



Building on the Globus Platform

Rachana Ananthakrishnan
rachana@globus.org

Vas Vasiliadis
vas@uchicago.edu

Penn State University — June 29, 2017





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

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
P R O D U C	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



Example: ARM Climate Research Facility

Data Selection Summary

Signed in as ANANTHAKRISHNANR1.

mergesonde1mace c1 @ fkb M1 [Generate Citation](#) **274 file(s) // 6014 MB**

Order Complete Datastream Extract Specific Measurements

Note: All variables will be delivered for this datastream.

Measurement : Atmospheric temperature
Variable : Temperature // temp

2007-04-01

2007-12-31

Combine files by datastream

File format

Remove data flagged by Data Quality Reports (DQR) of type Incorrect Suspect

Data Delivery Options

FTP

Globus

THREDDS

Dropbox

Extraction options only apply when "Extract Specific Measurements" is selected.

Original files will be delivered as part of all orders.

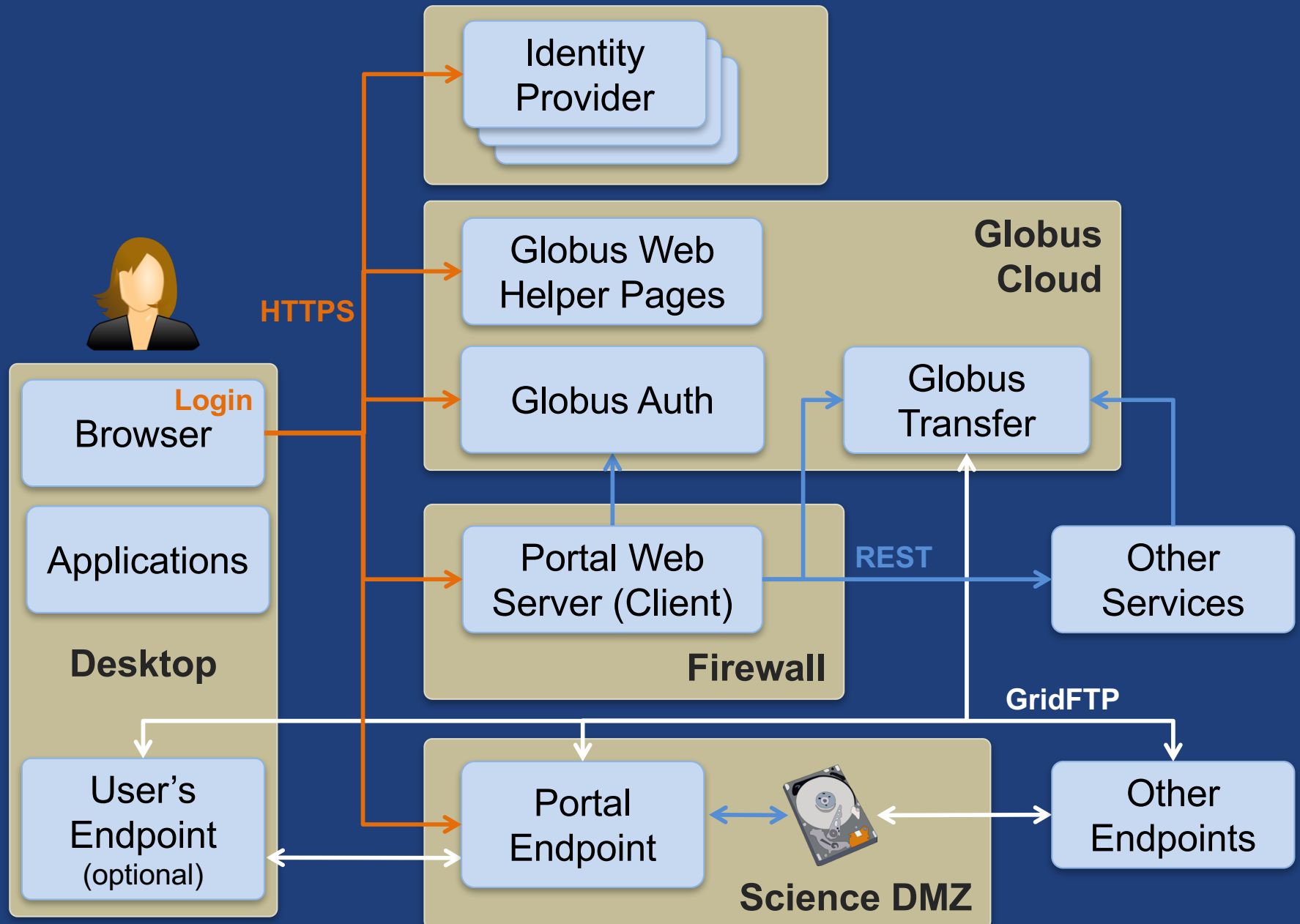


Demo

**Sample
Research Data Portal**

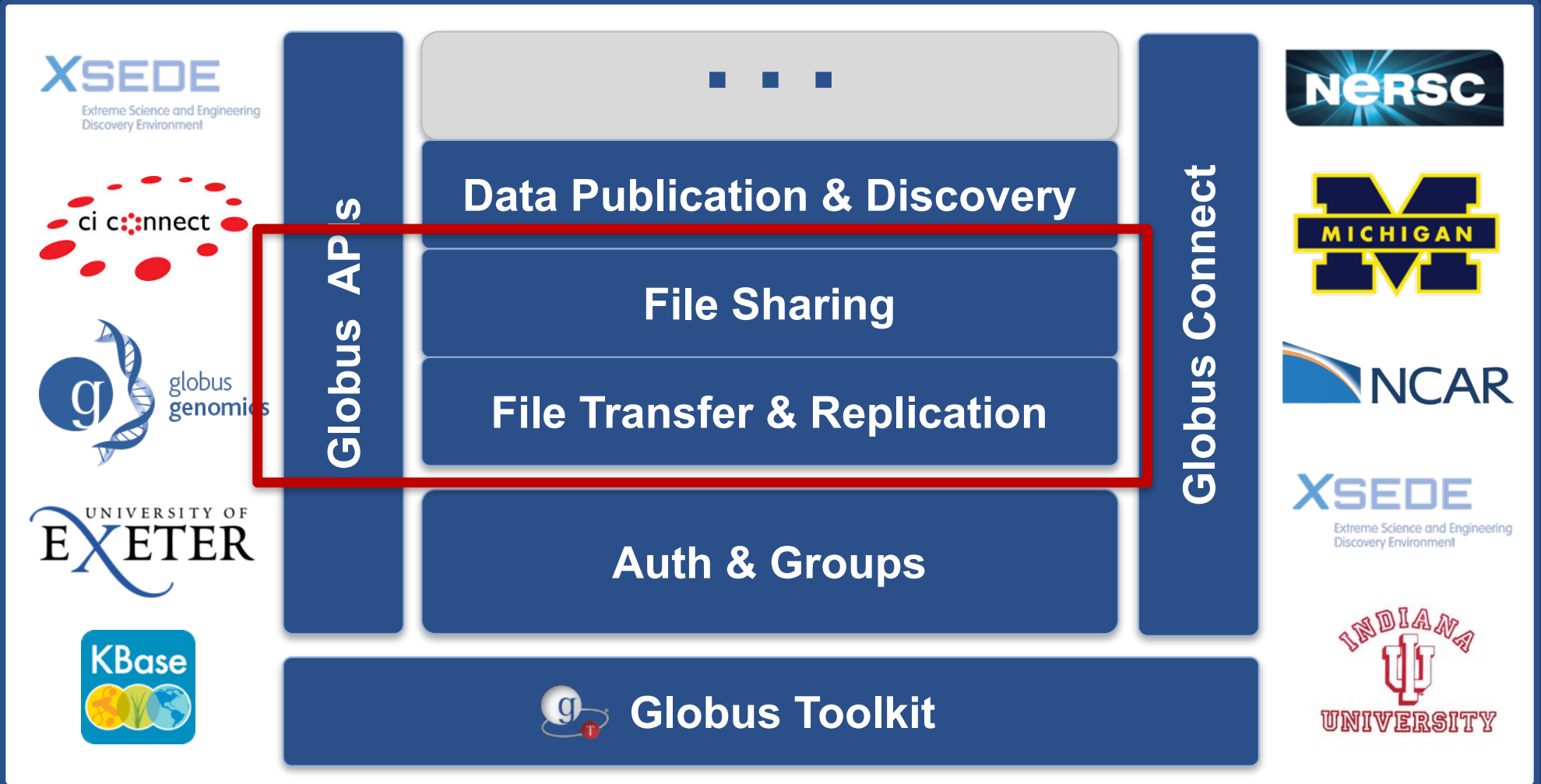


Prototypical research data portal



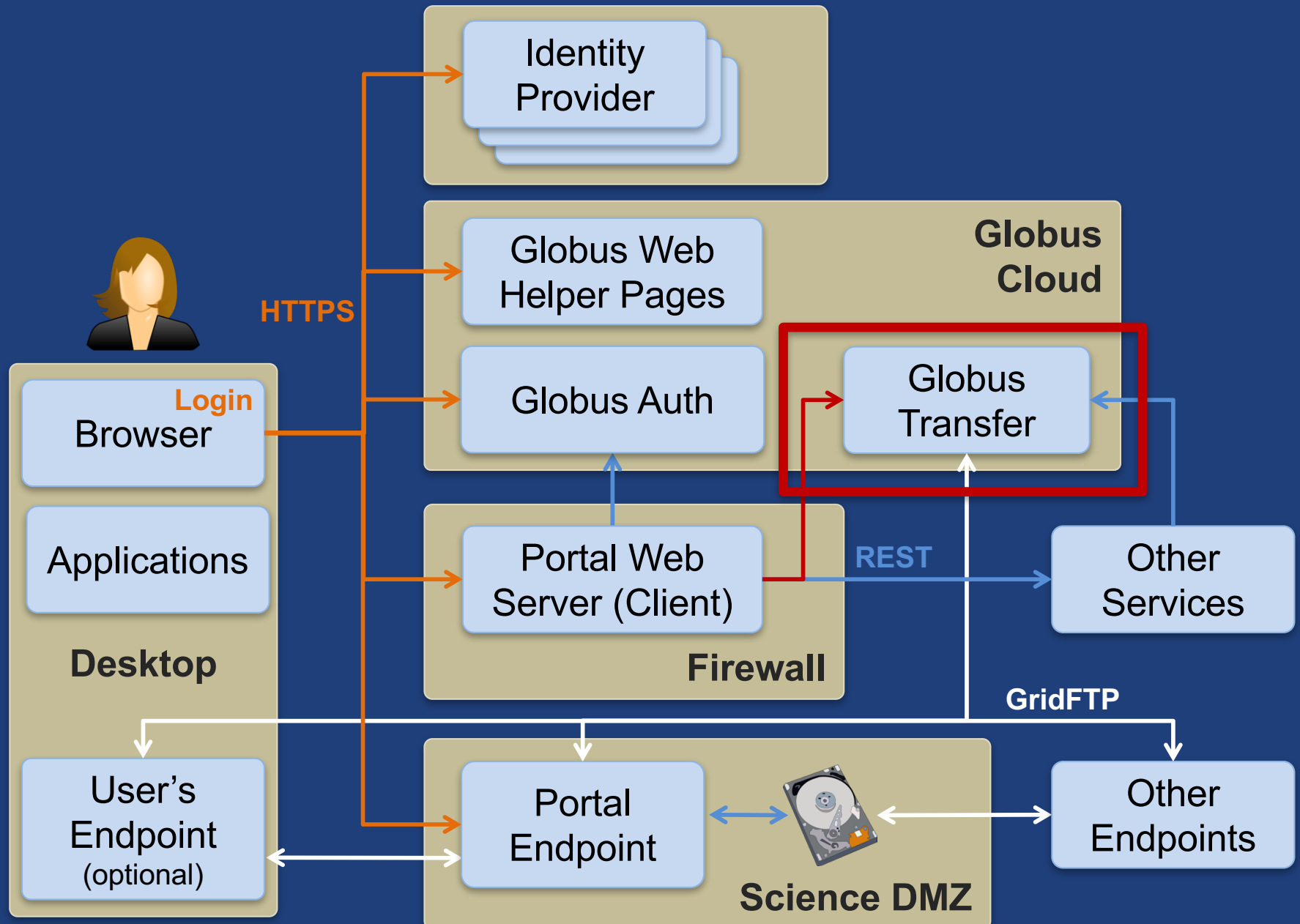


Globus PaaS





Prototypical research data portal





Introduction to REST APIs

- **Remote operations on resources via HTTPS**
 - POST ~ = Create (or other operations)
 - GET ~ = Read
 - PUT ~ = Update
 - DELETE ~ = Delete
- **Globus APIs use JSON for documents and resource representations**
- **Resource named by URL**
 - Query params allow refinement (e.g., subset of fields)
- **Requests authorized via OAuth2 access token**
 - Authorization: Bearer asdflkqhafsdafeawk



Globus Transfer API

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

docs.globus.org/api/transfer

- Fairly stable
- Deprecation policy



Globus Python SDK

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

globus.github.io/globus-sdk-python



TransferClient class

- `globus_sdk.TransferClient` **class**

```
from globus_sdk import TransferClient  
tc = TransferClient()
```

- **Handles connection management, security, framing, marshaling**



TransferClient low-level calls

- **Thin wrapper around REST API**

- `post()`, `get()`, `update()`, `delete()`

`get(path, params=None, headers=None, auth=None, response_class=None)`

- `path` – path for the request, with or without leading slash
 - `params` – dict to be encoded as a query string
 - `headers` – dict of HTTP headers to add to the request
 - `response_class` – class for response object, overrides the client's `default_response_class`
 - Returns: `GlobusHTTPResponse` object



TransferClient higher-level calls

- **One method for each API resource and HTTP verb**
- **Largely direct mapping to REST API**

```
endpoint_search(filter_fulltext=None,  
                filter_scope=None,  
                num_results=25,  
                **params)
```



Python SDK Jupyter notebook

- Jupyter (iPython) notebook demonstrating use of Python SDK

github.com/globus/globus-jupyter-notebooks

- Overview
- Open source, enjoy



Walk-through

Jupyter Notebook



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**



Exercise: Jupyter notebook

Install Jupyter notebook either locally or on EC2 instance

github.com/globus/globus-jupyter-notebooks.git

Modify Jupyter notebook to:

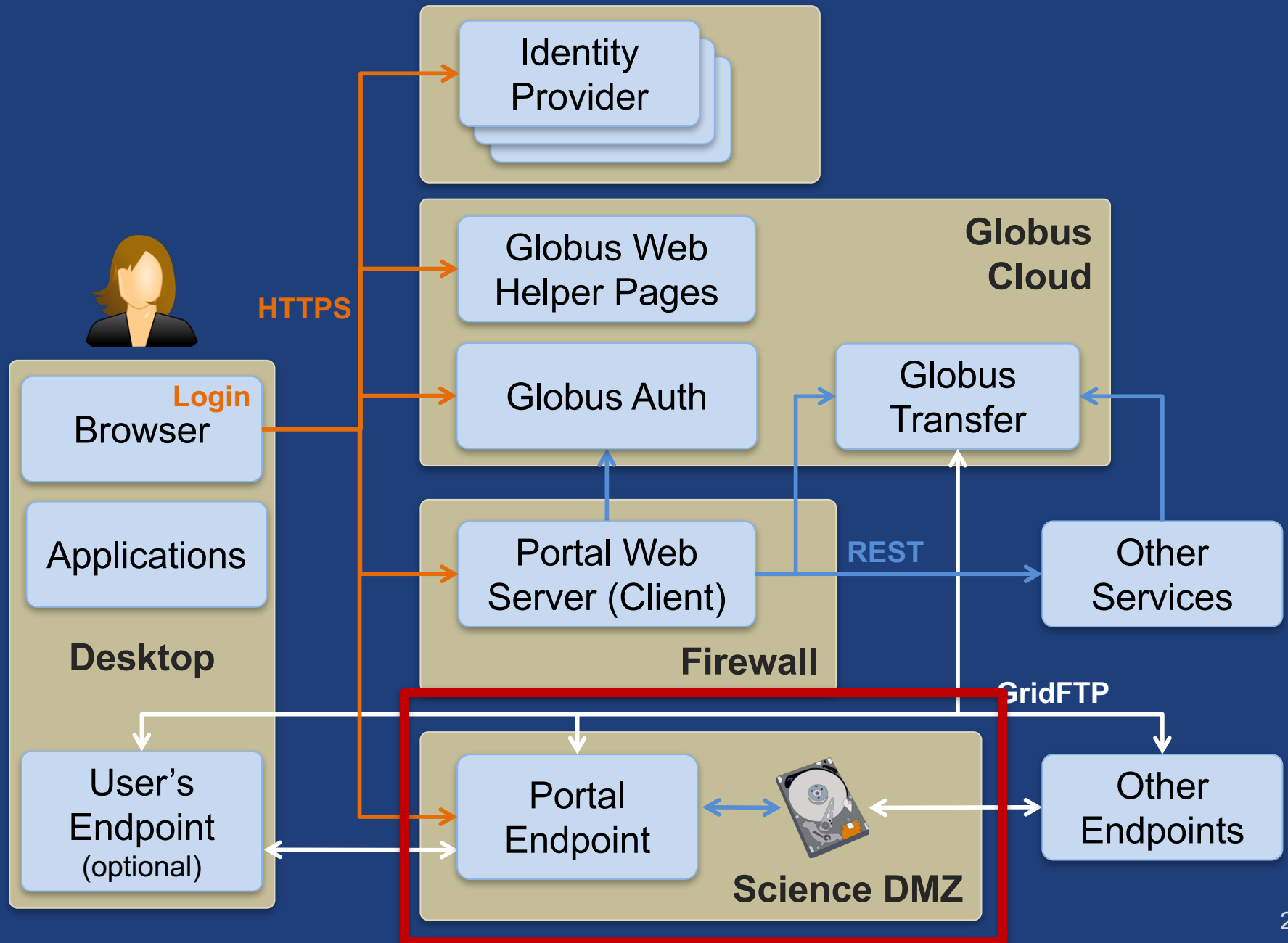
1. Find the endpoint id for XSEDE Comet
2. Set all the metadata fields on your shared endpoint
3. Set permissions to allow your neighbor to access your shared endpoint
4. Transfer all files *.txt from the tourexercise directory on the Globus Vault endpoint to any other endpoint.



Maximizing the value of the Science DMZ

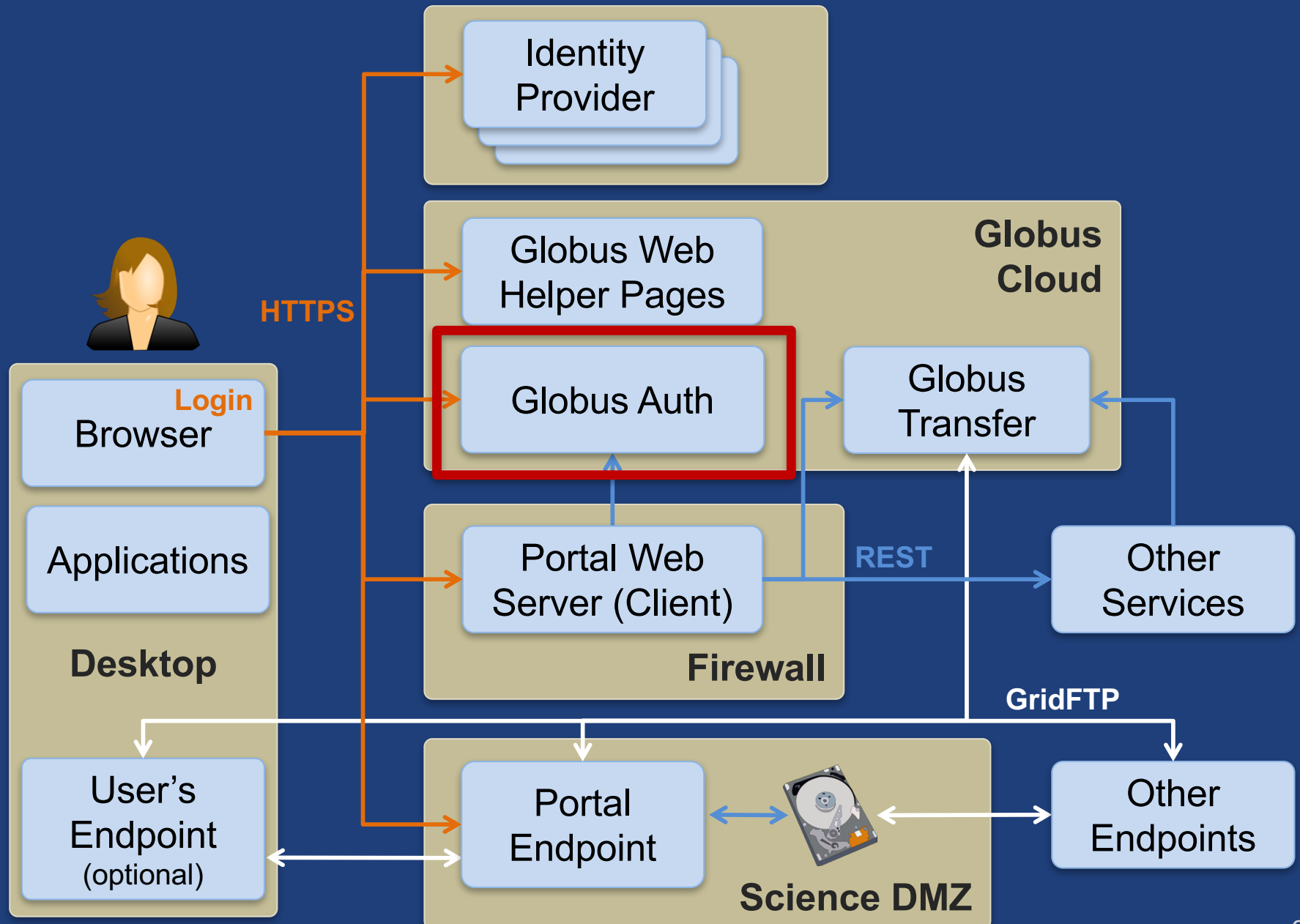


Prototypical research data portal



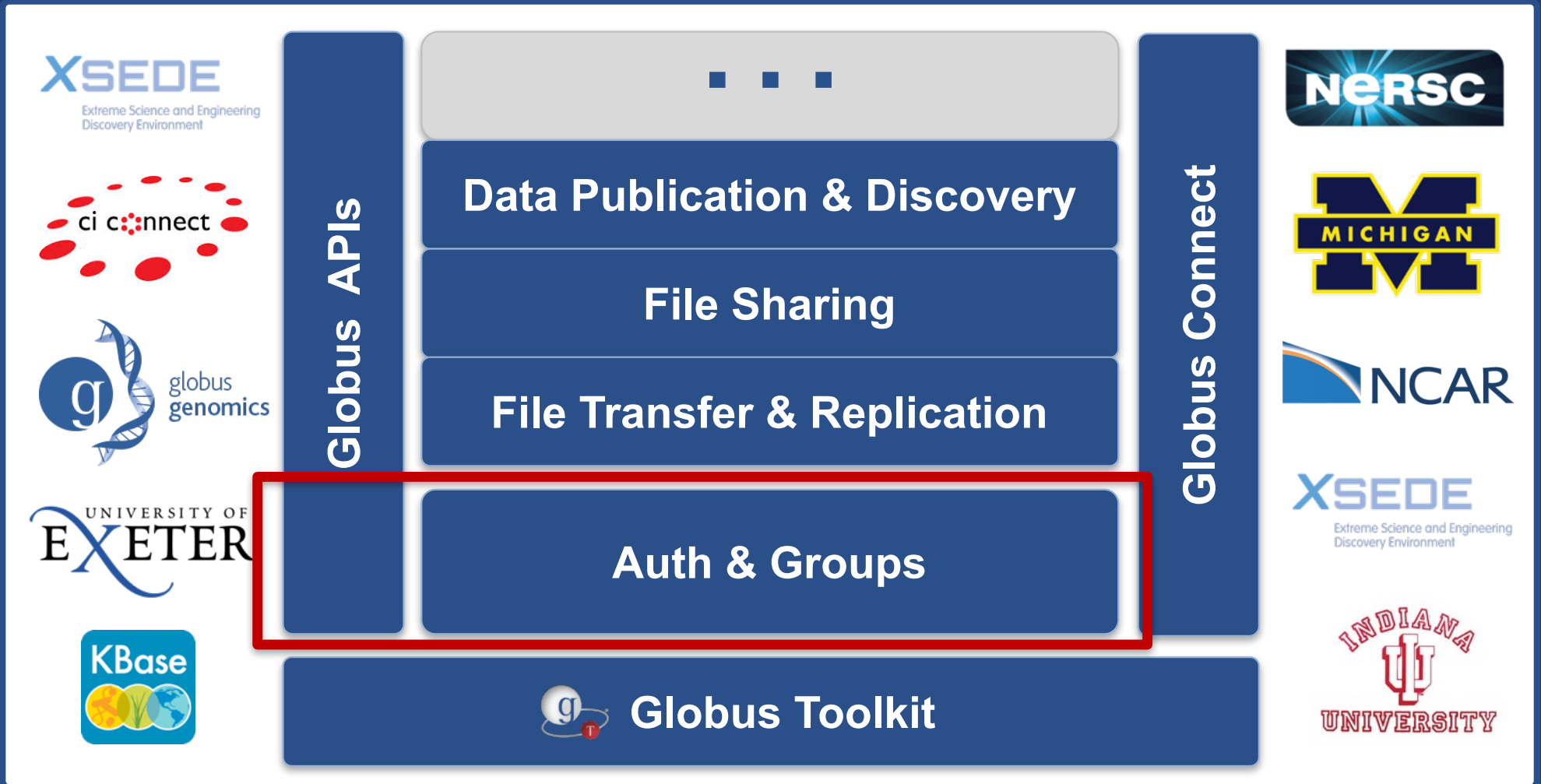


Prototypical research data portal





Globus PaaS





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

- Specification
- Developer Guide
- API 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



Scopes

- **APIs that client is requesting access to**
- **Scope syntax:**
 - OpenID Connect: openid, email, profile
 - urn:globus:auth:scope:<service-name>:<scope-name>
- **If client requests multiple scopes**
 - Token response has tokens for first scope
 - other_tokens field in response has list of token responses for other scopes
 - Client must use correct token with each request



Consent

- **Resource owner authorization that a client can request access to a service scope on the resource owner's behalf within a limited scope**
 - If service has dependent scopes, they are part of the consent
- **User can rescind a consent at any time**
 - Invalidates all access, dependent, and refresh tokens originating from the client



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

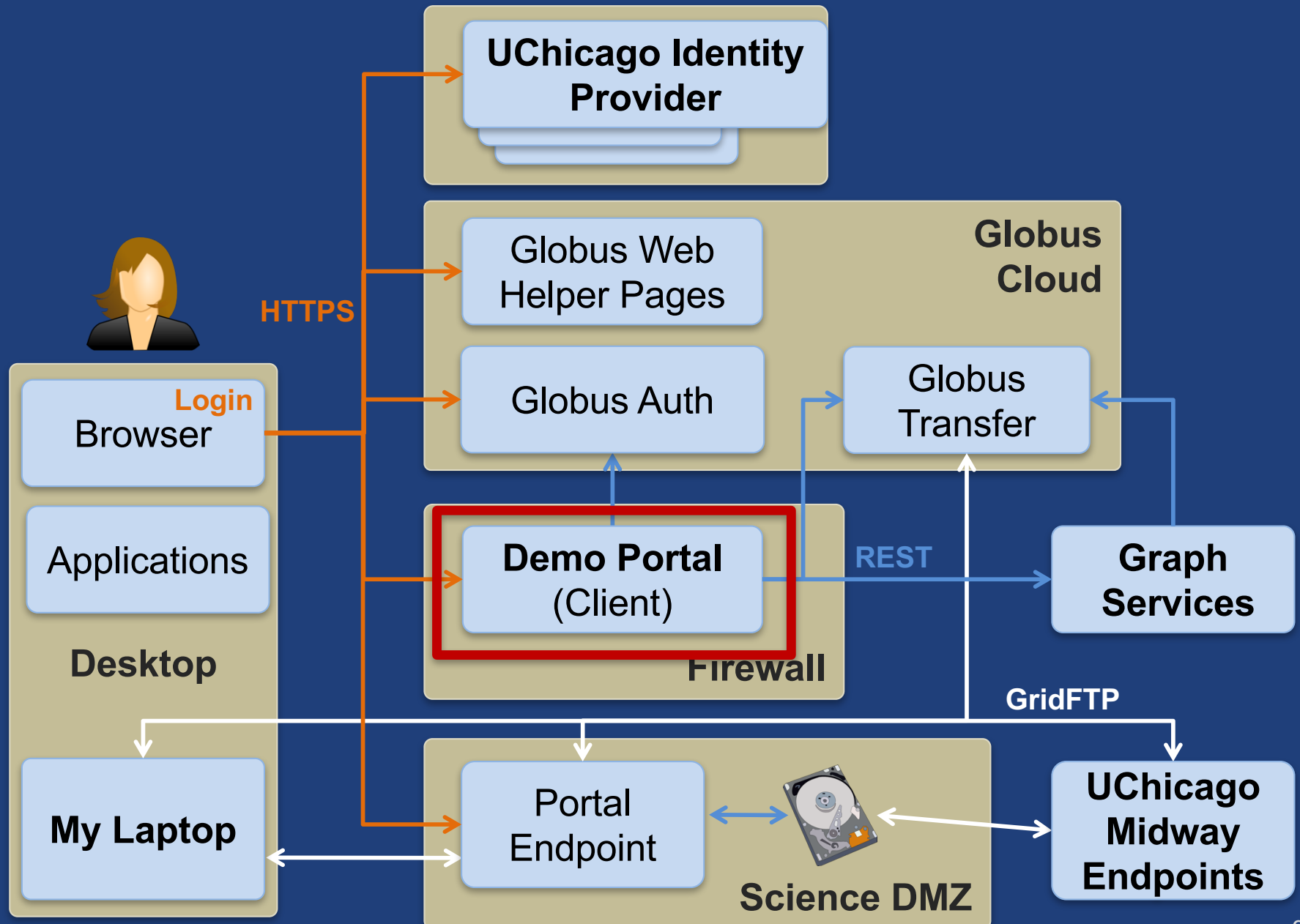


Identity id vs. username

- **Identity id:**
 - Guaranteed unique among all Globus Auth identities, and will never be reused
 - UUID
 - Always use this to refer to an identity
- **Identity username:**
 - Unique at any point in time
 - May change, may be re-used
 - Case-insensitive user@domain
 - Can map to/from id, for user experience
- **Auth API allows mapping back and forth**



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

KBase PREDICTIVE BIOLOGY

About - Data & Tools - Docs - Help

Maintenance Window - February 13, 2016 in 2 days Sat Feb 13 from 10:00am to 3:00pm

KBase: The Department of Energy Systems Biology Knowledgebase

Analyze your data with KBase apps

APPS & METHODS

- DUPLICATE PANGENOME ANALYSIS v0.1.0
- Compare Genomes from Pangenome v0.1.0
- Insert Genomes into Species Tree v0.1.0

Insert Genomes into Species Tree

Determine evolutionary relationships between organisms by calculating a tree c with closely related public genomes in KBase. more...

The "Insert Genomes into Species Tree" app allows a user to determine evolutionary rel on the differences in their genomic sequences. In this app, the user may either upload existing genomes already in KBase. KBase will then recruit these genomes into a specified number of phylogenetically close genomes from the KBase reference genom The tree object may be reported or viewed in KBase.

New to KBase?

Search Data

Sign In

Jetstream

Images Help

Login

globus

Products - Pricing - Developers - Support - Log In

Research data management simplified.

share transfer publish RESEARCH DATA

169,082,876,263 MB TRANSFERRED

Researchers
Focus on your research, not IT problems. We make it easy to move, manage, and share big data.

Resource Providers
Globus gives you more control over your data infrastructure, while providing excellent ease-of-use for you

Our Users
Researchers and resource providers are our greatest inspiration and we love it when they say nice things about

XSEDE USER PORTAL

Extreme Science and Engineering Discovery Environment

Search XSEDE...

SIGN IN

MY XSEDE RESOURCES DOCUMENTATION ALLOCATIONS TRAINING USER FORUMS HELP ECSS ABOUT

Summary Allocations/Usage Accounts Jobs Profile Publications Tickets Change Password Add User Community Accounts SSH Terminal

Get Started on XSEDE

Sign In

Create Account

Quick Links

- System Monitor
- Allocations
- User News
- Scheduled Downtimes
- Software Search

XSEDE USER PORTAL ON THE GO

App Store Google play

In The Past 7 Days

XD SUs Charged: Total: by Field of Science

Administrative Sciences

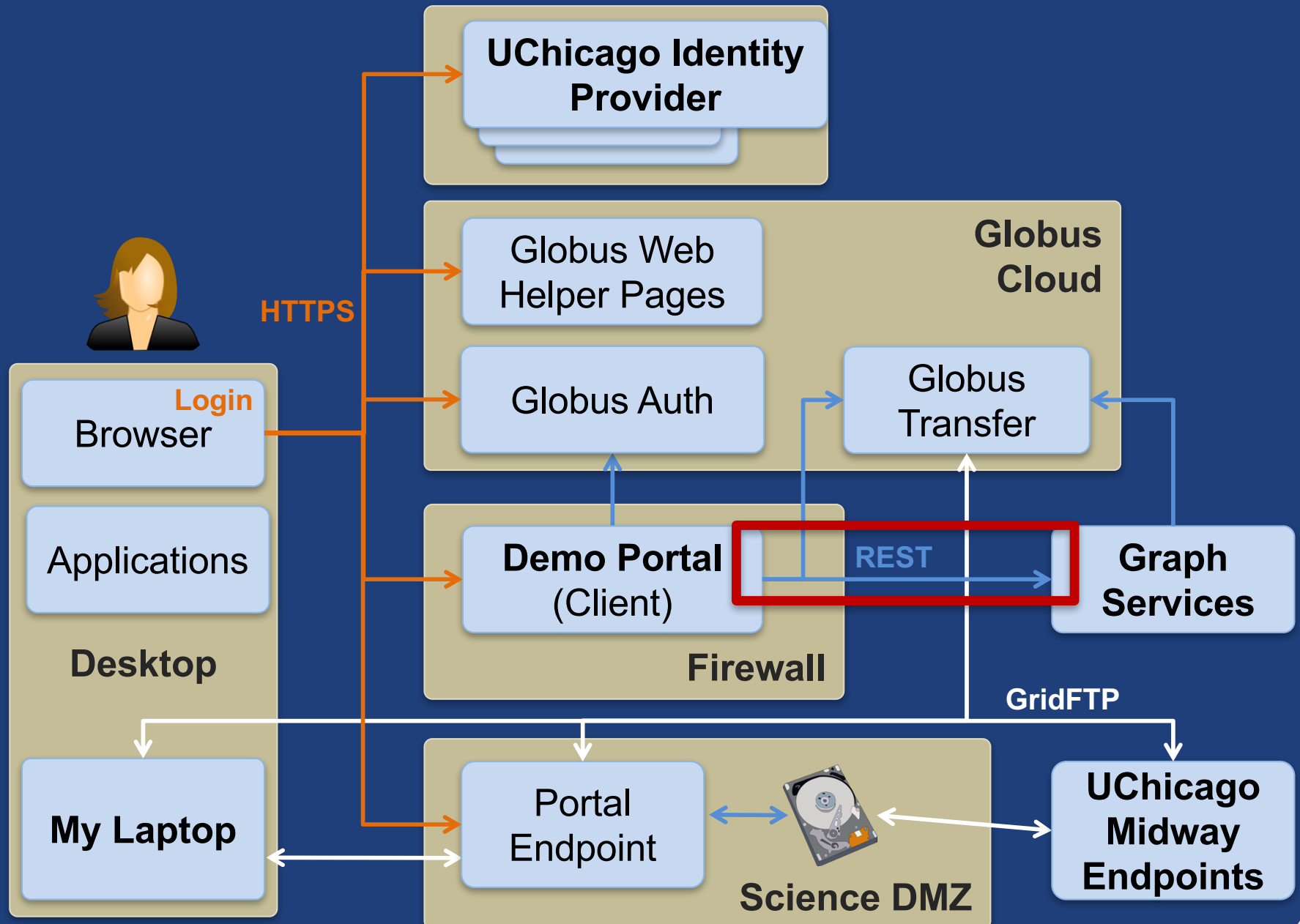


Demo

Jetstream App use of Globus Auth



Sample Research Data Portal



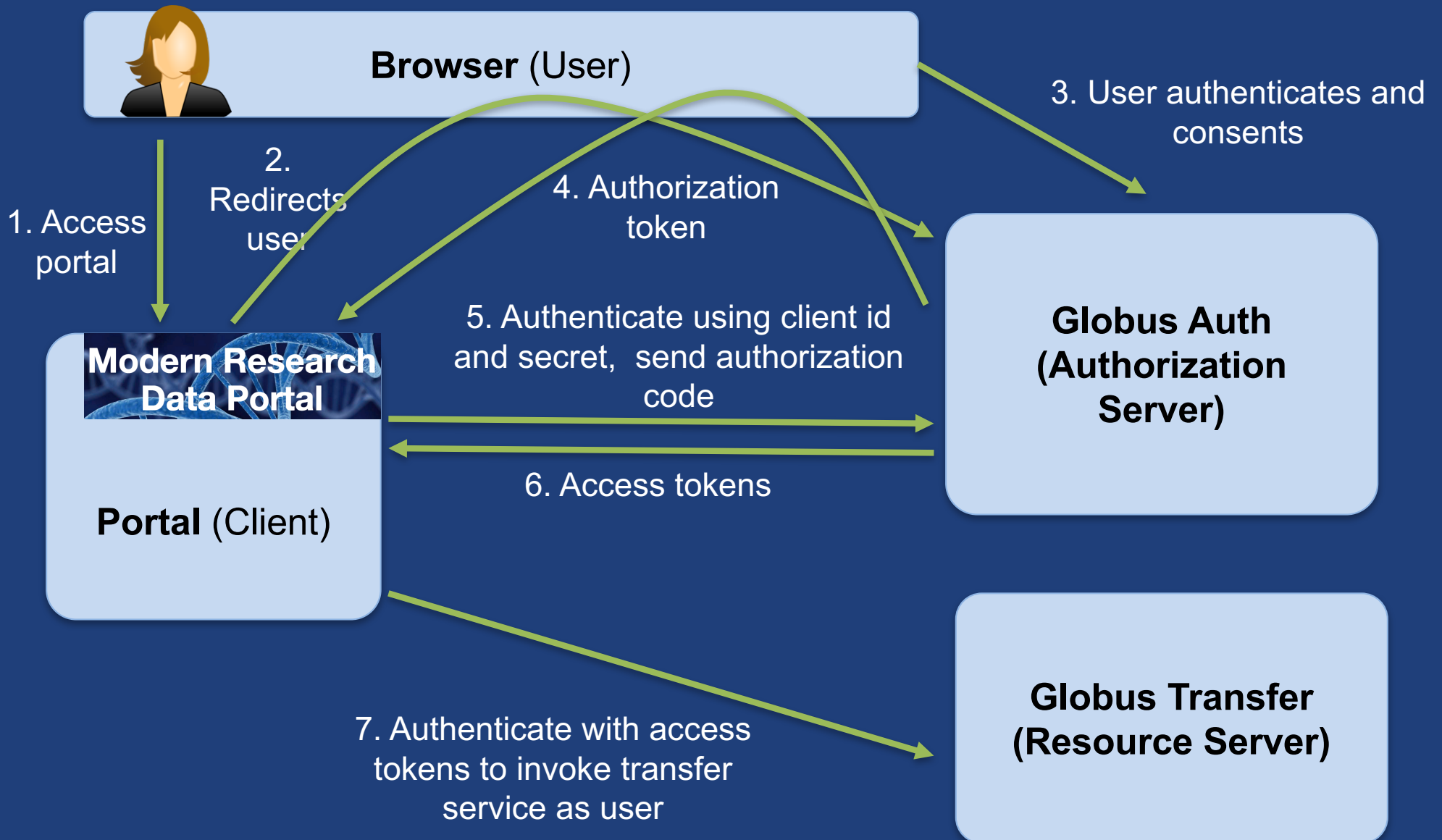


Use case: Portal calling services on user's behalf

- **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



Authorization Code Grant





App registration

- **Client_id and client_secret for service**
- **App display name**
- **Declare required scopes**
 - Need long-term, offline refresh tokens?
 - May require authorization from scope admin
- **OAuth2 redirect URIs**
- **Links for terms of service & privacy policy**
- **Effective identity policy (optional)**

developers.globus.org

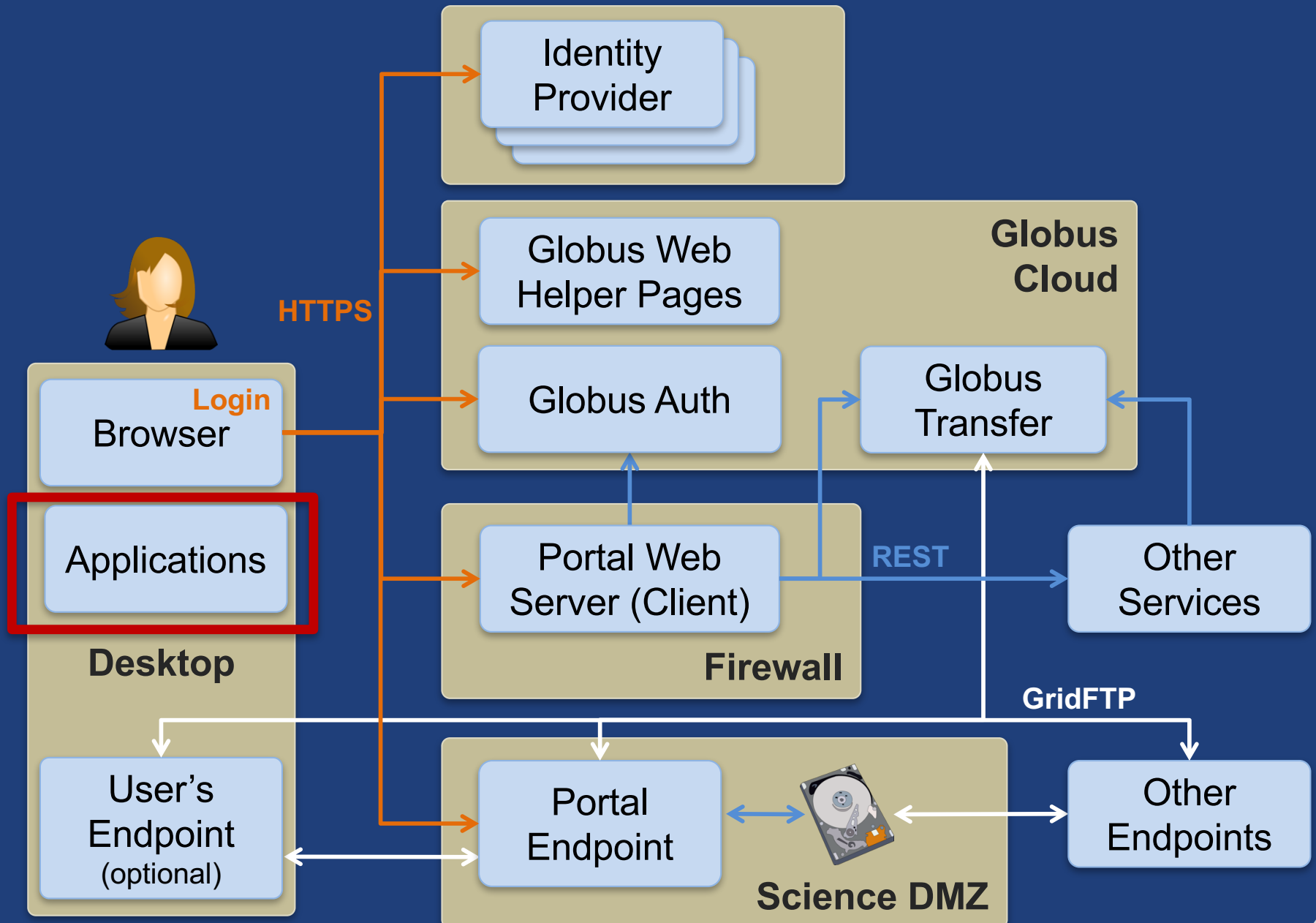


Sample Research Data Portal

Demo: Install and Register
Code walk through



Prototypical research data portal



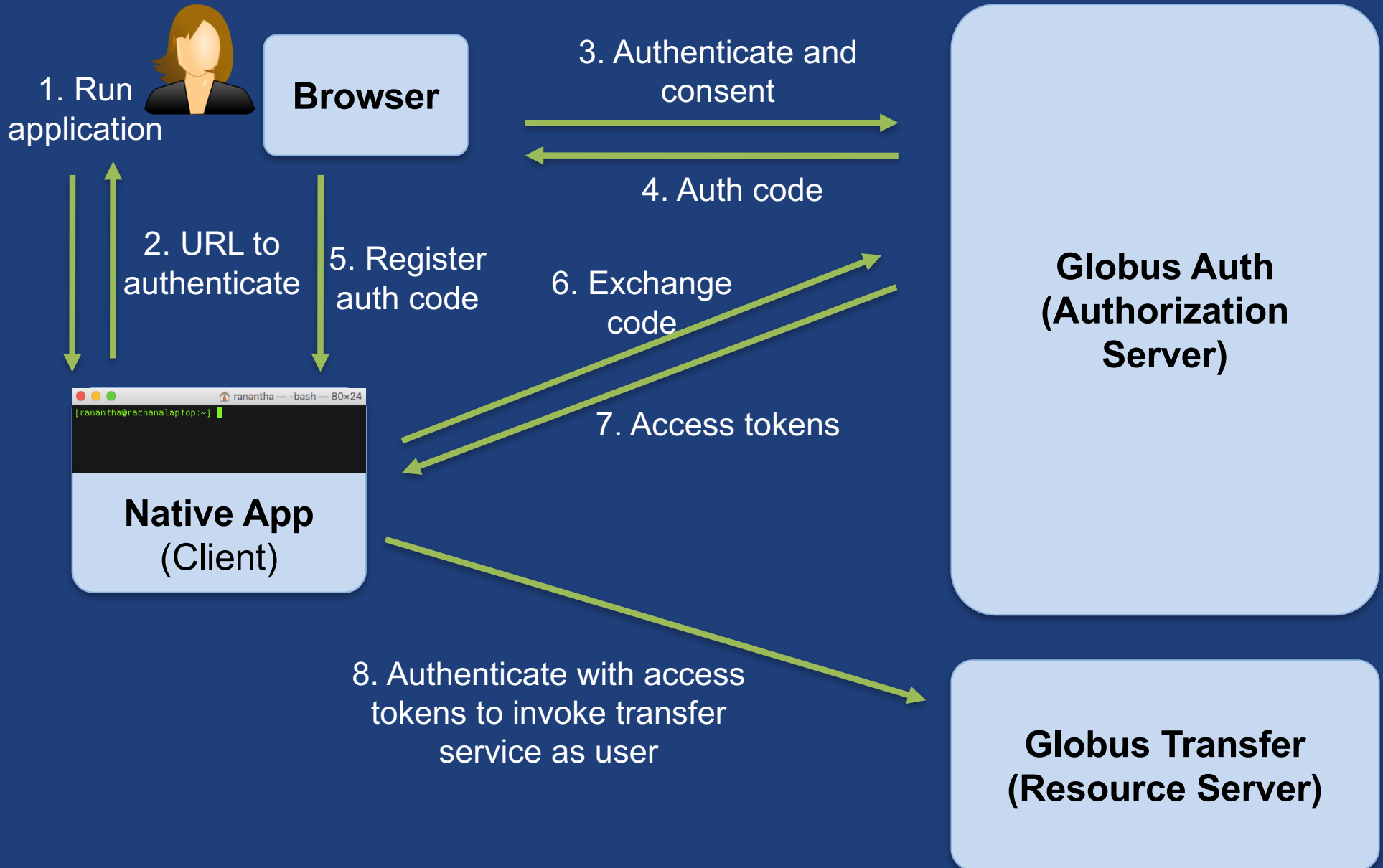


Use case: Native apps

- **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`



Native App grant





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

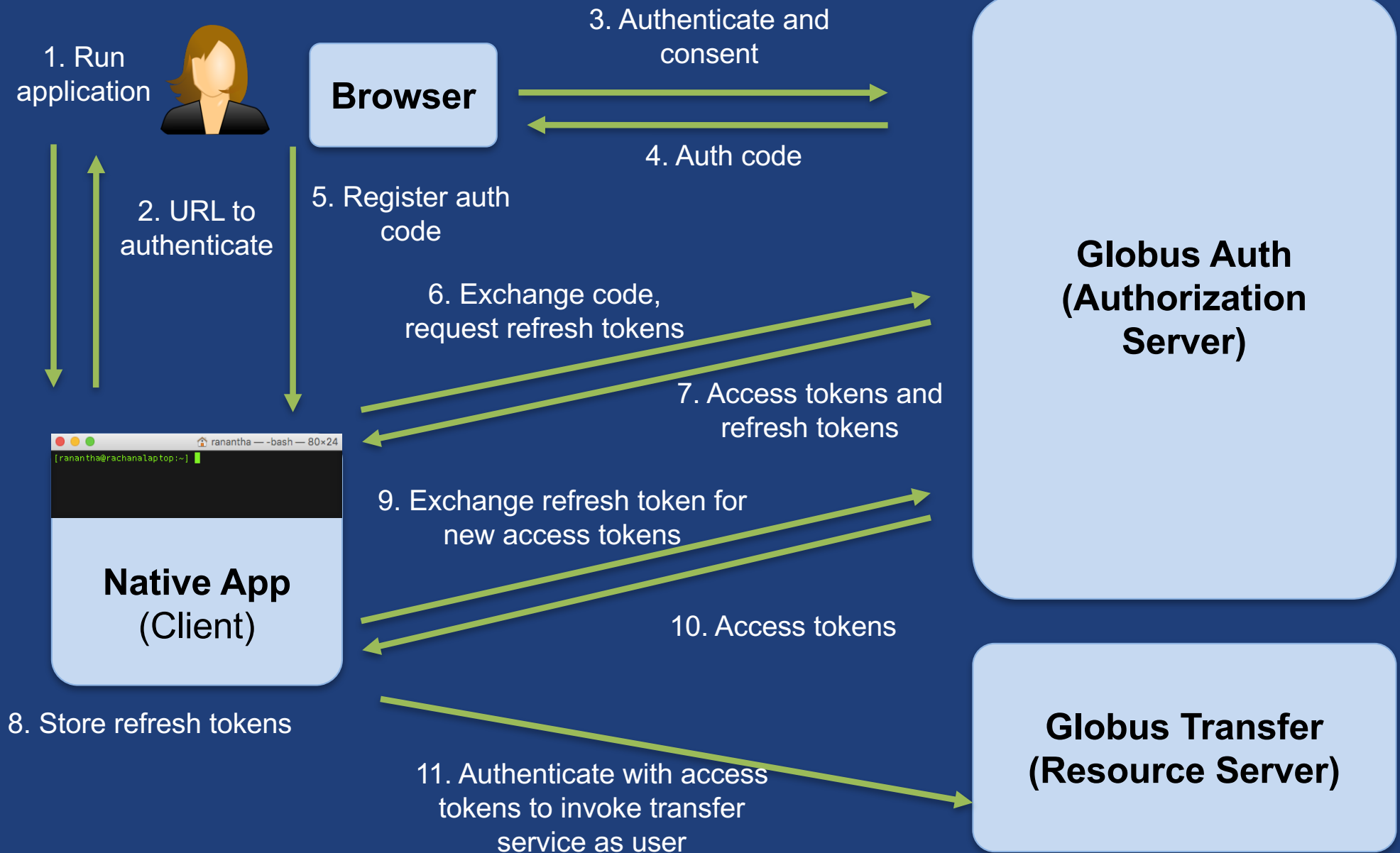


Refresh tokens

- **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**



Refresh tokens





Demo: Native App/Refresh Tokens

<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

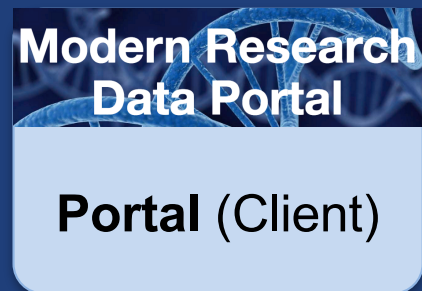


Use case: App invoking services as itself

- **Examples**
 - Sample portal invoking graph service and accessing endpoints as itself
 - Robots, agents, services
- **Every app is/has an identity in Globus Auth**
(`<client_id>@clients.auth.globus.org`)
- **App registers with Globus to get client id/secret**
 - Native app cannot do this (no `client_secret`)
- **Client Credential Grant is used**
- **Can use the `client_id` just like any other `identity_id`**
 - Sharing access manager role, permissions, group membership, etc.
- **Globus SDK:**
 - To get tokens: `ConfidentialAppAuthClient`
 - To use tokens: `AccessTokenAuthorizer`



Client credential grant



1. Authenticate with portal
client id and secret

2. Access Tokens

3. Authenticate
as portal with
access tokens to
invoke service

**Globus Auth
(Authorization
Server)**

**Globus Transfer
(Resource Server)**



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, e.g. set an ACL on a shared endpoint



Exercise: Using Client credential grant

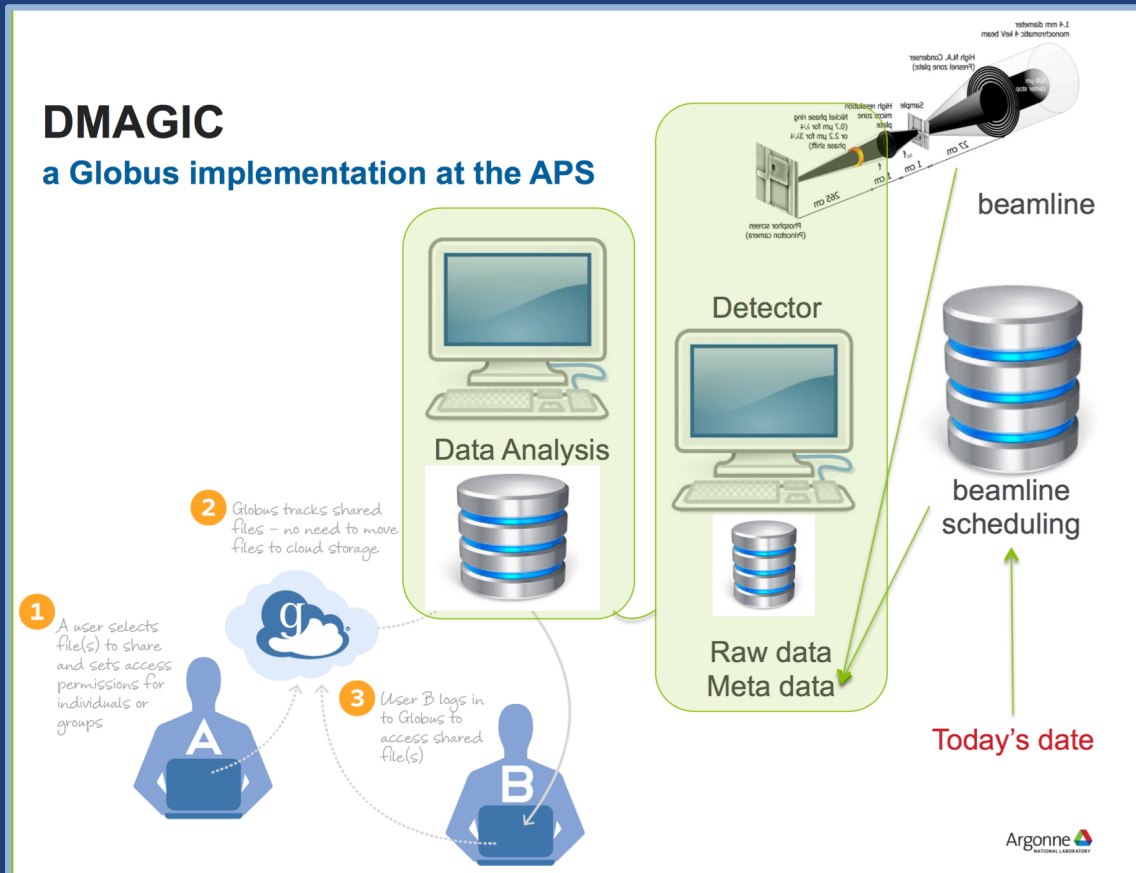
- **Start with native app examples**
- **Register a new app to get client id and secret**
- **Globus SDK:**
 - ConfidentialClientApp
 - AccessTokenAuthorizer
- **Using the Globus webapp:**
 - Create a shared endpoint
 - Set Access Manager role for the new client id
- **List files on the shared endpoint as the client identity**
- **Change permissions on the shared endpoint as the client identity**
- **Hint: Look at Jupyter notebook for SDK calls for the transfer operations**



Automating Common Tasks with Globus



Example: APS data distribution



DMagic ANL APS

Courtesy of Francesco De Carlo, Argonne National Laboratory (2016)

DMagic latest

Docs » DMagic Edit on GitHub

DMagic is an open-sourced Python toolbox to perform data management and data sharing for users of the Imaging Group of the Advanced Photon Source.

API reference
Examples
Frequently asked questions

dmagic.readthedocs.io



1. Scheduled replication

Recurring transfers
with sync option



Copy /ingest
Daily @ 3:30am



- **Using Globus CLI or SDK**
- **Meant to be run via cron or other task manager**
- **Native app grant**



2. Data distribution using sharing

Data distribution

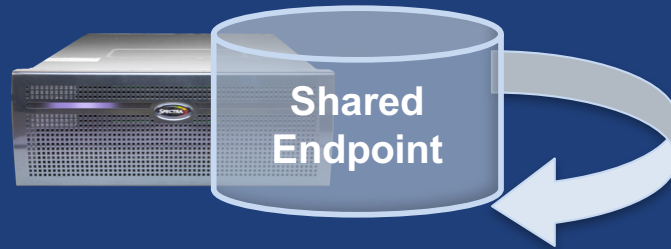


- **Uses Auth and Transfer API via SDK**
- **Native app grant**
- **Client credential grant**
 - portal or service
 - Permission for the client id



3. Monitor and clean up

Staging
area
cleanup

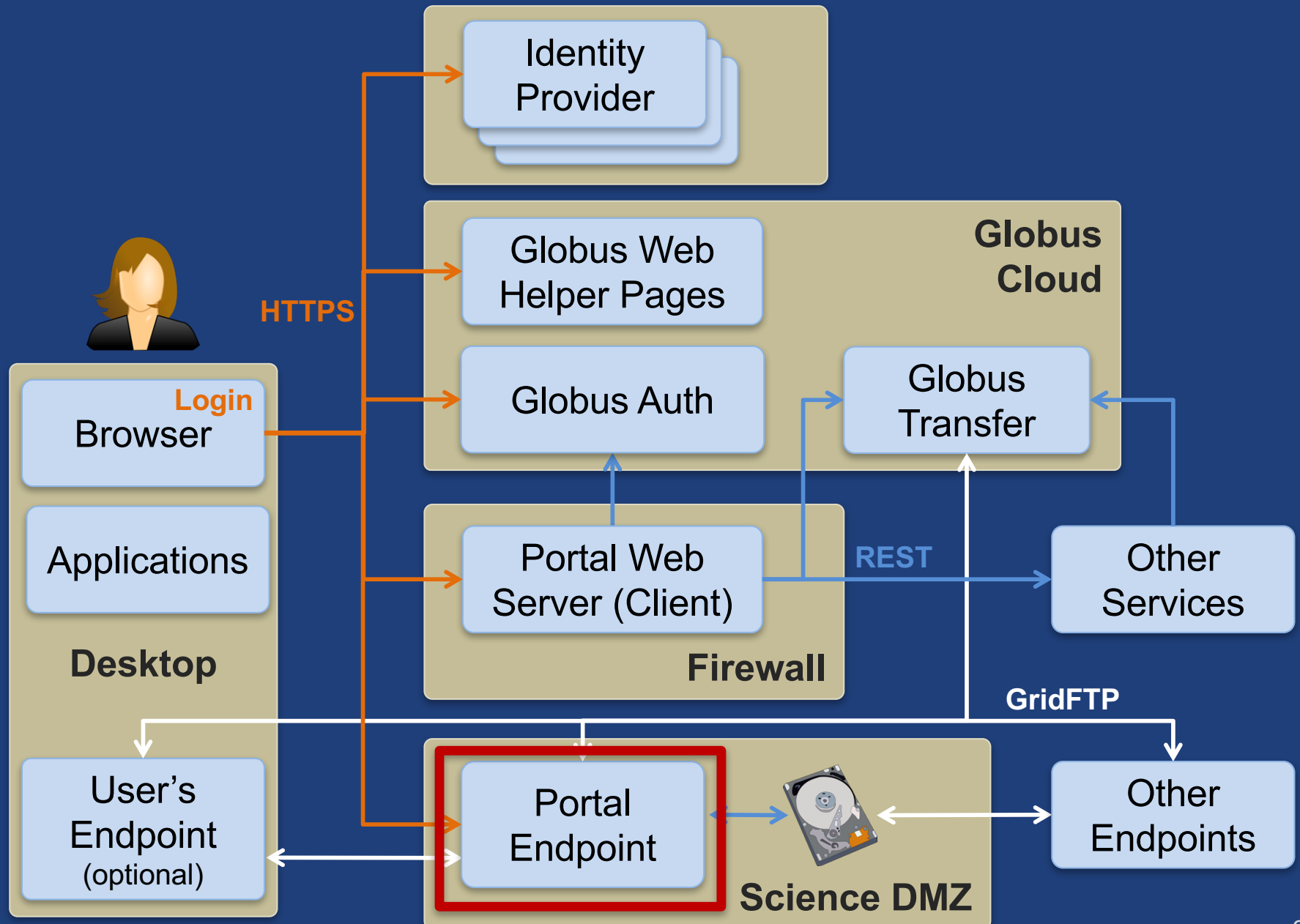


1. Check if successful transfer
2. Delete data from staging area

- **Poll model to get status**
- **Delete files**



Prototypical research data portal



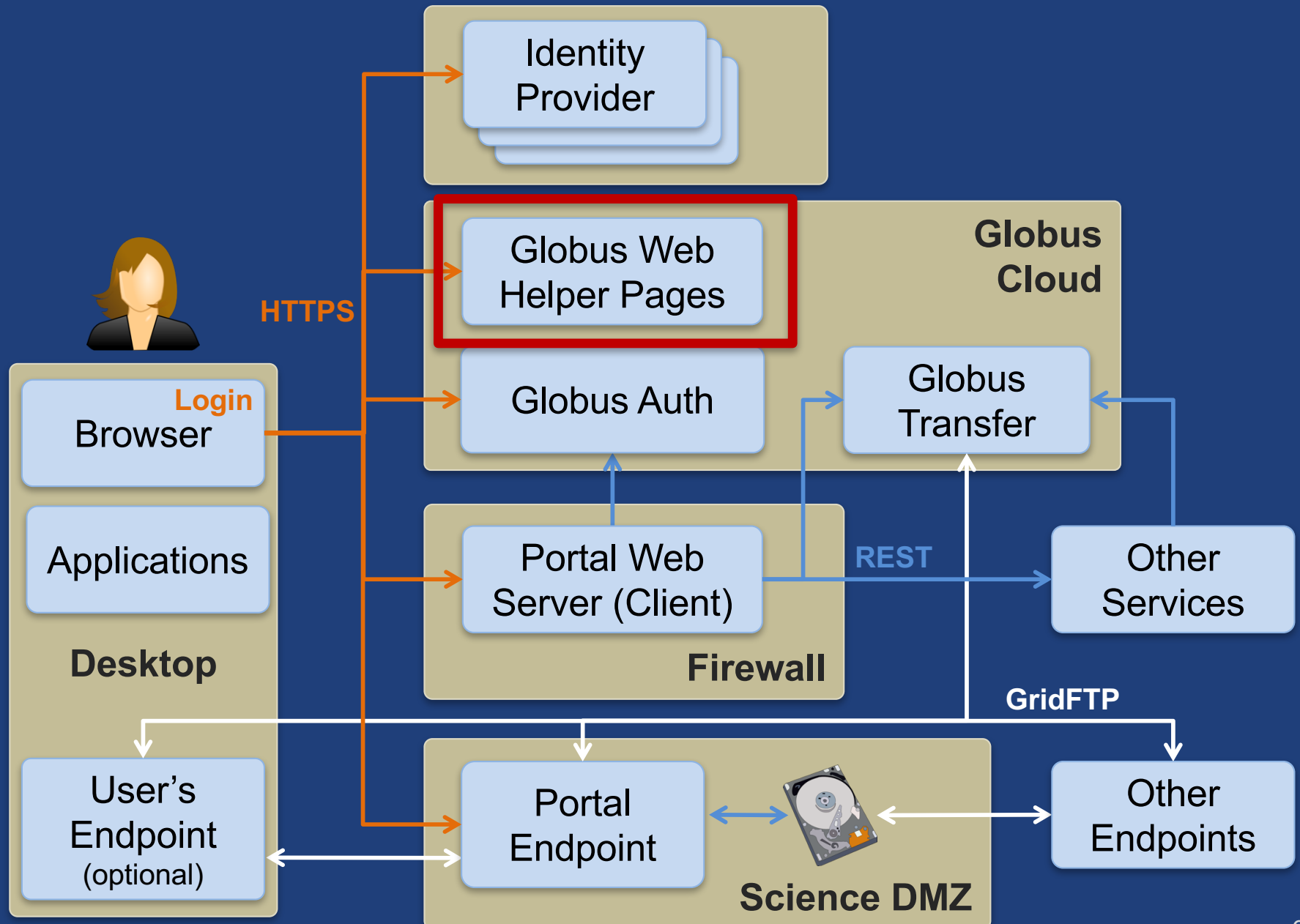


HTTPS to Endpoints

- **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



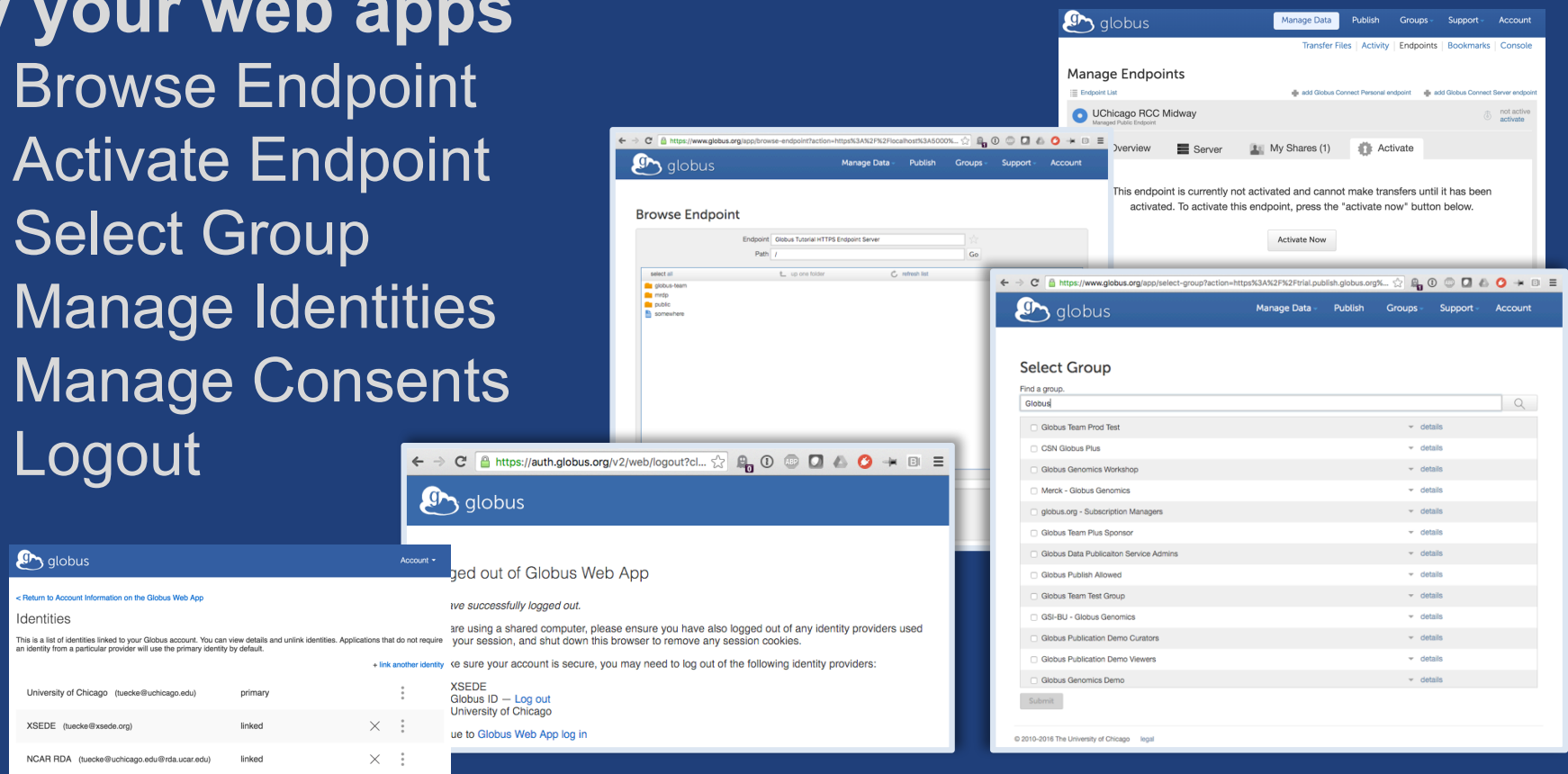
Prototypical research data portal





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

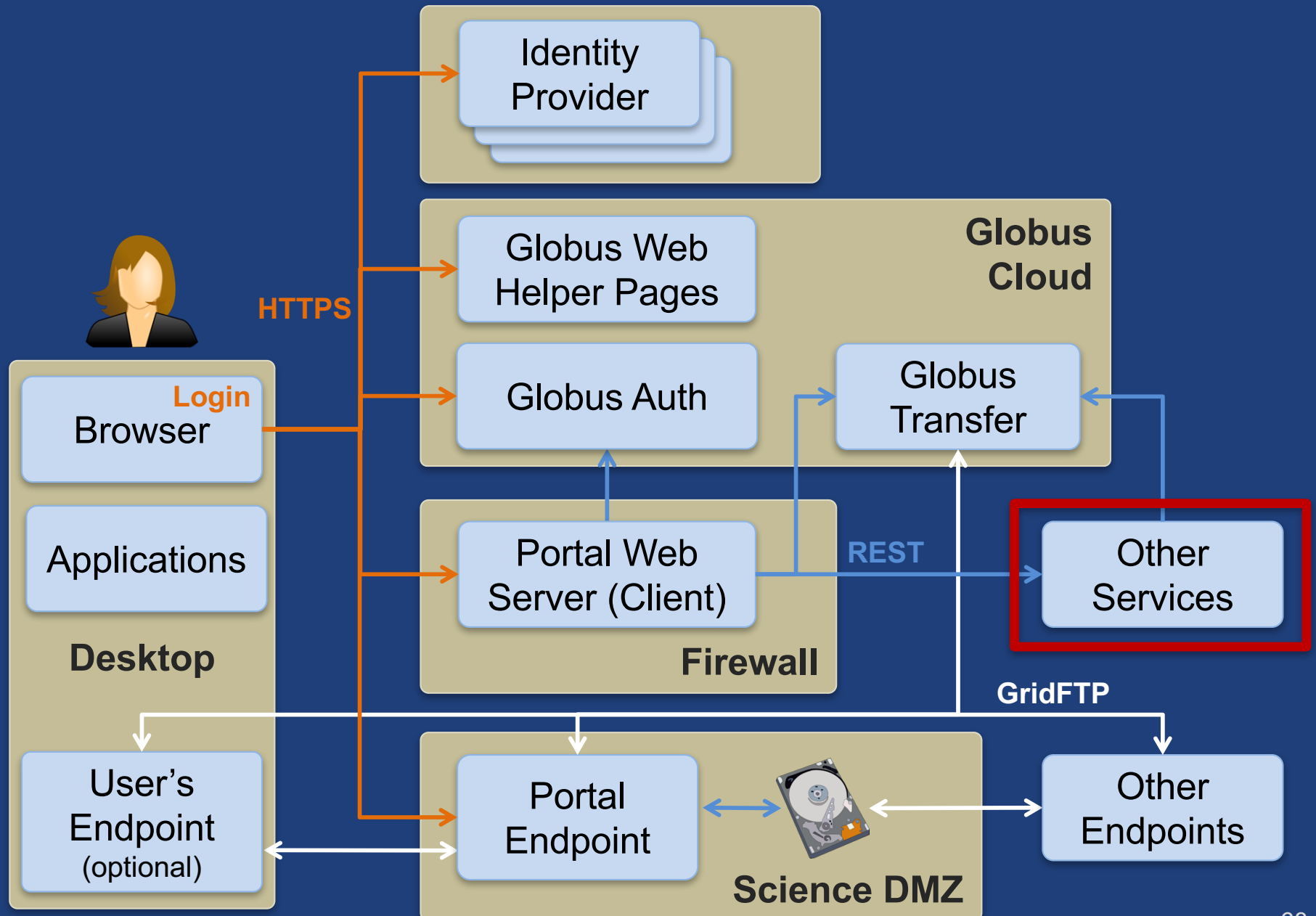


Client Logout

- **Call token revocation on access tokens**
 - <https://auth.globus.org/v2/oauth2/token/revoke>
 - Doc: docs.globus.org/api/auth/reference
 - Note: Does not revoke dependent tokens
- **Delete access tokens**
- **Redirect to logout helper page**
 - <https://auth.globus.org/v2/web/logout>
 - Doc: docs.globus.org/api/helper-pages



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 app/portal with a public REST API, so that other developers can integrate with and extend it**



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



Portal to Graph service interaction



1. Login and consent for portal and use of graph & transfer service.

2. Client credential grant to get access tokens

Globus Auth (Authorization Server)

3. Authenticate with access tokens to invoke graph service: HTTPS with access token as header

4. Authenticate with graph service client id and secret to introspect token

5. Return validity, client, scope, effective identity, identity set (for the portal)

Portal (Client)

Graph Service (Resource Server)

7. Graph service response

6. Verifies token, authorization checks





Summary of how resource works

- **Registration of resource servers**
 - Scopes
- **Dependent services**
- **Validation**



Additional Features for Service Developers



Service registration

- **Client_id and client_secret for service**
- **Service display name**
- **Validated DNS name for service**
- **One or more scopes**
- **Authorize clients to use each scope**
 - All clients (public API), or specific clients
- **Declare dependent scopes**
 - Need long-term, offline refresh tokens?
 - May require authorization from scope admin
- **Links for terms of service & privacy policy**
- **Effective identity policy (optional)**
- **Email: support@globus.org**



Effective identity

- **App or service can choose to operate only with identities from a particular identity provider**
 - Globus Auth login will require an identity from that provider to be linked to user's account
 - OIDC id_token uses this “effective identity”
- **If app or service does not set an effective identity policy, then the primary identity of the account is used as the effective identity for that app**



Branding

- Can skin Globus Auth pages

Header

Text

Default IdP

Log In using Globus

https://auth.globus.org/p/login?client_name=globus_webapp_compute canada&state=e...

globus Globus Account Log In

compute | calcul
canada | canada

Compute Canada has partnered with Globus to offer this high performance file transfer service.

Calcul Canada s'est associé à Globus pour vous offrir ce service de transfert de fichier à haute performance.

Log in to use Compute Canada Globus Web App

Use your existing organizational login
e.g. university, national lab, facility, project, Google or [Globus ID](#)
(Your Globus username and password used prior to February 13, 2016 is now Globus ID)

Compute Canada

Continue

[Why has this page changed?](#)

Didn't find your organization? Then use Globus ID to [sign up](#).



Token caching

- **Service should cache tokens and related information**
 - Improves performance of service
 - Reduces load on Globus Auth
- **Access token -> introspect response**
 - Cache timeout: 1-30 seconds recommended
 - To improve performance and load related to bursty use of REST API
 - Validity: Timeout duration determines responsiveness to token revocation and rescinding consent
 - client, scope, effective_identity: these will never change for an access token
- **Refresh tokens**
 - For however long they are needed for specific operations.



Join the Globus developer community

- Join developer-discuss@globus.org mailing lists: globus.org/mailling-lists
- Python SDK is open source
 - github.com/globus/globus-sdk-python
 - Submit issues, pull requests
 - Discussions on developer-discuss@globus.org
- All tutorial materials are open source on github
- Documentation: docs.globus.org