New Globus Features for Campuses
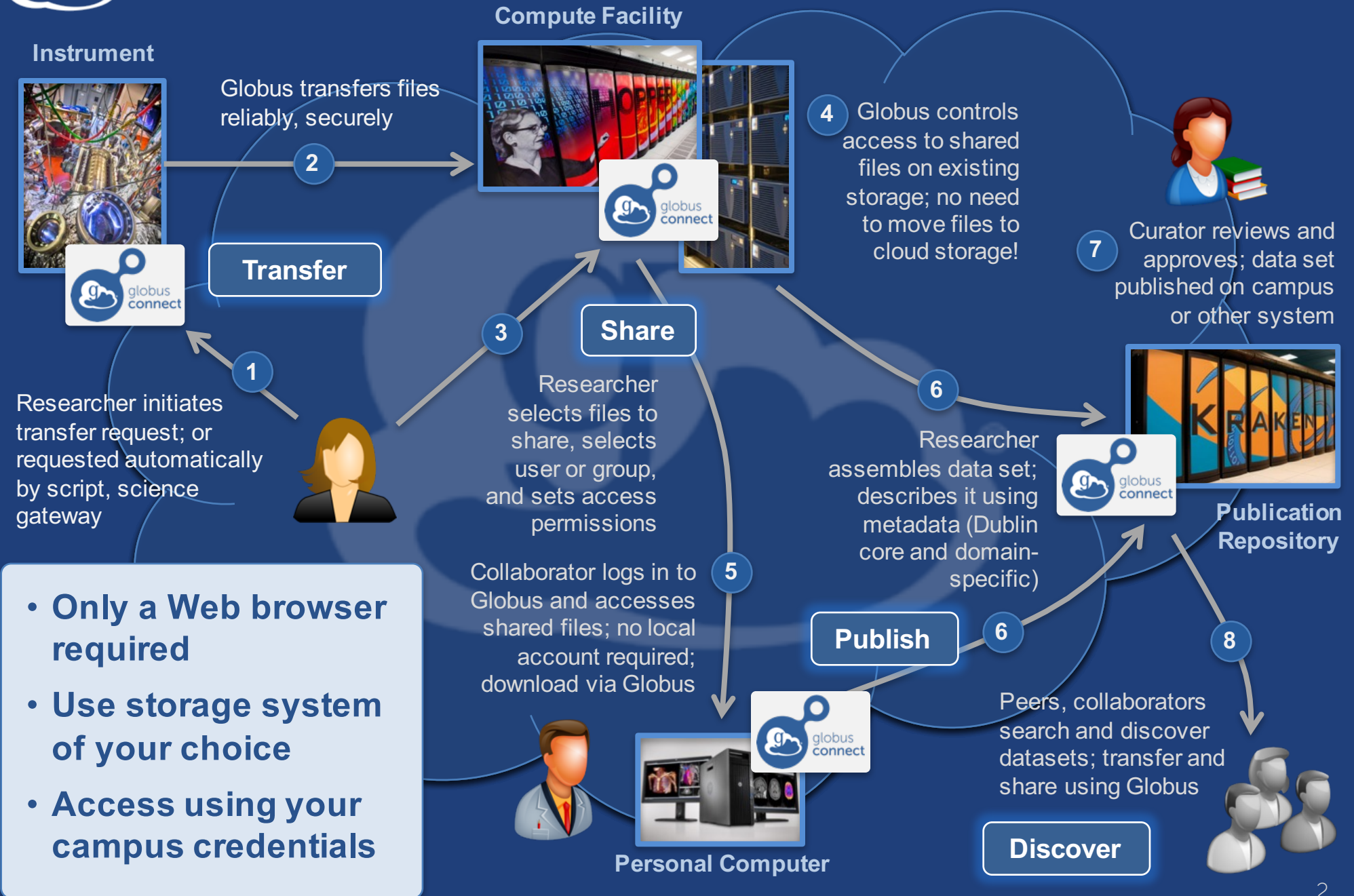
Steve Tuecke

The University of Chicago

globus

Globus SaaS: Research data lifecycle

Instrument

Compute Facility

Globus transfers files reliably, securely

2

Transfer

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

7 Curator reviews and approves; data set published on campus or other system

1

Researcher initiates transfer request; or requested automatically by script, science gateway

3

Share

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

6

Researcher assembles data set; describes it using metadata (Dublin core and domain-specific)

Publication Repository

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

5

Publish

6

Peers, collaborators search and discover datasets; transfer and share using Globus

8

- **Only a Web browser required**
- **Use storage system of your choice**
- **Access using your campus credentials**

Personal Computer

Discover

2

Globus by the numbers

4 major services	**135 PB** transferred	**20 billion** files processed	**31,000** registered users
13 national labs use Globus	**10,000** active endpoints	**~400** active daily users	**99.9%** uptime
35+ institutional subscribers	**1 PB** largest single transfer to date	**3 months** longest continuously managed transfer	**130** federated campus identities

No Globus usernames required!

- **Globus users no longer require a Globus username & password**
  - Old Globus usernames moved to separate, optional "Globus ID" identity provider

- **Any identity recognized by Globus is now sufficient to access Globus**

- **Globus Account is a primary identity plus a set of linked identities**
  - Verified email address can be a linked identity

- **Sharing & groups use any identity**

Adding your campus identity provider to Globus

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

- **OpenID Connect identity provider supported by Globus Auth** *(subscription)*

Demo new features

- **Logging into Globus with any identity**

- **Endpoint search**

- **Bookmarks**

- **Sharing with any identity**

- **Group membership with any identity**
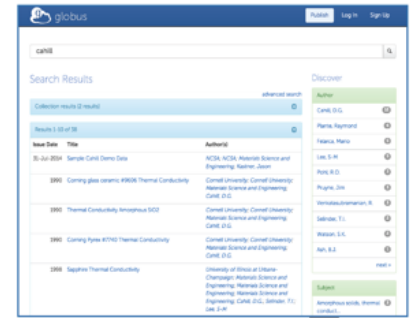
- **Management Console**

Data Publication and Discovery

Globus data publication framework
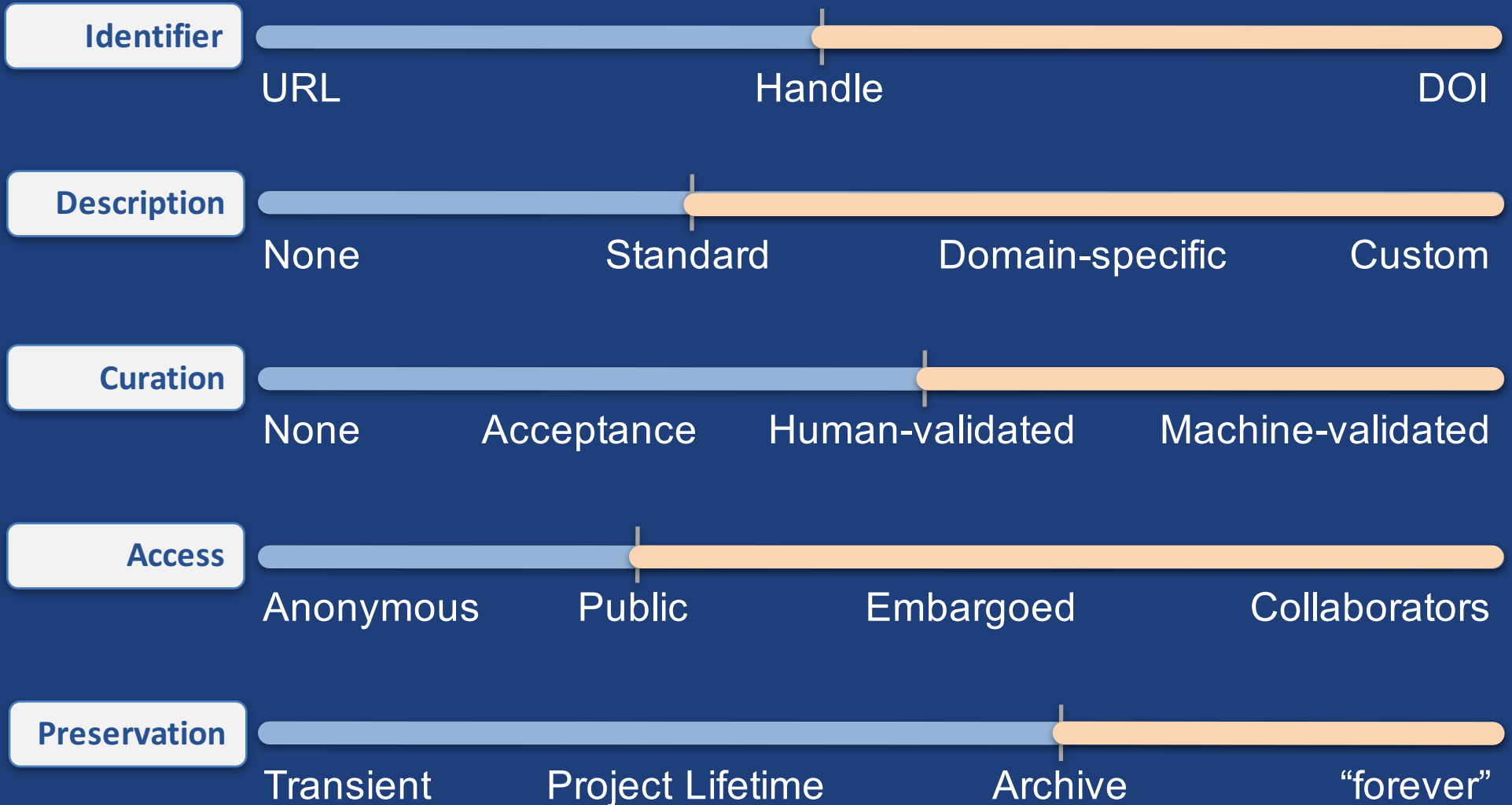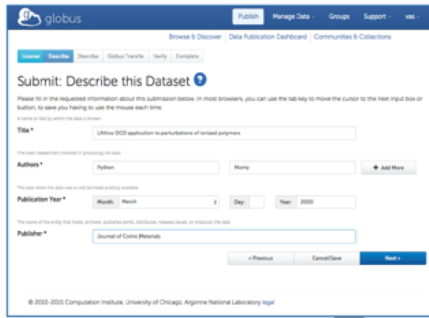
Identifier

URL   Handle   DOI

Description

None   Standard   Domain-specific   Custom

Curation

None   Acceptance   Human-validated   Machine-validated

Access

Anonymous   Public   Embargoed   Collaborators

Preservation

Transient   Project Lifetime   Archive   "forever"

Raw NGS output

Minimal metadata…

- Source environment
  - Instrument, timestamp,…
- Unique ID

High durability, low cost store

No curation

- Automated dataset acceptance

Identify…

- Handle

Glacier

Analysis results

Optional metadata…

- "Implicit" metadata
- Description through organization

Widely accessible stores

Team review

- Any collaborator may approve

Identify…

- Globus share

Peer reviewed paper data

(Re)format…

- PDF/A
- HDF
- …

Fully described…

- Dublin core metadata
- Domain metadata
- Provenance info

Replicated, public repositories

Formal, multi-step review

- Review → Update → Resubmit cycle

Persistent identifier

- DOI
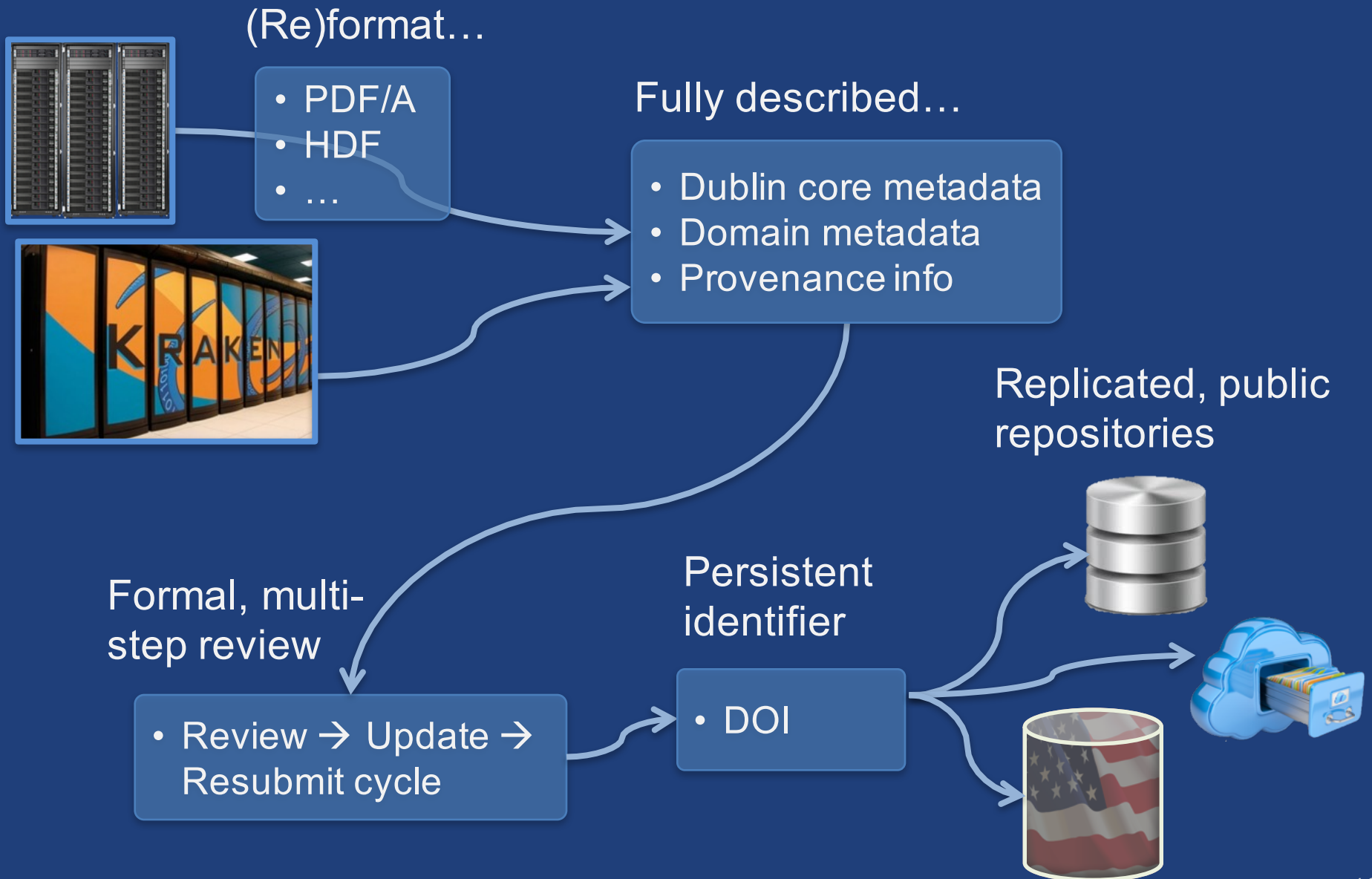
Demo

- **Publication**

- **Discovery**

HTTPS Access to Endpoints
(coming soon)

- **Enhanced use of research storage:**
  - asynchronous, bulk transfer via GridFTP
  - (soon) synchronous remote access via HTTPS

- **Enhanced Globus web app**
  - Browser-based upload/download
  - Inline file viewer

- **Integration with clients and web apps**

Cloud has transformed how platforms and software are delivered

Software as a service: SaaS
(web & mobile apps)

Platform as a service: PaaS

Infrastructure as a service: IaaS

PaaS enables more rapid, cheap, and scalable delivery of powerful apps—as SaaS

A science CI platform can spur creation of a science CI ecosystem

Software as a service: SaaS

(web and mobile apps)

Platform as a service: PaaS

Infrastructure as a service: IaaS

In so doing, we can slash costs, improve quality, and accelerate discovery across the sciences

16

A science CI platform can spur creation of a science CI ecosystem

Software as a service: SaaS

(web and mobile apps)

TripIt from Concur · NETFLIX · salesforce

2010-

Platform as a service: PaaS

amazon web services · Microsoft Azure · SOFTWARE force.com cloud platform

2014-

Infrastructure as a service: IaaS

amazon web services EC2 · amazon web services S3 · Microsoft Azure · Google Compute Engine

XSEDE Extreme Science and Engineering Discovery Environment · Open Science Grid · CASC · BLUE WATERS SUSTAINED PETASCALE COMPUTING · MICHIGAN · THE UNIVERSITY OF CHICAGO Research Computing Center · NCAR

In so doing, we can slash costs, improve quality, and accelerate discovery across the sciences

Globus PaaS: Ecosystem enabler
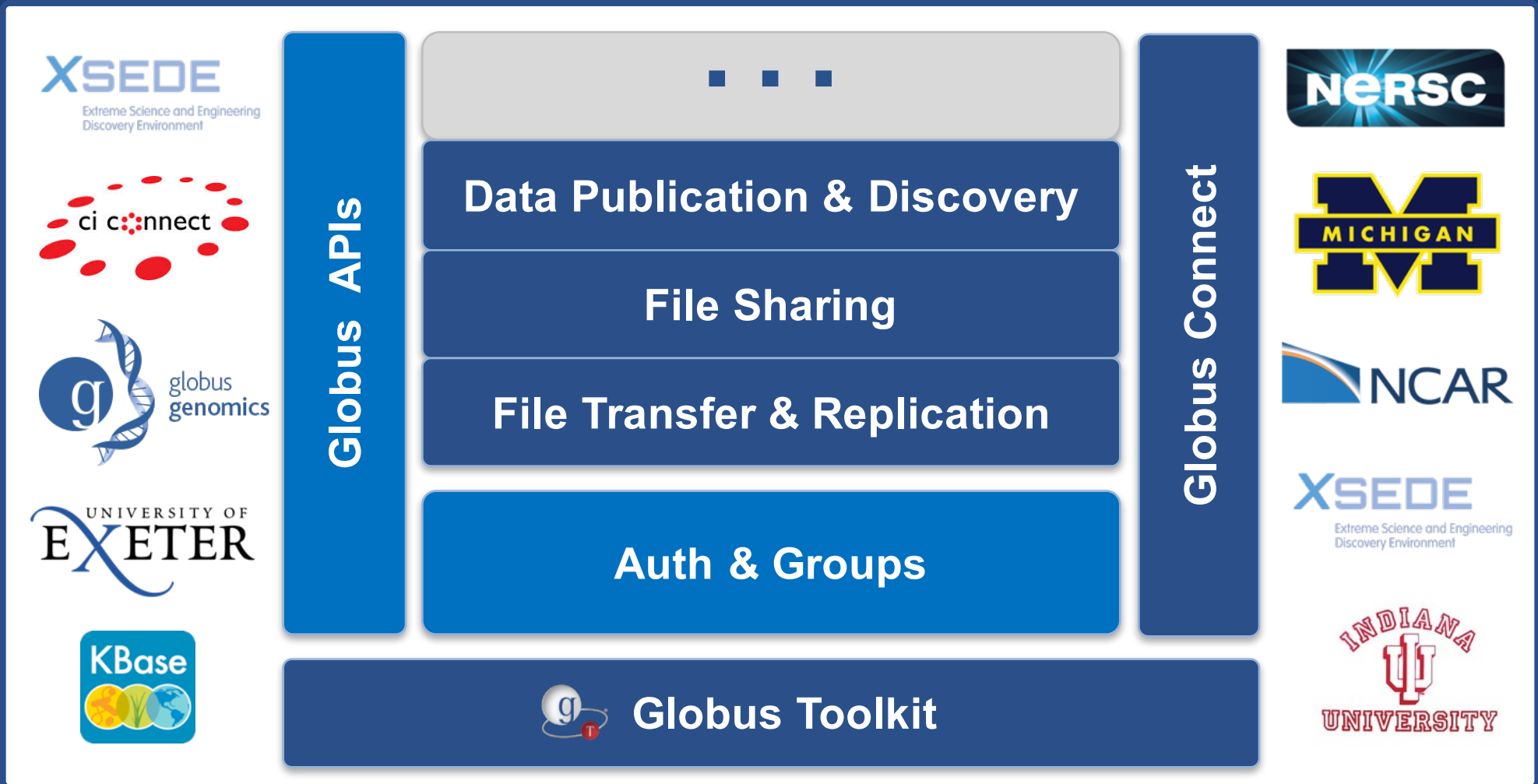
Globus PaaS at NCAR

- **Research Data Archive at NCAR**

- **Integrate Globus for data downloads**

- **Shared endpoint with subfolder per request**

- **Single sign on via streamlined account provisioning**

Find Data | **Ancillary Services** | **About/Contact**

All Datasets | Recently Added/Updated | Browse the RDA

- GCMD Topic:
  Agriculture • Atmosphere • Biosphere • Climate
  Oceans • Paleoclimate • Solid Earth • Spectral/En

- Atmospheric Reanalysis Data:
  All Reanalysis Datasets • BPRC Arctic System Reanal
  ECMWF ERA15 Reanalysis (ERA15) • ECMWF ERA40
  ECMWF Interim Reanalysis (ERA-I) • JMA Japanese 2
  JMA Japanese 55-year Reanalysis (JRA55) • NCEP Cl
  NCEP North American Regional Reanalysis (NARR) •
  NCEP/NCAR Reanalysis Project (NNRP) • NOAA-CIRE

- Station Observations:
  Land Surface Air Temperature: Hourly, Monthly

Find Platform Observations datasets

CISL Research Data Archive

Managed by NCAR's Data Support Section
Data for Atmospheric and Geosciences Research

RDA

Globus Auth

- **Foundational identity and access management (IAM) platform service**

- **Brokers authentication and authorization interactions between:**
  - end-users
  - identity providers: XSEDE, InCommon, web apps
  - resource servers: services with REST APIs
  - clients: web, mobile, desktop, command line apps
  - resource servers acting as clients to other resource servers

Based on widely used web standards

- **OAuth 2.0 Authorization Framework**
  - Globus Auth is an OAuth2 authorization server

- **OpenID Connect Core 1.0**
  - Globus Auth is an OIDC claims provider

- **Allows use of standard OAuth2 and OIDC libraries**
  - E.g., Google OAuth Client Libraries (Java, Python, etc.), Apache mod_auth_openidc
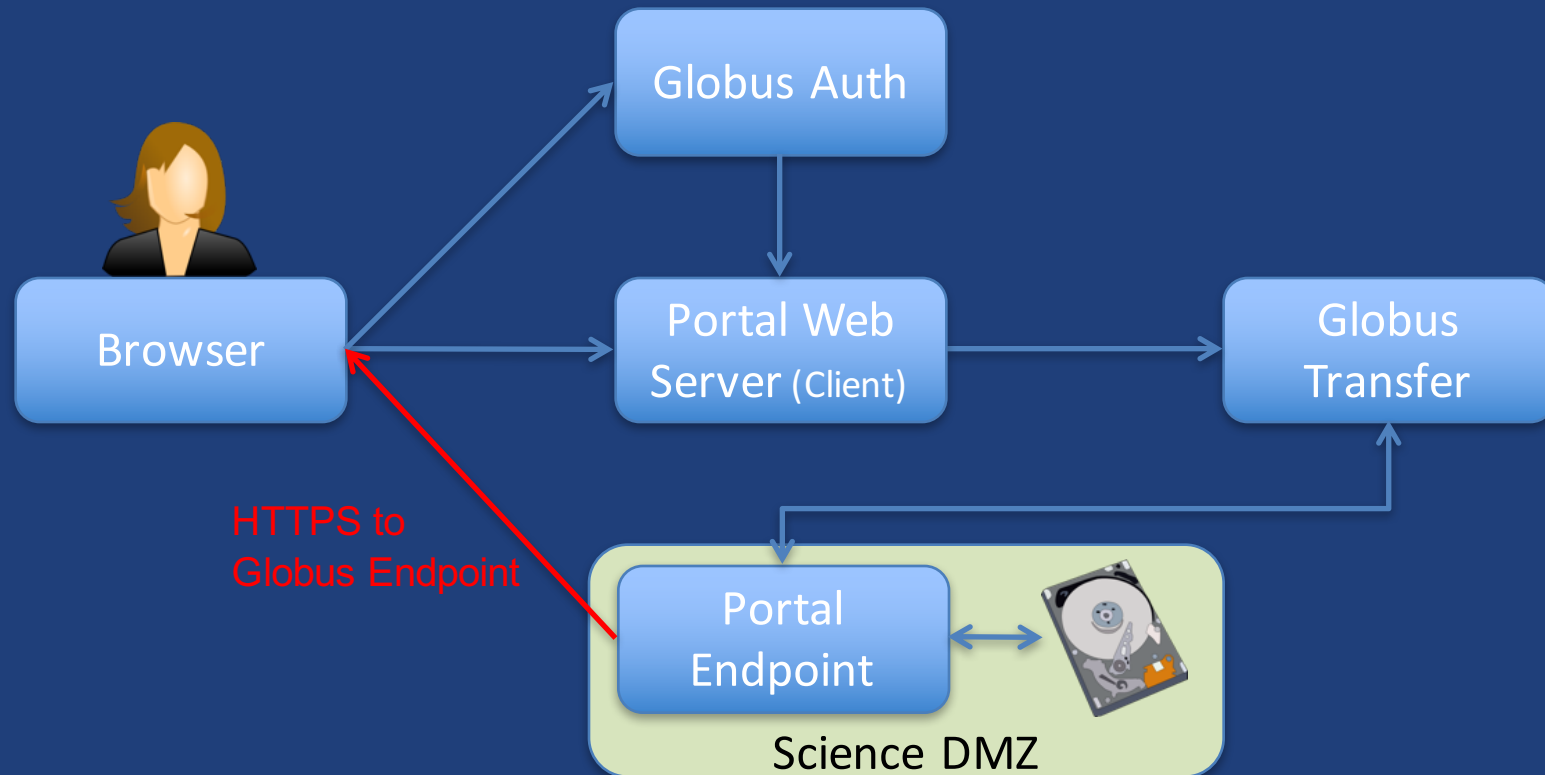
Research data portal

Research data portal

- **Move portal storage into Science DMZ, with Globus endpoint**
  - High performance, managed storage

- **Leave Portal Web server behind firewall**

- **Globus handles the data heavy lifting**

- **Globus Auth handles the security glue**

HTTPS to endpoints (coming soon)

- **Globus Connect Server will soon allow HTTPS access to endpoint storage**

- **Your web application can directly link to files on the Portal Endpoint**

- **Globus Auth and Transfer mediated security**
  - Restrict HTTPS access to files by particular users and groups

Other resource servers

XSEDE services integration

Web App Server (Client)

(2) access_token$_1$

Globus Transfer

Browser

(1) id_token, access tokens

Globus Auth

(3) access_token$_2$

XSEDE services

Globus Web App
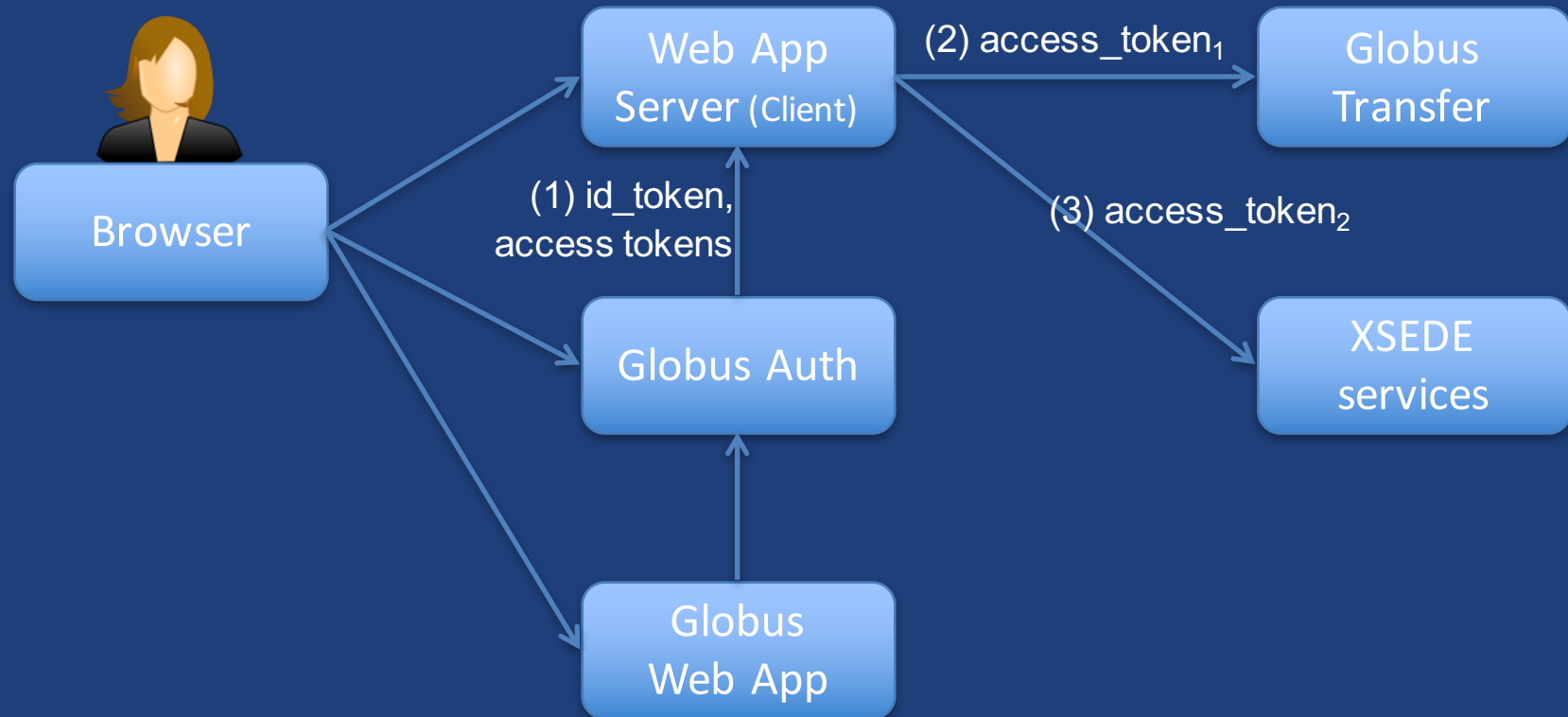
- **OAuth2 Authorization Code Grant with Globus Auth**
  - Scoprofile urn:globus:auth:scope:transfer.api.globus.org:all
    urn:globus:auth:scope:api.xsede.org:all
  - Globus Auth returns OIDC id_token & OAuth2 access tokens to clpes: openid email ient
- **Globus Auth returns different access tokens for different resource servers**
- **Web App Server calls each resource server with appropriate access token**
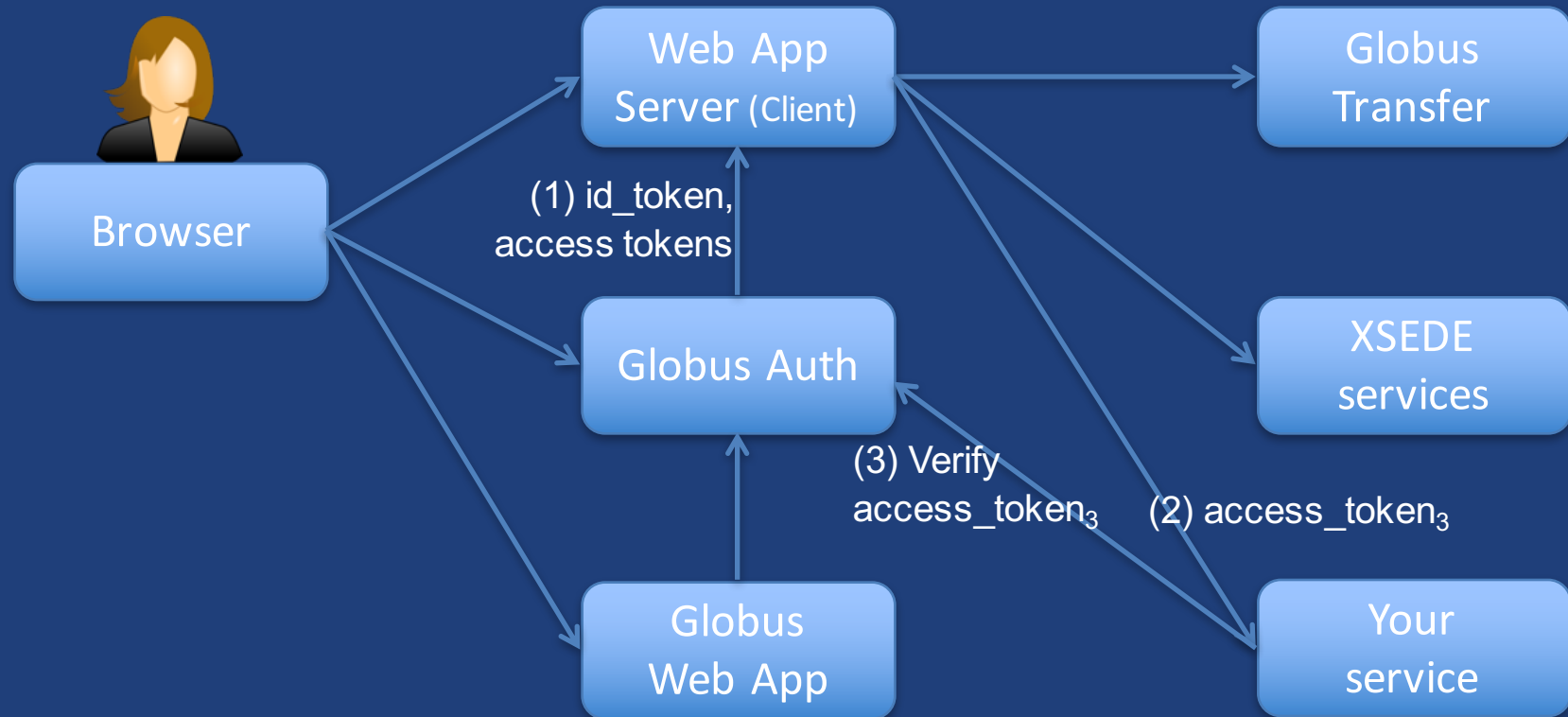
Add your own resource servers

- **OAuth2 Authorization Code Grant with Globus Auth**
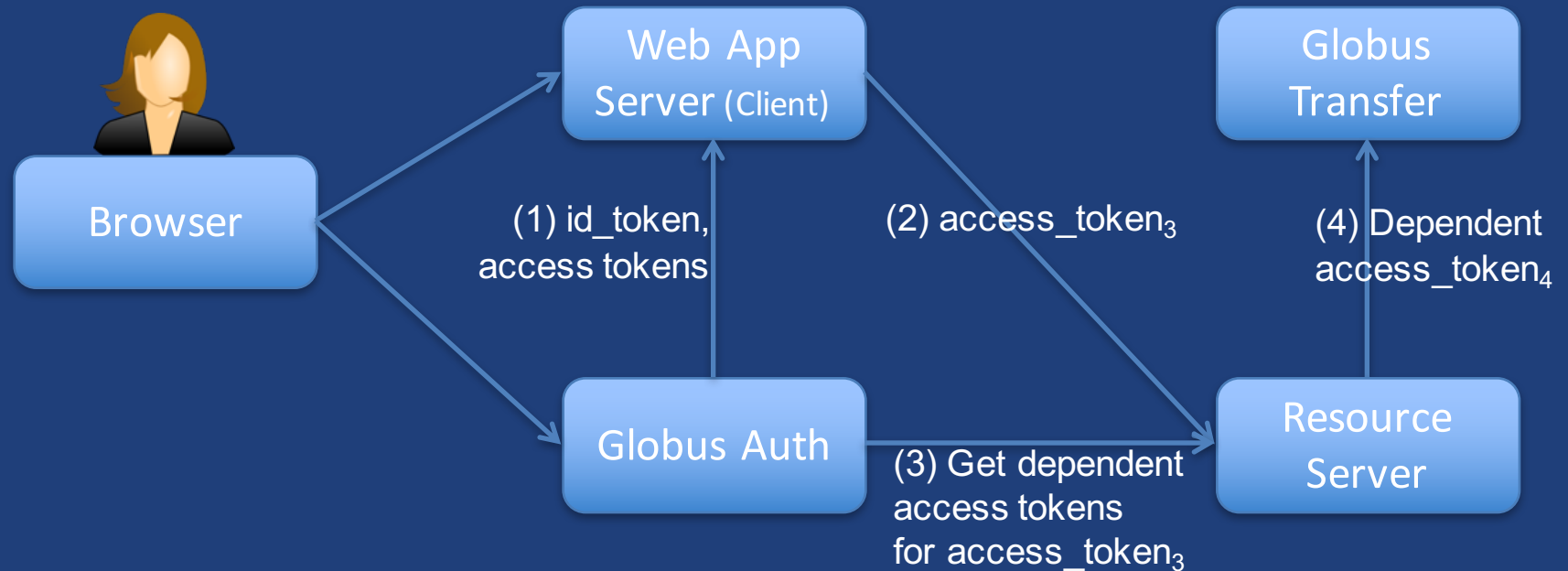 - Scopes: openid email profile urn:globus:auth:scope:transfer.api.globus.org:all urn:globus:auth:scope:api.xsede.org:all urn:globus:auth:scope:api.example.com:all
  - Globus Auth returns OIDC id_token & OAuth2 access tokens to client

- **Resource Server must register with Globus Auth**
  - Resource server policy can require identity from a particular identity provider

Dependent resource servers

Diagram: Browser → Web App Server (Client); Browser → Globus Auth. Web App Server (Client) → (1) id_token, access tokens. Web App Server → Resource Server: (2) access_token₃. Globus Auth → Resource Server: (3) Get dependent access tokens for access_token₃. Resource Server → Globus Transfer: (4) Dependent access_token₄.

- **OAuth2 Dependent Token Grant with Globus Auth**
  - Scopes: openid email profile urn:globus:auth:scope:api.example.com:all

- **Resource Server registers its Globus Transfer dependency with Globus Auth**

- **Resource Server uses request access token to get dependent access tokens**

- **Resource Server uses dependent access token to call Globus Transfer**

Mobile applications

- **Globus Auth will be adding support for mobile apps**
  - "Log in with Globus" in mobile apps
    - RFC 7636: Extension to OAuth2 to allow OAuth2 Authorization Code Grant to work from mobile apps
  - Mobile apps can call any resource server REST APIs that use Globus Auth
  - iOS and Android

An extensible platform for CI

Apps

Domain-independent and domain-specific services

AWS, Google, Azure services

Foundation services: Globus Auth, Groups, etc.

- Globus provides foundation and other services
- Community can extend to meet domain-specific and domain-independent needs

Developer Workshop: Building the Modern Research Data Portal

New high-speed networks make it possible, in principle, to transfer and share research data at tremendous speeds and scales—but have also proved challenging to use in practice. Two new technologies now allow us to translate this potential into reality: Science DMZ architectures provide frictionless end-to-end network paths; and Globus APIs allow programmers to create powerful research data portals that leverage these paths for data distribution, synchronization, and other useful purposes.

Come to Chicago in April to learn more!

Introduction, Concepts, and Components

IMPERIAL 2

Led by: **TBD**

We will introduce the Modern Research Data Portal and set the context for how Globus and the ScienceDMZ combine to deliver unique data management capabilities. This will include:

- Overview of use cases: Common patterns like data publication/distribution, orchestration of data flows, etc.

- Overview of the Globus platform: Architecture and brief overview of available services

- Introduction to the Globus Auth API: Authenticating and authorizing a client

- Introduction to the Globus Transfer API: Make your first call and move data with Globus

- Introduction to the Python SDK for using Globus Auth and Transfer
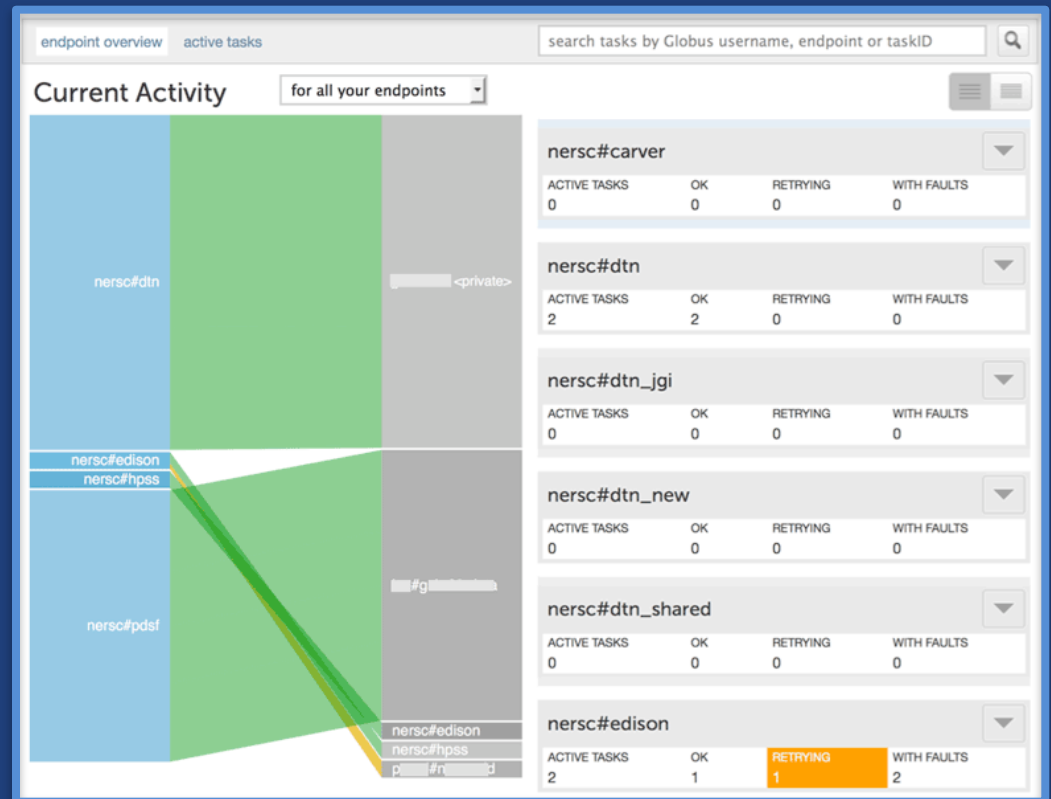
Globus Provider Subscriptions

- **Globus Provider Plan**
  - Shared endpoints
  - Data publication
  - Amazon S3 endpoints
  - Management console
  - Usage reporting
  - Priority support
  - Application integration

- **Branded Web Site**

- **Alternate Identity Provider (InCommon is standard)**

- **Mass Storage System optimization**

globus.org/provider-plans

Summary

- Globus SaaS continues to improve its capabilities for researchers and campus research computing centers

- Globus PaaS opens new opportunities for developers creating web and mobile applications for researchers

Thank you to our sponsors!

U.S. DEPARTMENT OF **ENERGY**

NSF

THE UNIVERSITY OF CHICAGO

NATIONAL INSTITUTES OF HEALTH

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

ALFRED P. SLOAN FOUNDATION
1934

Argonne
NATIONAL LABORATORY

powered by
amazon
web services