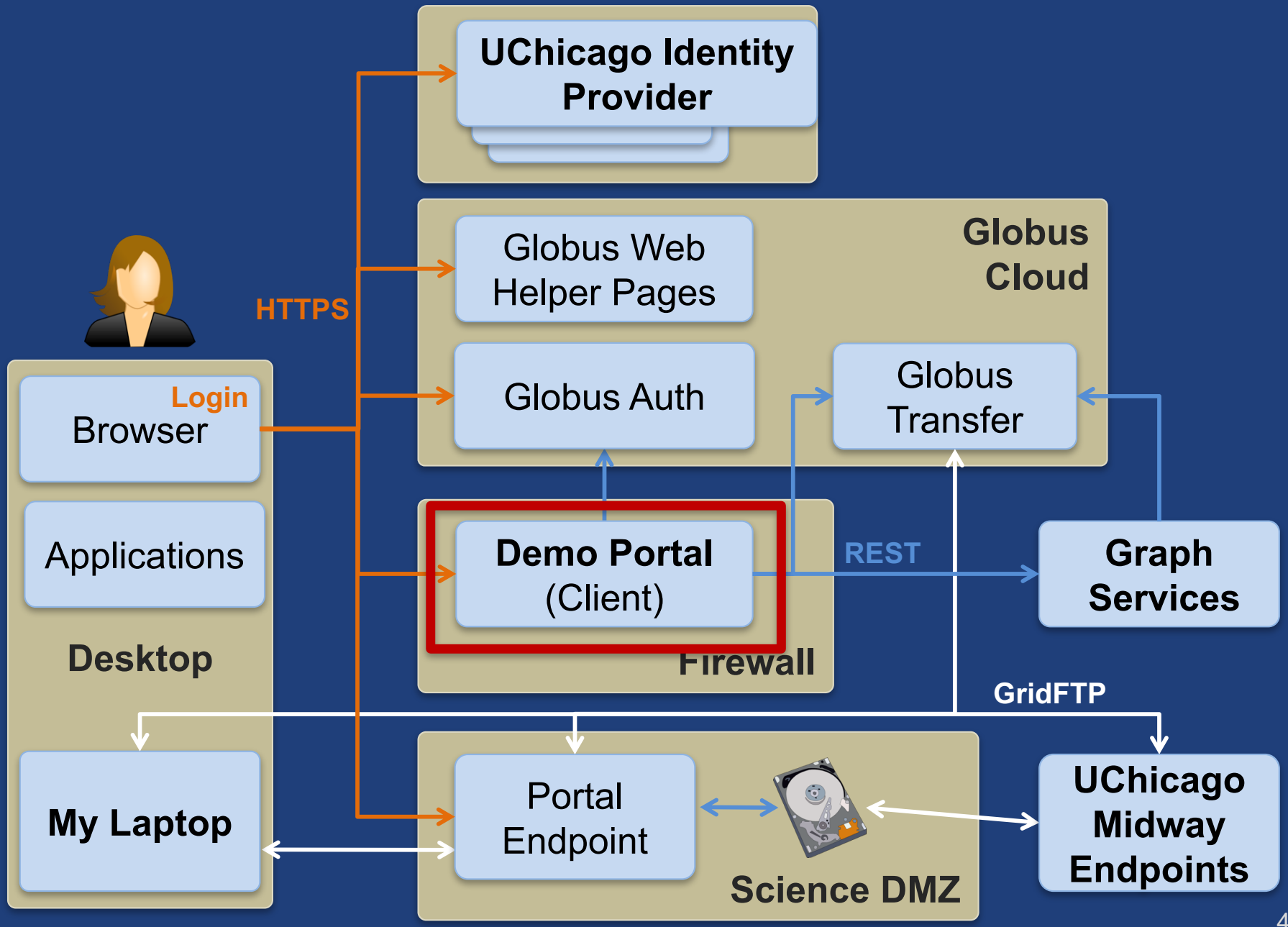
Globus Auth interactions



1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 `access_token` to client
5. May issue OIDC `id_token` to client with resource owner identity
6. HTTPS/REST call with `access_token`
7. Validates `access_token` for resource server, gets additional info
8. Issues dependent access tokens to resource server

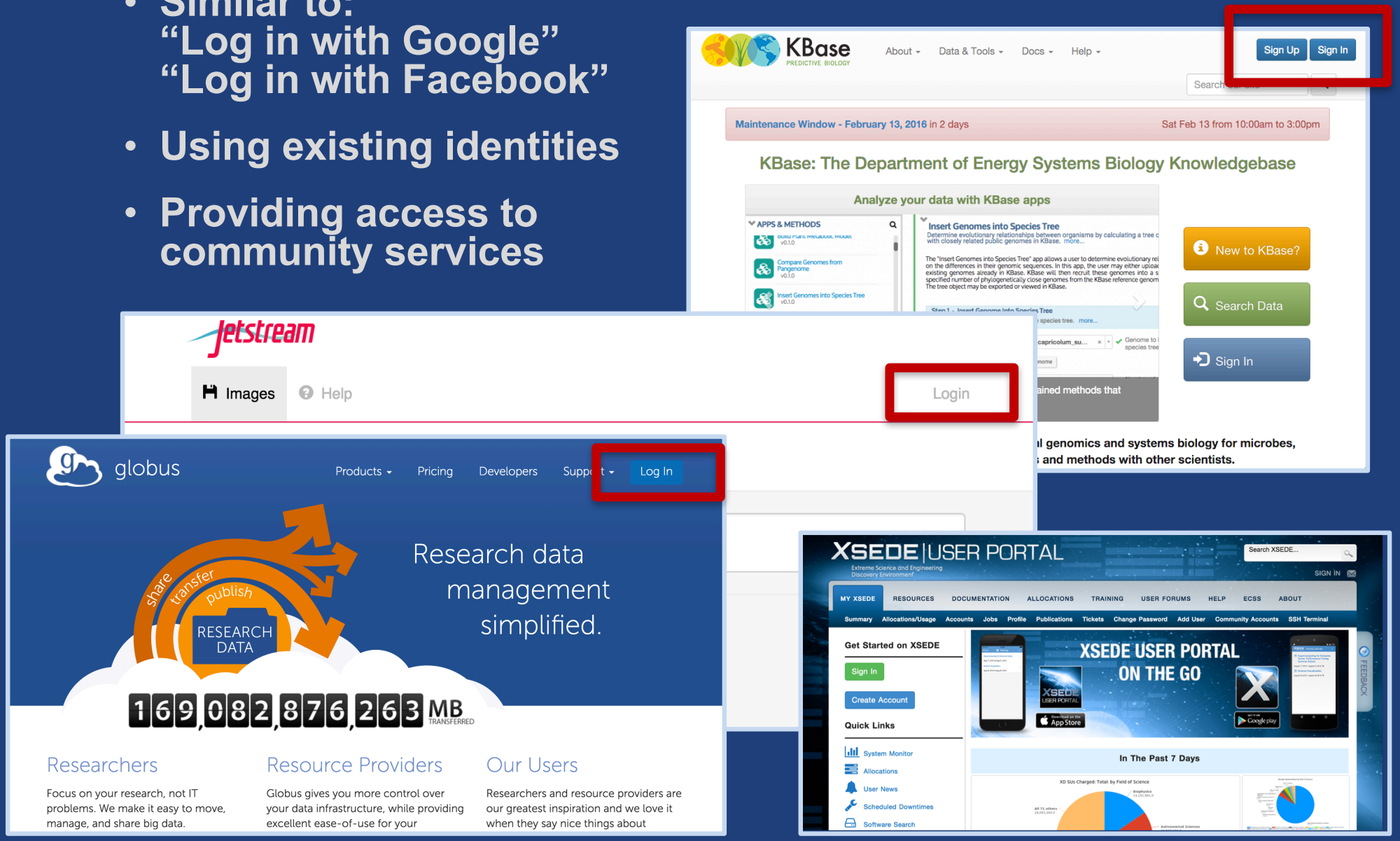
- Allows resource server to act as client to other resource servers
- Service uses request `access_token` to get a *dependent* `access_token` for each dependent service
- Service acts as client to its dependent services

Sample Research Data Portal



Use case: Log in with Globus

- **Similar to:**
  "Log in with Google"
  "Log in with Facebook"
- **Using existing identities**
- **Providing access to community services**



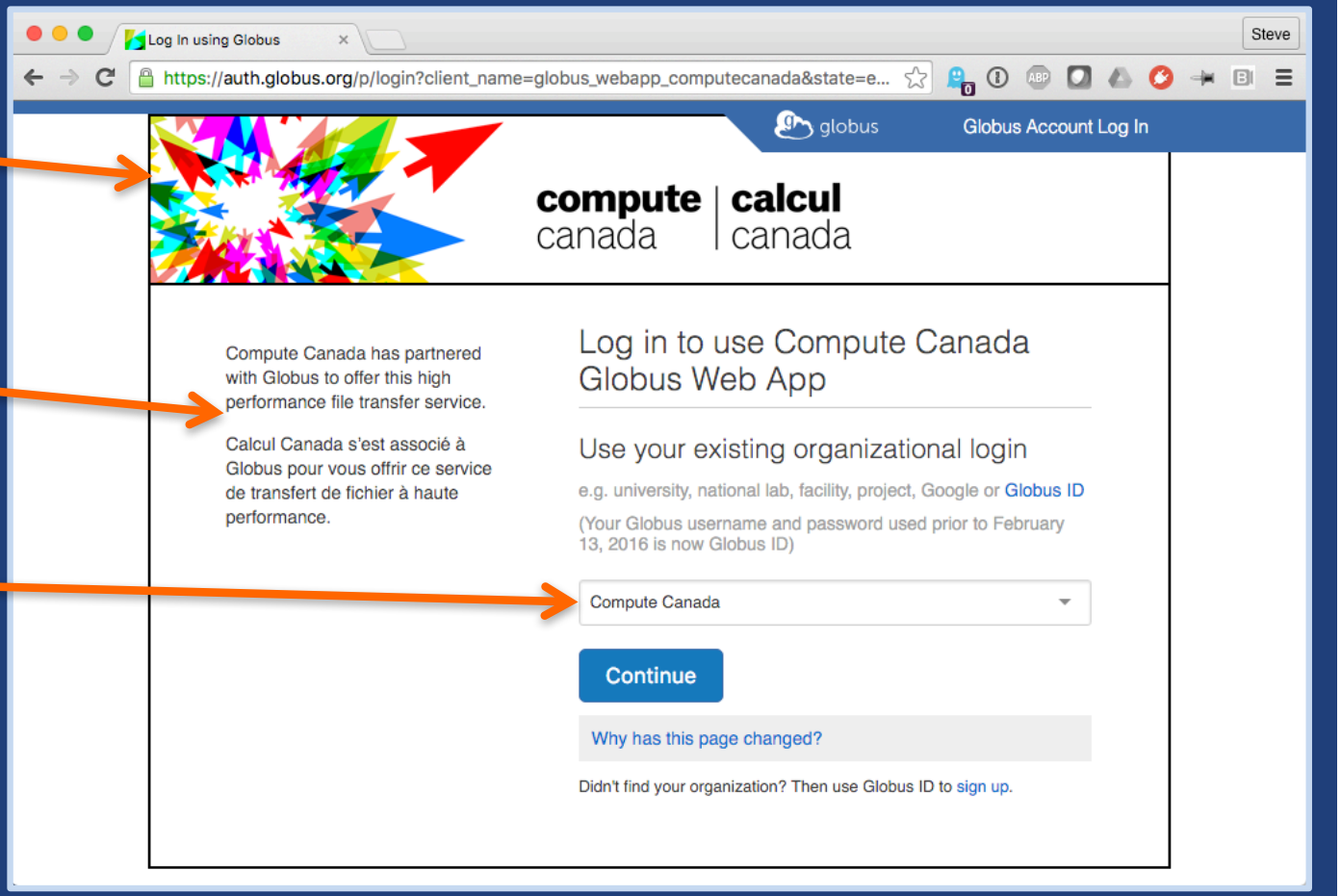
Branding

- **Can skin Globus Auth pages**

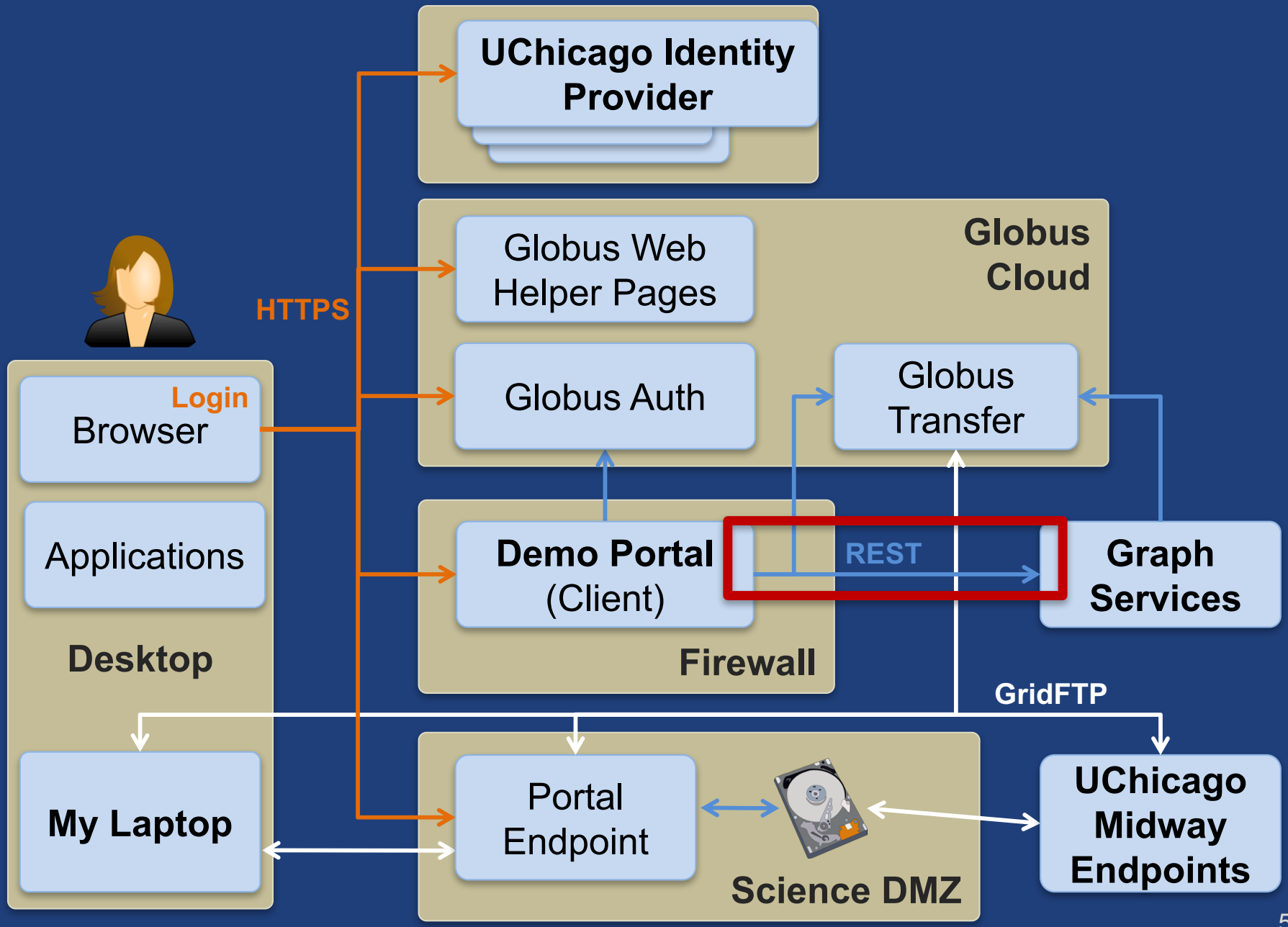
Header

Text

Default IdP



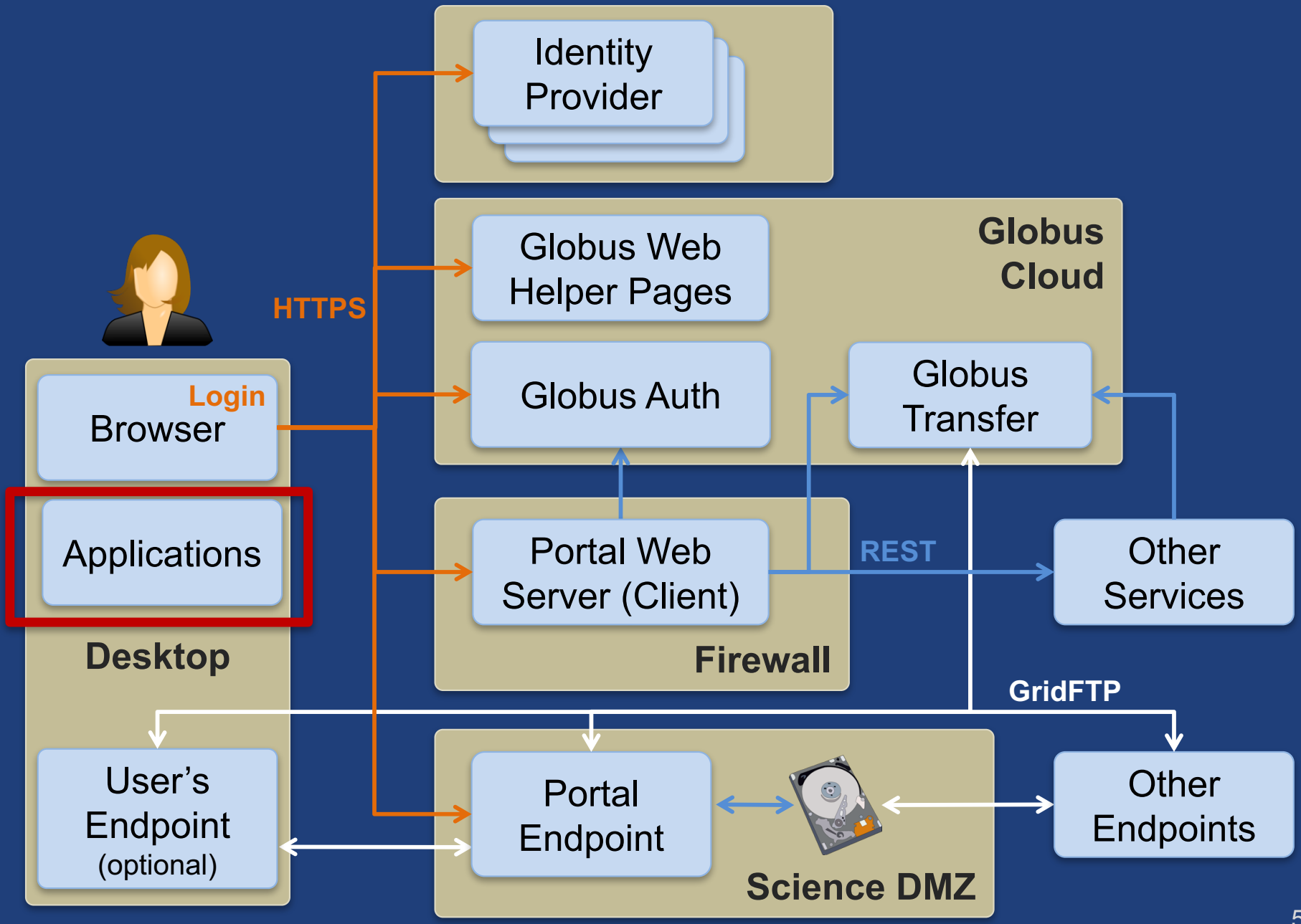
Sample Research Data Portal



Use case: Portal calling services on user's behalf

- **Examples:**
  - Portal starting transfer for user
- **Authorization Code Grant**
  - With service scopes
  - Can also request OIDC scopes
- **Confidential client**
- **Globus SDK:**
  - To get tokens: ConfidentialAppAuthClient
  - To use tokens: AccessTokenAuthorizer

Prototypical research data portal



Use case: Native apps

- **Examples**
  - Command line, desktop apps
  - Mobile apps
  - Jupyter notebooks
  - Any client that cannot keep a secret (downloaded)
- **Native app is registered with Globus Auth**
  - Not a confidential client
- **Native App Grant is used**
  - Variation on the Authorization Code Grant
- **Globus SDK:**
  - To get tokens: NativeAppAuthClient
  - To use tokens: AccessTokenAuthorizer

Mobile apps

- **Globus Auth supports mobile apps**
  - "Log in with Globus" in mobile apps
    - RFC 7636: Proof Key for Code Exchange by OAuth Public Clients (PKCE, pronounced "pixy")
    - Extension to OAuth2 to allow OAuth2 Authorization Code Grant to work from mobile apps
  - Uses mobile browser for web-based login
  - Mobile apps can call any service REST APIs that use Globus Auth
  - iOS and Android
  - Same approach as used by Google, Facebook, etc.

Desktop & command line apps

- **Globus Auth "Native App" PKCE support**
- **Use browser if possible**
  - "OAuth 2.0 for Native Apps"
    - draft-ietf-oauth-native-apps-02
    - Use external browser if possible
    - Embed browser in app
    - Embed mini web server in app
- **Allows copy-n-paste of authorization code**
  - A little like app passwords, but OAuth2 compliant
- **Globus Python SDK and CLI support Native App login**

- **Limited support for username/password authentication**
  - Not recommended

Use case: Apps that need access tokens for long time

- **Examples:**
  - Portal checks for transfer status when user is not logged in
  - Run command line app from script
- **App requests refresh tokens**
- **Globus SDK:**
  - To get token: ConfidentialAppClient or NativeAppClient
  - To use tokens: RefreshTokenAuthorizer

Refresh tokens

- **For “offline services”**
  - E.g., Globus transfer service working on your behalf even when you are offline
- **Refresh tokens issued to a particular client for use with a particular scope**
- **Client uses refresh token to get access token**
  - Confidential client: client_id and client_secret required
  - Native app: client_secret not required
- **Refresh token good for 6 months after last use**
- **Consent rescindment revokes resource token**

Native App Examples

https://github.com/globus/native-app-examples

- **README for install instructions**
- **./example_copy_paste.py**
  - Copy paste code to the app
- **./example_local_server.py**
  - Local server to get the code
- **./example_copy_paste_refresh_token.py**
  - Stores refresh token locally, uses it to get new access tokens

New Globus CLI (beta)

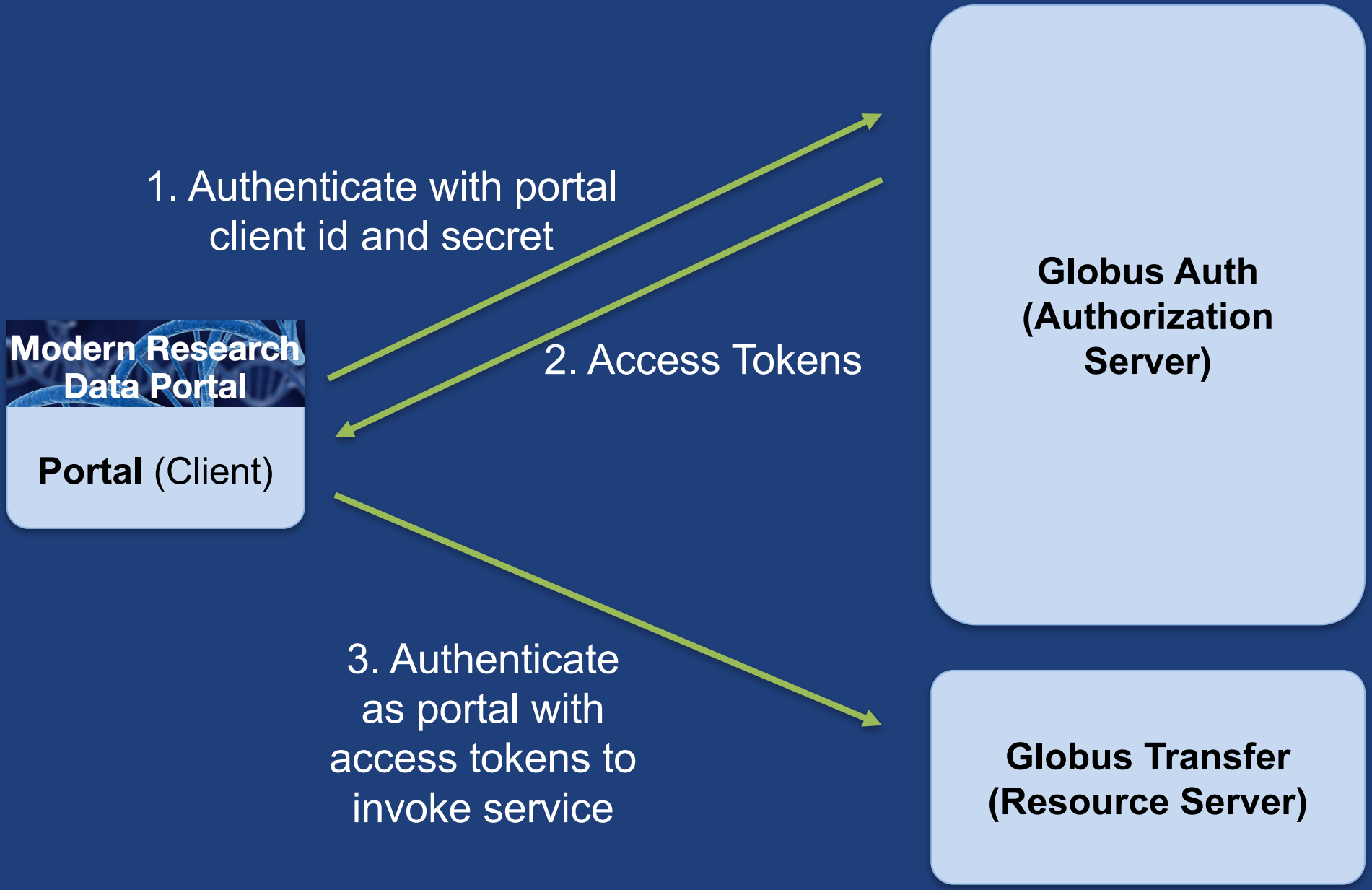
```
$ globus transfer
Usage: globus transfer [OPTIONS] COMMAND [ARGS]...

  Interact with Globus Transfer API. Transfer, Delete, List, and Rename
  files on Endpoints, manage your Endpoints and Shares, and monitor your
  ongoing Transfer Tasks

Options:
  -F, --format [json|text]  Output format for stdout. Defaults to text
  -h, --help                Show this message and exit.
  --version                 Show the version and exit.

Commands:
  acl             Manage Endpoint Access Control Lists
  async-delete    Submit a Delete Task
  async-transfer  Submit a Transfer Task
  bookmark        Manage Endpoint Bookmarks
  endpoint        Manage Globus Endpoint definitions
  ls              List Endpoint directory contents
  mkdir           Make a directory on an Endpoint
  rename          Rename a file or directory on an Endpoint
  task            Manage asynchronous Tasks
```

https://github.com/globus/globus-cli

Use case: App invoking services as itself

- **Examples**
  - Sample portal invoking graph service and accessing endpoints as itself
  - Robots, agents, services
- **App registers with Globus to get client id and secret**
  - Native app cannot do this, because no client_secret
- **Client Credential Grant is used**
- **Globus SDK:**
  - To get tokens: ConfidentialAppAuthClient
  - To use tokens: AccessTokenAuthorizer

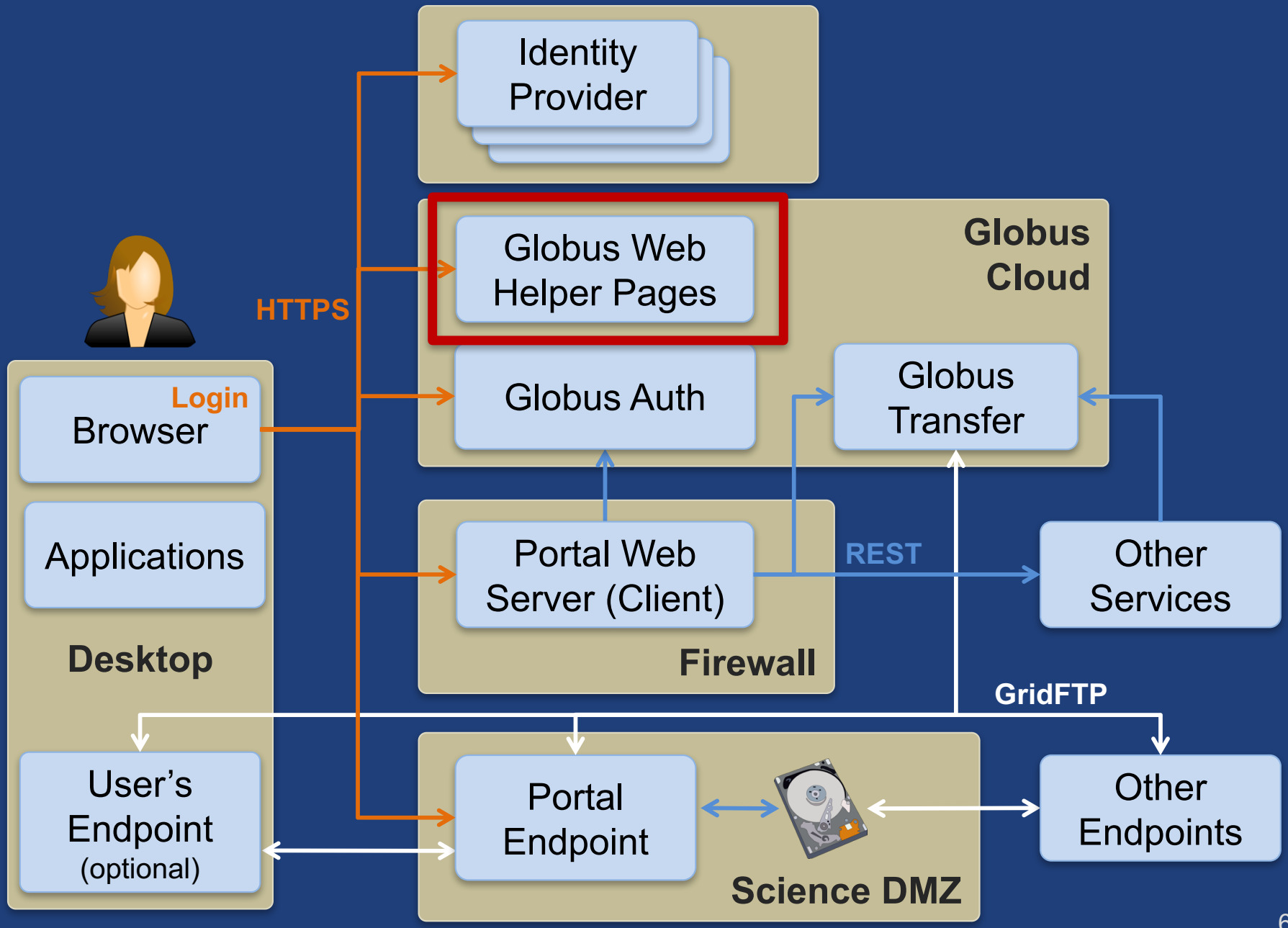
Client credential grant



User identity vs. portal identity

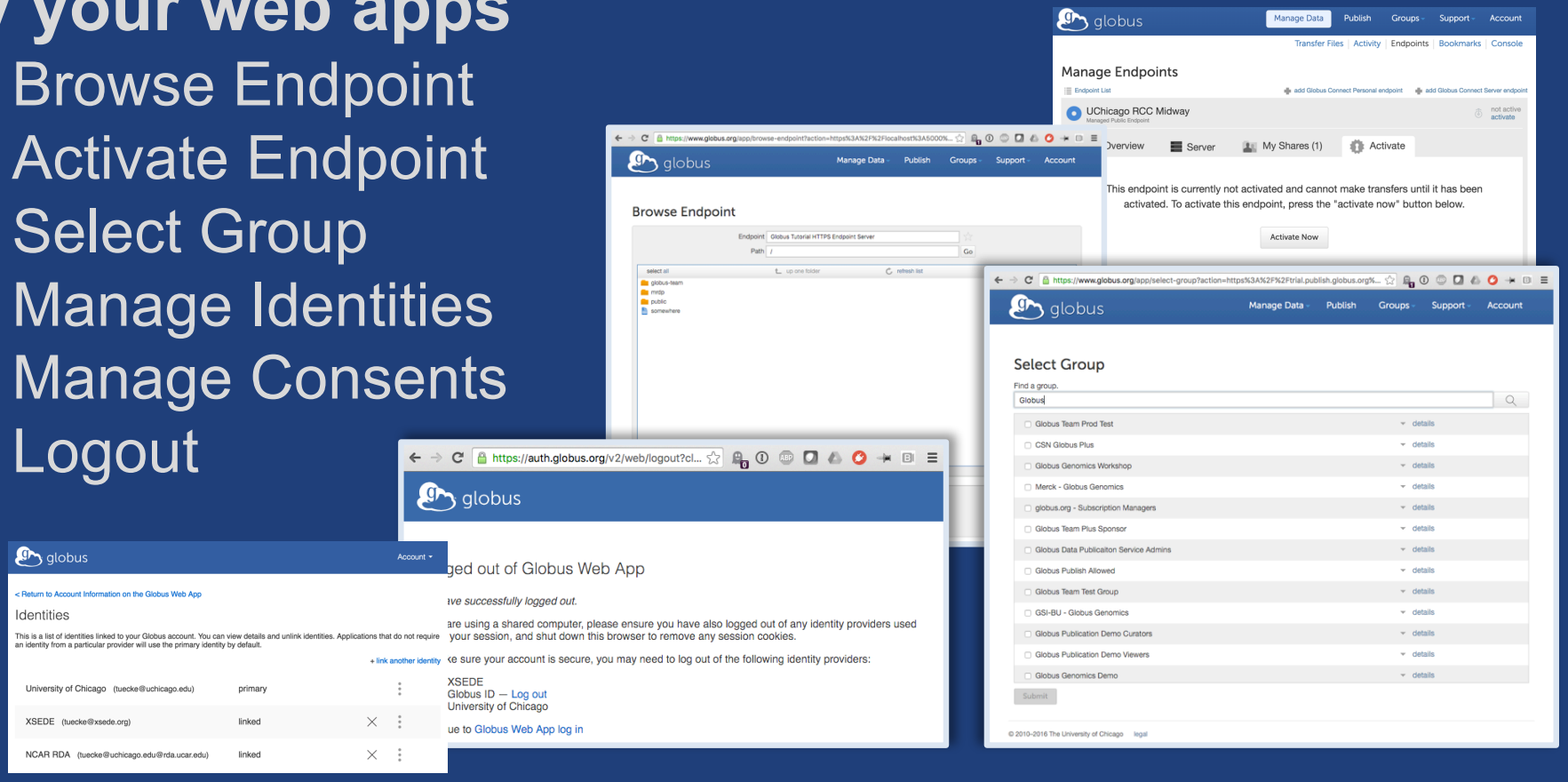
- **User logging into portal results in portal having user's identity and access token**
  - Used to make requests on the user's behalf
- **Portal may also need its own identity**
  - Access and refresh tokens for this identity
  - Used to make requests on its own behalf

Prototypical research data portal



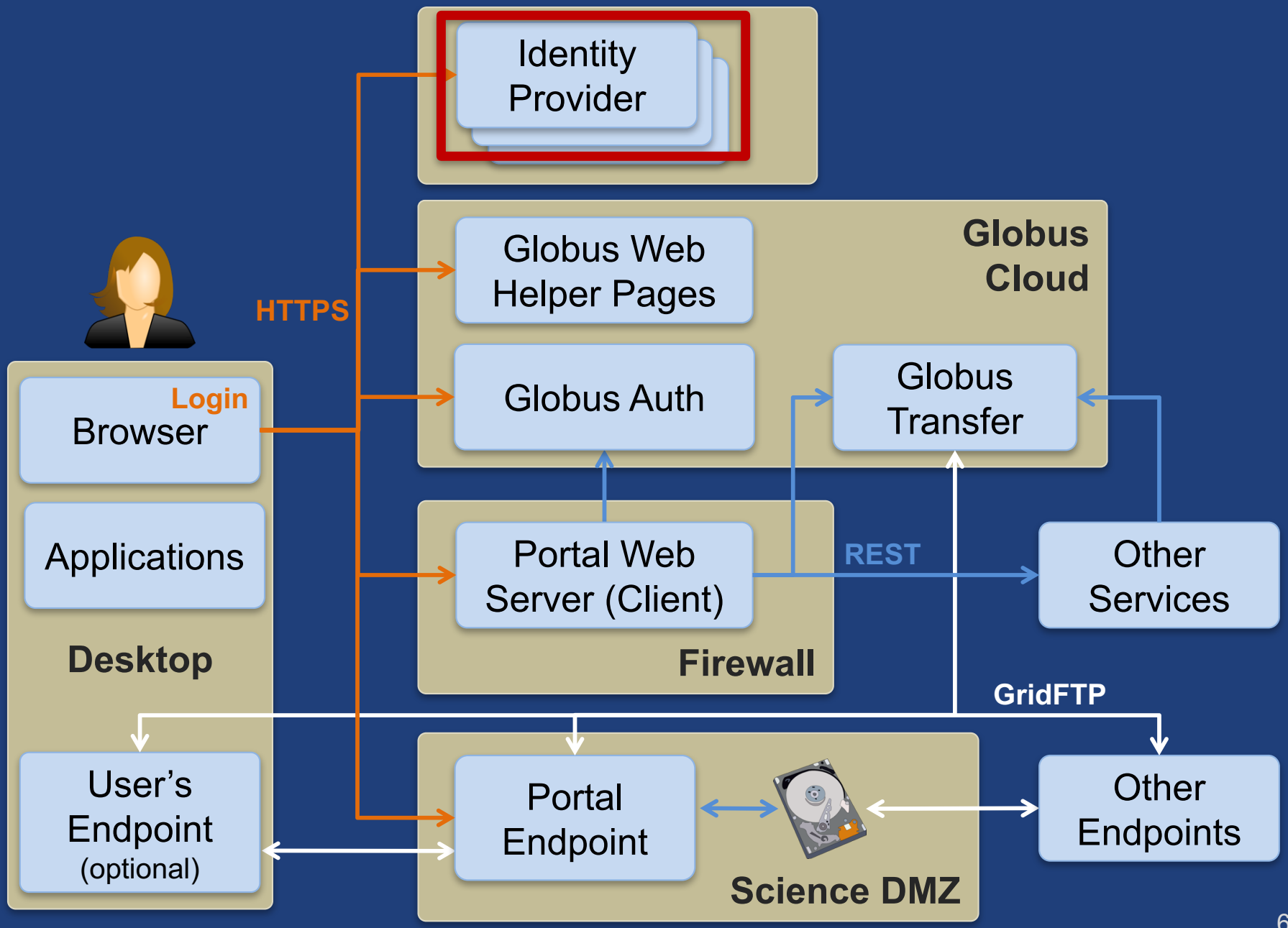
Globus Helper Pages

- **Globus provided web pages designed for use by your web apps**
  - Browse Endpoint
  - Activate Endpoint
  - Select Group
  - Manage Identities
  - Manage Consents
  - Logout



docs.globus.org/api/helper-pages

Prototypical research data portal



Adding your identity provider

- **InCommon identity providers that release research & scholarship attributes to CILogon** ***(free)***

- **Any other OpenID Connect identity provider** ***(subscription)***

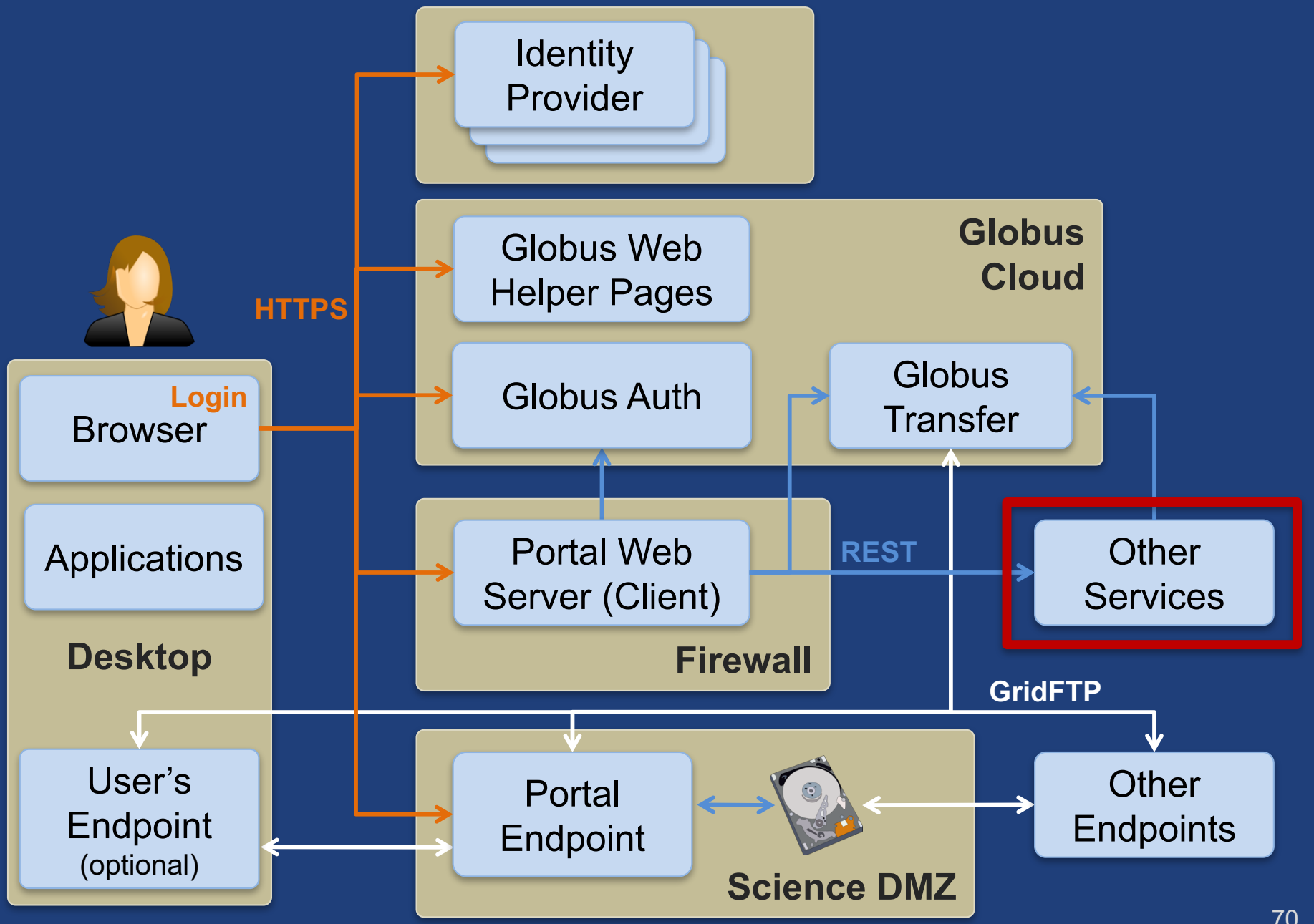
Adding an identity provider

- **If your portal has identities already:**
  - Deploy OIDC server in front of it
    - Globus Python OIDC (coming soon)
    - Any standard OIDC server should work
    - Requires claim that can map to username
    - Optional claims: name, email, organization
  - Can register apps and services with an effective identity policy
    - Requires account to have identity from your identity provider when logging into your app

Portal accounts

- **Your app portal can still have portal accounts for users**
- **Tie portal account to Globus account identity, rather than username/password**
- **Associate your profile with this account**
- **Globus Auth handles authentication of that identity, in order to log user into your portal account**

Prototypical research data portal



Why create your own services?

- **Front-end / back-end within your portal**
  - Remote backend for portal
  - Backend for pure Javascript browser apps

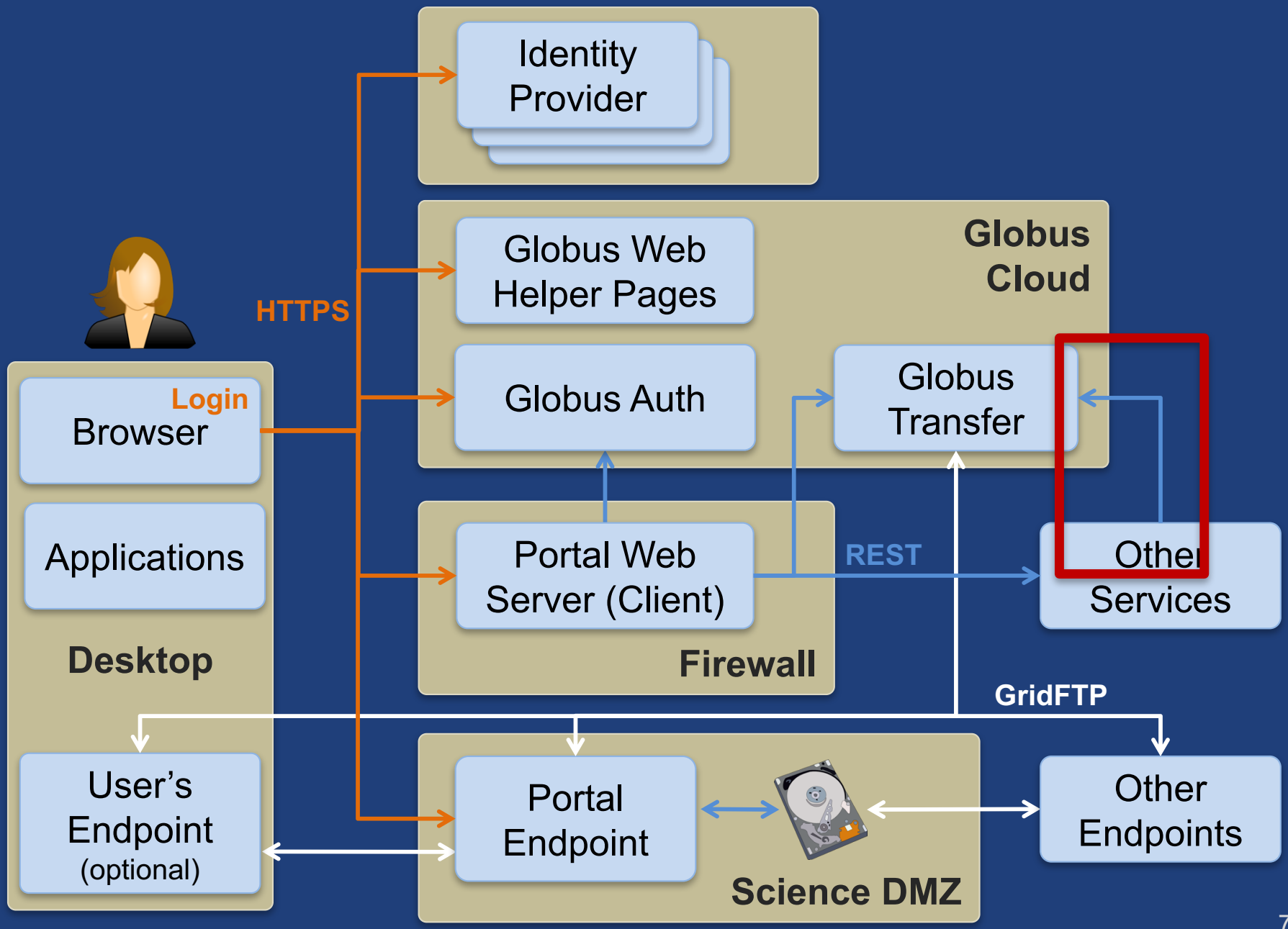
- **Extend your portal with a public REST API, so that other app and service developers can integrate with and extend your portal**

Why Globus Auth for your service?

- **Outsource all identity management and authentication**
  - Federated identity with InCommon, Google, etc.
- **Outsource your REST API security**
  - Consent, token issuance, validation, revocation
  - You provide service-specific authorization
- **Apps use your service like all others**
  - Its standard OAuth2 and OIDC
- **Your service can seamlessly leverage other services**
- **Other services can leverage your service**
- **Implement your service using any language and framework**

Add your service to the science cyberinfrastructure platform

Prototypical research data portal



Dependent tokens

- **Your service can act as client to other services (scopes)**
  - Globus Transfer and Auth
  - XSEDE (e.g., Jetstream, XUP)
  - Other community services
  - Future: Commercial services (e.g., Google Drive)
- **Entire service call tree consented by user and service owners**
  - Rescinding consent revokes all dependent tokens
- **Dependent tokens are restricted to a particular client, calling a particular scope, on behalf of a particular resource owner (e.g., user)**
  - Restricted delegation!

Join the Globus developer community

- **Join developer-discuss@globus.org mailing lists: globus.org/mailing-lists**
- **Python SDK is open source**
  - **github.com/globus/globus-sdk-python**
  - Submit issues, pull requests
  - Discussions on **developer-discuss@globus.org**
- **Jupyter notebook, sample data portal and native applications are open source on github**
- **Documentation: docs.globus.org**

https://www.globusworld.org/

Thank you for participating!

- Please offer your feedback with our 30-second evaluation: http://sciencegateways.org/webinareval

- Join us next month (April 12) for

  Gateway Showcase featuring Greg Madey of the Ensayo project and another gateway TBD