globus labs

Globus as a Case Study in Sustainable Research Software Development

Kyle Chard (chard@uchicago.edu)

Globus is a …

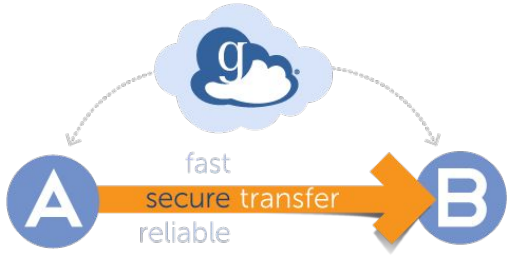
non-profit service developed and operated by

THE UNIVERSITY OF CHICAGO

Our mission is to…

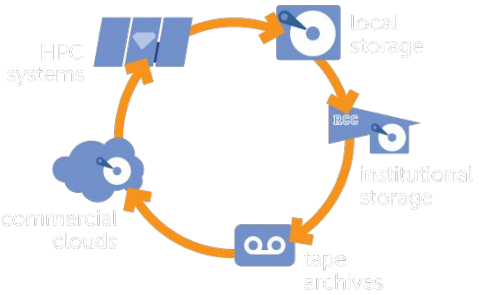
increase the efficiency and effectiveness of researchers engaged in data-driven science and scholarship through *sustainable* software.

Globus: Platform for data-driven research

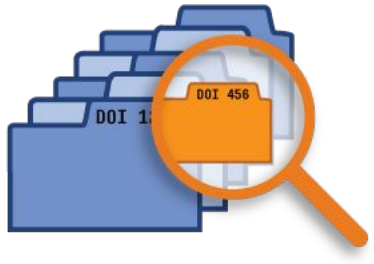


Managed transfer & sync

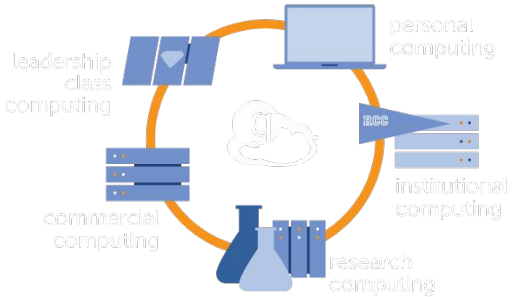
Collaborative data sharing



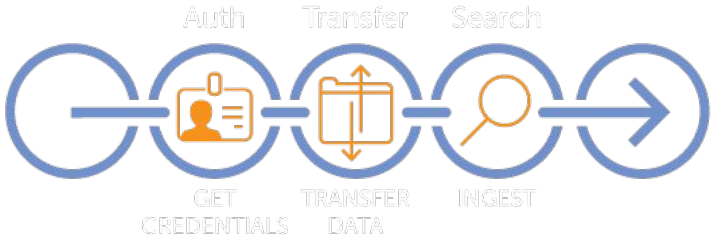
Unified data access



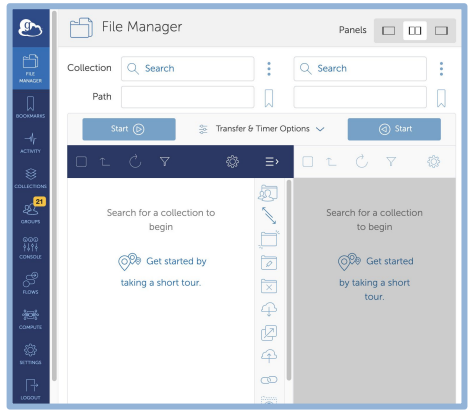
Publication & discovery



Managed remote execution

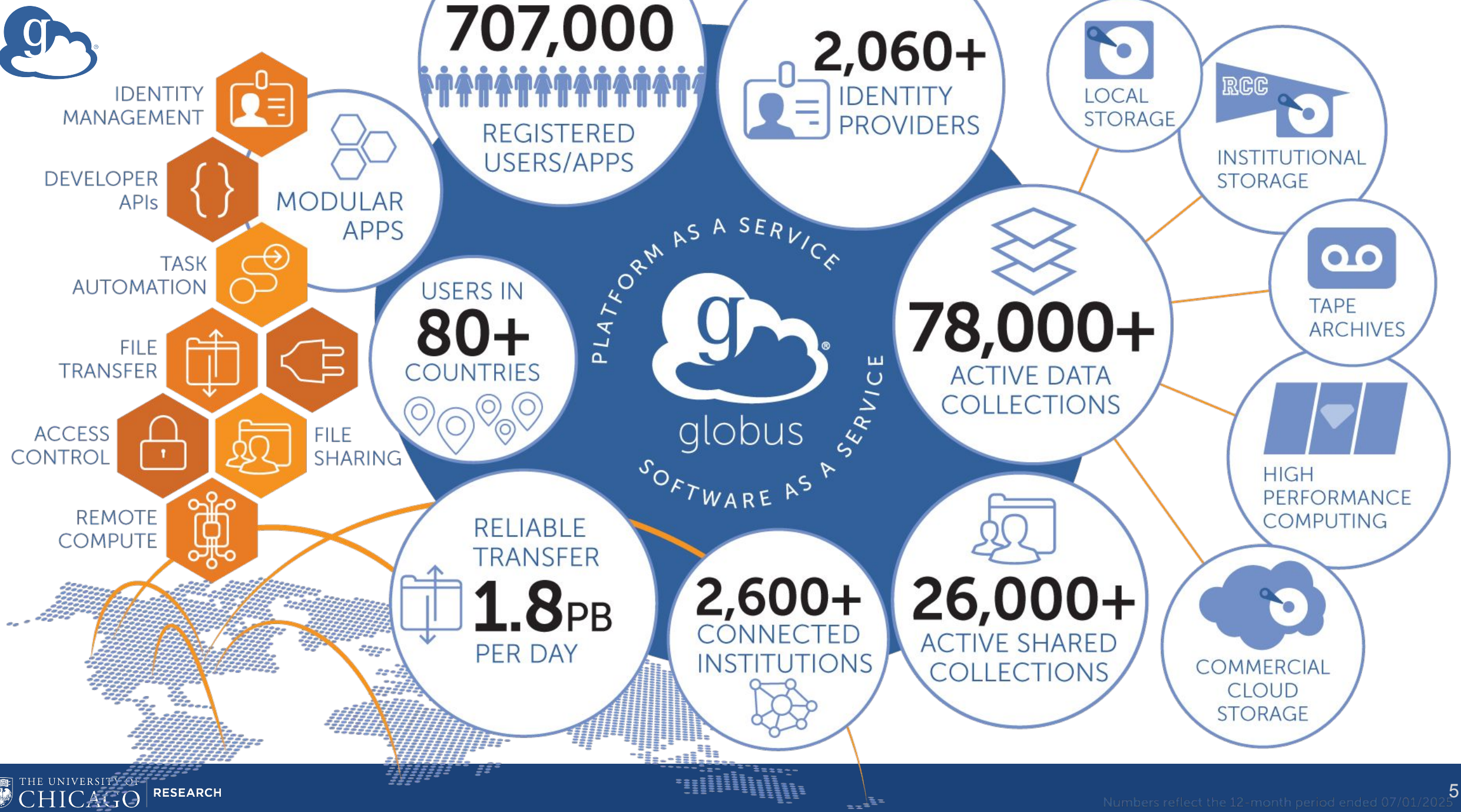


Reliable automation



Software-as-a-Service

Platform-as-a-Service


707,000
REGISTERED USERS/APPS
2,060+
IDENTITY PROVIDERS
LOCAL STORAGE
RCC
INSTITUTIONAL STORAGE
IDENTITY MANAGEMENT
DEVELOPER APIs
MODULAR APPS
TASK AUTOMATION
FILE TRANSFER
ACCESS CONTROL
FILE SHARING
REMOTE COMPUTE
USERS IN 80+ COUNTRIES
PLATFORM AS A SERVICE
globus
SOFTWARE AS A SERVICE
78,000+
ACTIVE DATA COLLECTIONS
TAPE ARCHIVES
HIGH PERFORMANCE COMPUTING
RELIABLE TRANSFER
1.8PB
PER DAY
2,600+
CONNECTED INSTITUTIONS
26,000+
ACTIVE SHARED COLLECTIONS
COMMERCIAL CLOUD STORAGE

The Grid: A vision for seamless on-demand computing

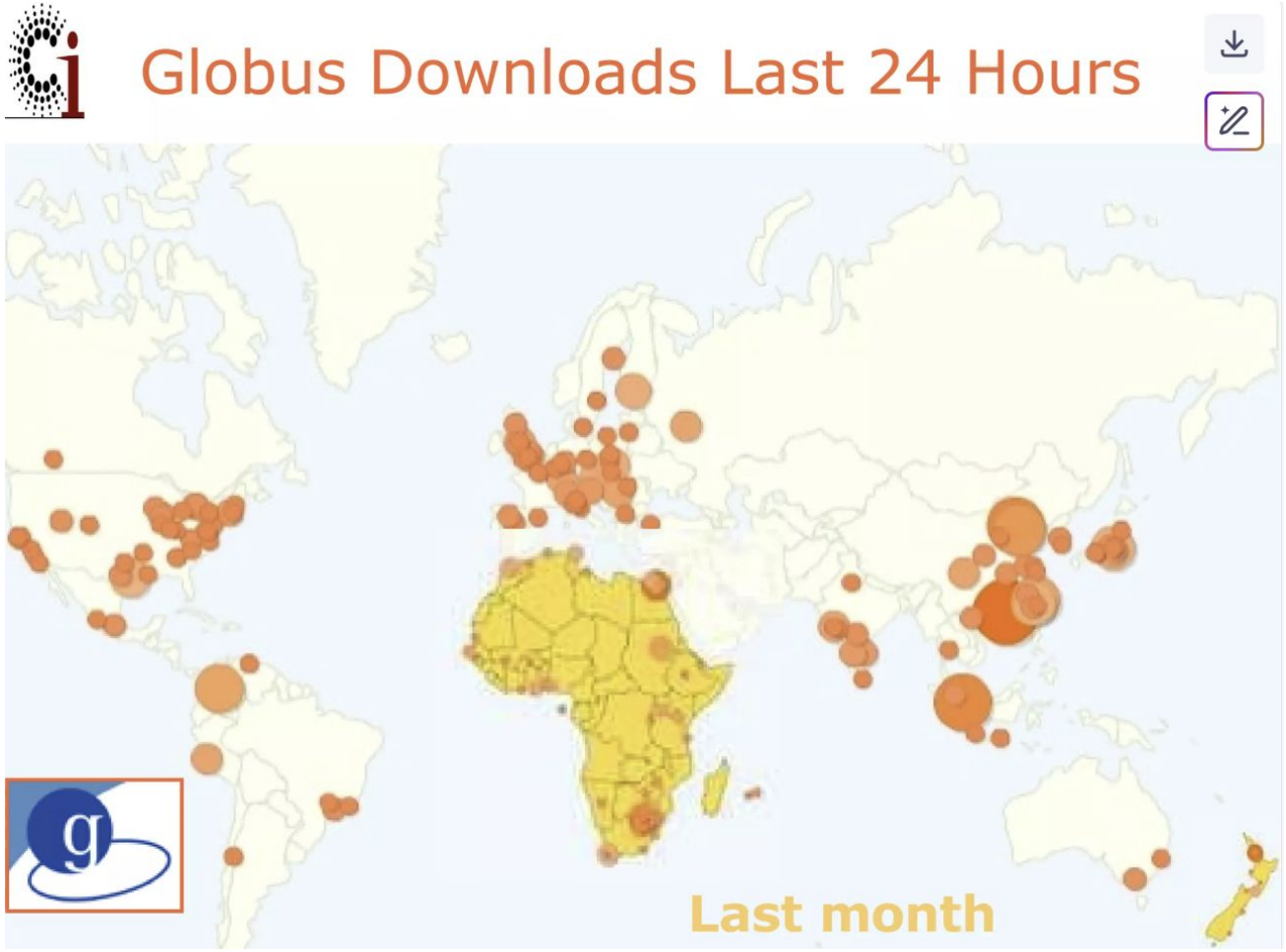
the globus toolkit®
www.globustoolkit.org

Accelerate discovery & innovation by providing **on-demand access to computing**

"if mechanisms are in place to allow reliable, transparent, and instantaneous access to high-end resources, then it is as if those resources are devoted to them"
(The Grid, Chapter 2)

Ian Foster
Carl Kesselman
Steve Tuecke

Globus Toolkit adoption and impact



Foster, 2008

Grid Computing impact

Discovery of Higgs Boson: Physics, 2013
"The results today are only possible because of the extraordinary performance of the accelerators, including the infrastructure, the experiments, and the **Grid computing**."

IPCC climate assessment: Peace, 2007

Earth System Grid enables sharing of simulation outputs

Detection of gravitational waves: Physics, 2017

LIGO scientific collaboration uses grid technologies to pool data and computing

What does sustainable software really mean?

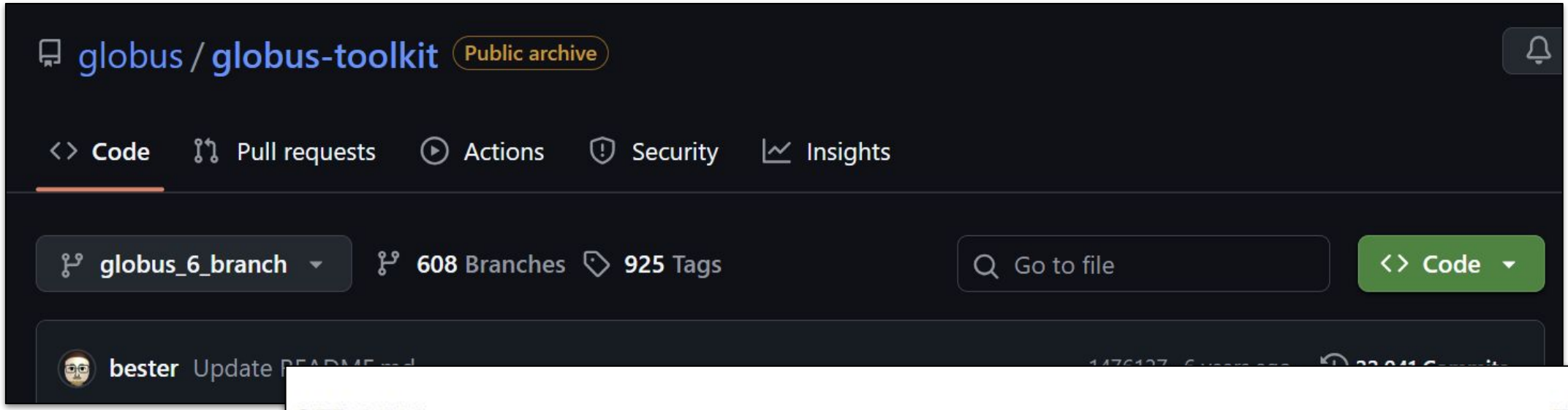
*"Sustainable software means that an existing product remains **viable in the future** such that it makes sense to continue **using, adapting, and expanding** its capabilities..."*

- *Gupta and Miller, Better Scientific Software, https://bssw.io/items/what-is-software-sustainability*

Technical: Address evolving user needs while minimizing the effort required to scale (growing user numbers, diverse use cases, heterogeneous hardware/software environments …)

Economic: Balance the competing demands of day-to-day operations, long-term maintenance, and continued innovation within available resources.

Open-source software and standardization



GFD-I.053

Category: Informational

Open Grid Services Architecture Working Group

Editors:

H. Kishimoto, Fujitsu

J. Treadwell, Hewlett-Packard

September 16, 2005

Defining the Grid: A Roadmap for OGSA™ Standards
Version 1.0

c|net

BEST PRODUCTS | REVIEWS | NEWS | VIDEO | HOW TO | SMART HOME | CARS | DEALS | DOWNLOAD

TECH INDUSTRY

Grid gurus launch a start-up

If you want to invest in a grid project, hiring Univa would be like paying Tim Berners-Lee to set up your home page.

BY **STEPHEN SHANKLAND** / DECEMBER 13, 2004 7:46 AM PST

The researchers who spawned the idea of grid computing on Monday launched a company to commercialize what so far has been a very academic software proje sharing computing resources.

The Chicago-based company, called Univa, is building its business on the Toolkit, grid software that serves as an important foundation to dozens of supercomputing projects.

As earlier reported, the company will sell support and services for those w to integrate Globus with their own products or computing operations, said Miller, chief operating officer of the new company.

InfoWorld FROM IDG

INSIDER Sign In | Register

Home > Technology Industry

Globus Consortium grid pact forged by IBM, Sun, Intel

By Paul Krill

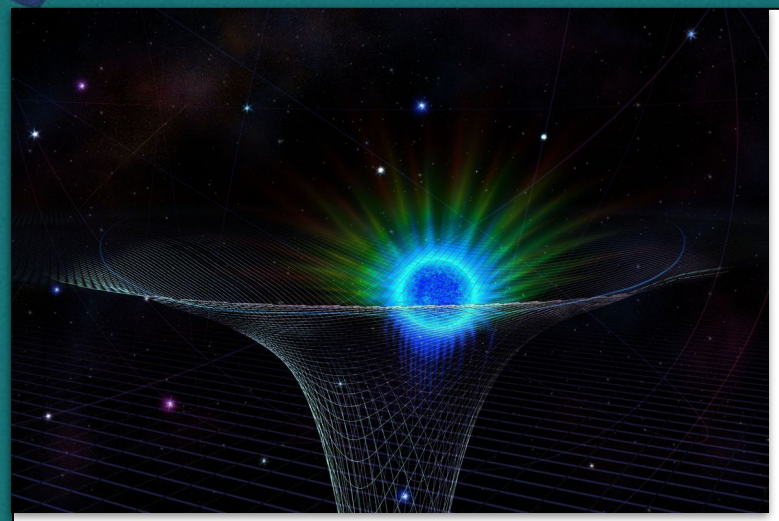
Editor at Large, InfoWorld | JAN 24, 2005

A new industry organization, the Globus Consortium, is being formed today to promote grid computing in commercial enterprises. Featured participants include Hewlett-Packard, IBM, Intel, Sun Microsystems and Univa. Intel's participation was revealed just this morning.

MORE LIKE THIS

Grid pioneers launch company

HP, IBM, Sun boost enterprise grids



!=

“Tell me again how you’re

[curing cancer, eradicating famine, ...]”

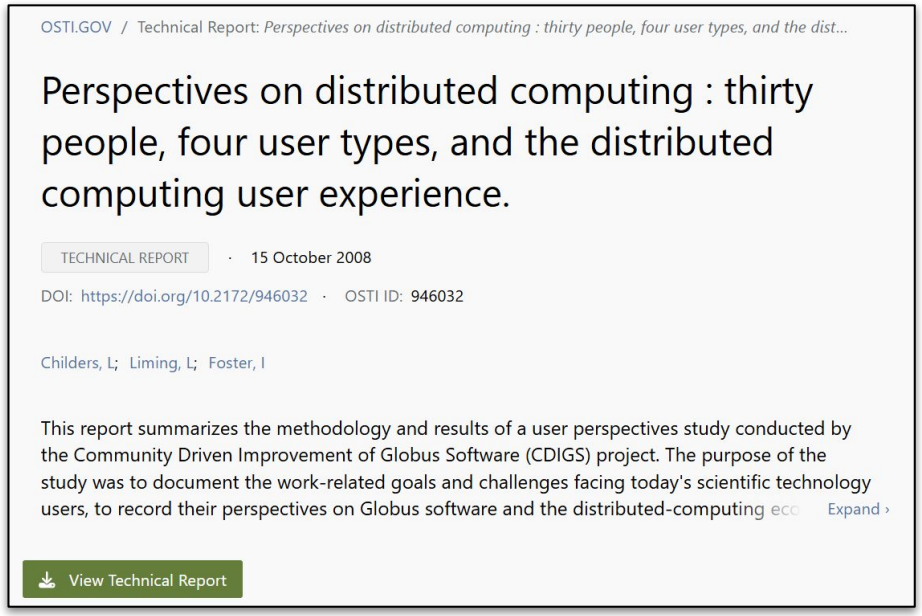
Understanding our “market”

Interviews with 30 community members:

- HPC scientist, HPC domain-specific developer, general HPC provider, general HPC developer

Findings and recommendations

- Focused on developers not users
  → design for all stakeholders
- Focused on siloed components
  → holistic product development
- Ease of use is paramount
  → “solutions not toolkits”



Childers, Liming, Foster. https://doi.org/10.2172/946032

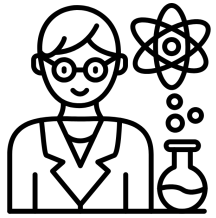
What does sustainable software really mean?

Technical: Address evolving user needs while minimizing the effort required to scale (growing user numbers, diverse use cases, heterogeneous hardware/software environments …)

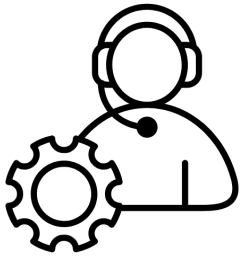
Economic: Balance the competing demands of day-to-day operations, long-term maintenance, and continued innovation within available resources.

Foster, Ian; Vas Vasiliadis; Tuecke, Steven (2013). Software as a Service as a path to software sustainability. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.791604.v1

Technical: Researchers need “solutions not toolkits”



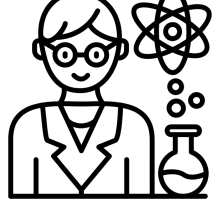
How do we serve researchers



How do we serve administrators

How do we serve developers

Globus Online: Software-as-a-Service for researchers



SaaS means:

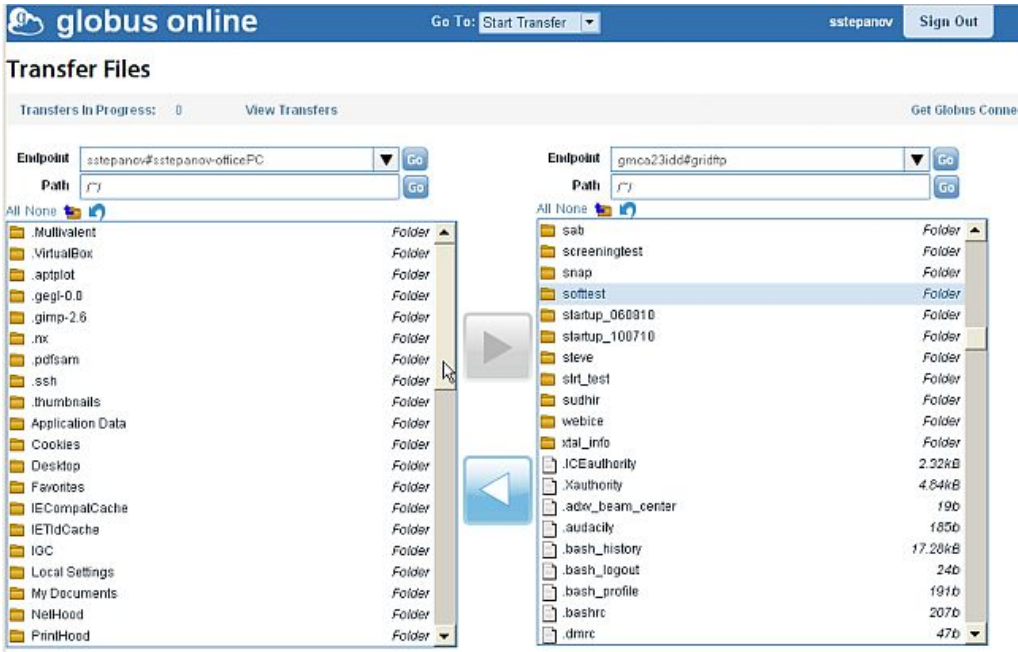
1. Application owned, delivered, managed by provider
2. Single code base supports many users at once
3. Web-native application using cloud technologies

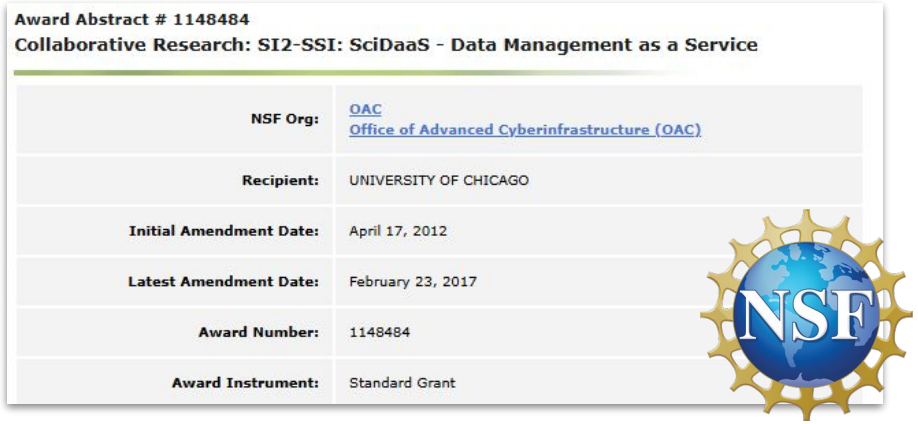
Has many potential advantages

- Leverage Web 2.0 for extreme ease of use
- Substantial economies of scale
- Expert operations and support
- Rapid software update

As well as challenges

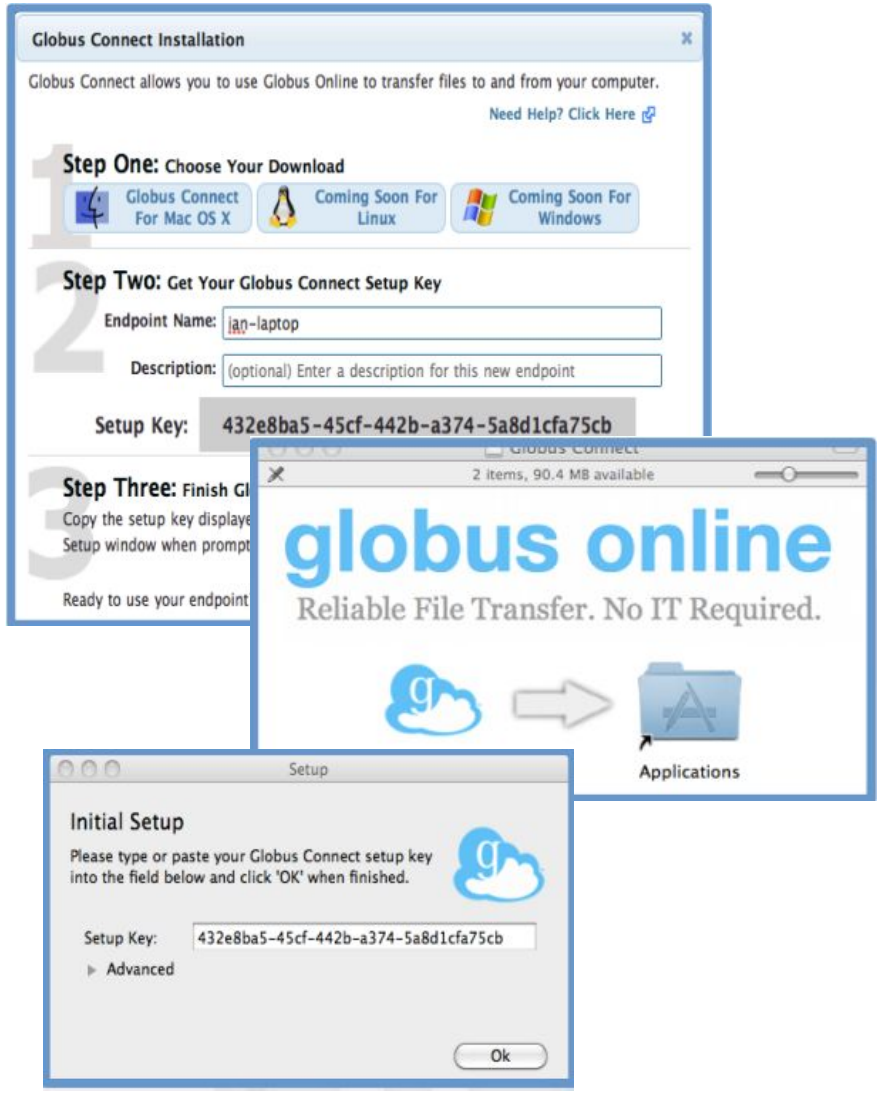
- Paying for it; availability, security and privacy …

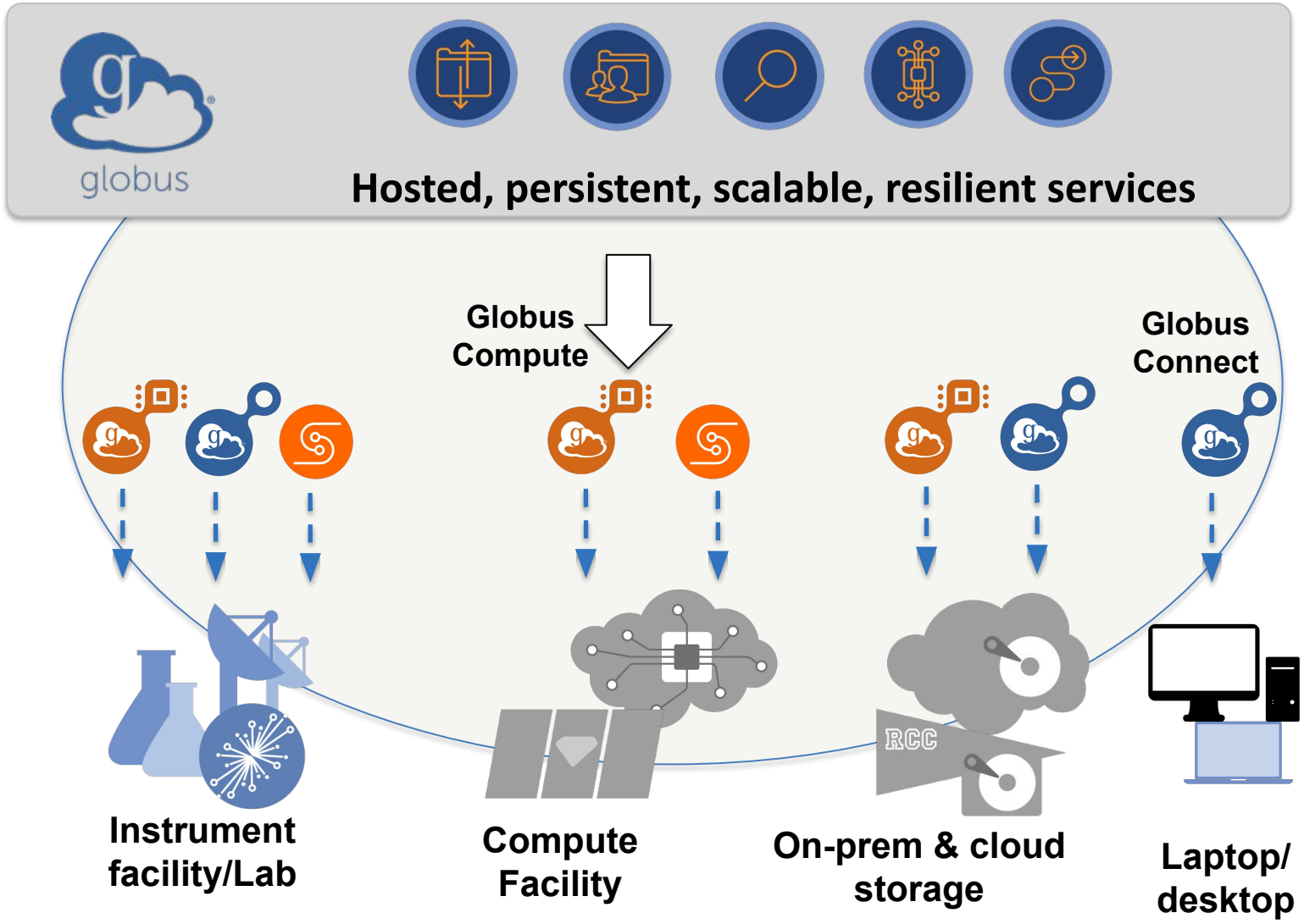




Foster, Globus World Keynote, 2011

A seamless and simple hybrid model





Openness: Platform, APIs, and Code

Open APIs - enable others to build upon and integrate in their applications

Plugin APIs - enable customization to different environments and partnership with other providers

Licensing: Agent and SDK code is open source (or available to subscribers); hosted service code remains UChicago IP

https://www.globus.org/legal/source-license

What does sustainable software really mean?

Technical: Address evolving user needs while minimizing the effort required to scale (growing user numbers, diverse use cases, heterogeneous hardware/software environments …)

Economic: Balance the competing demands of day-to-day operations, long-term maintenance, and continued innovation within available resources.

Foster, Ian; Vas Vasiliadis; Tuecke, Steven (2013). Software as a Service as a path to software sustainability. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.791604.v1

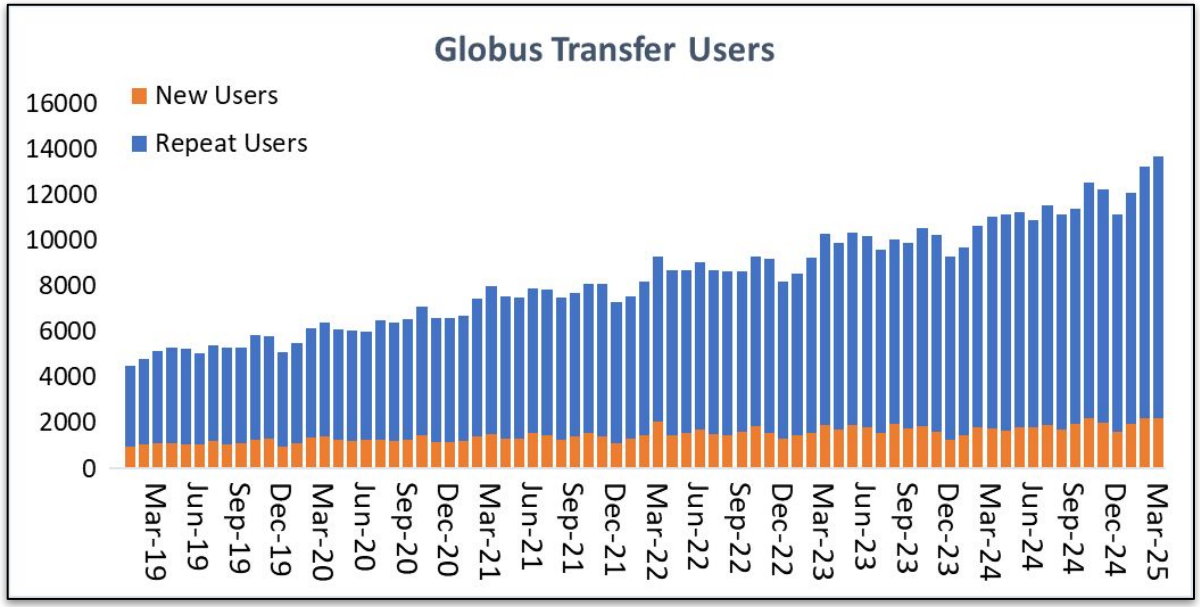
Economic: Cloud model changes the calculus

Delivering a highly-available service requires:

- Replication, production support teams, monitoring, automated deployment/recovery, …

→ Build on cloud services where possible

→ Implement reasonable limits to protect the service



3,334,363,989,143 MB TRANSFERRED

powered by aws

Freemium – Web 2.0 business models

Freemium is a business model that offers some features of a product or service for free, while charging for premium or advanced features

Who pays? And how much? And for what features?

- User, Institutional; Deployment, Usage, Fixed/unlimited

Finding our way with subscription models

End User Plans

- **Basic: Free**
  - File transfer and synchronization to/from servers
  - Server endpoints with Globus Connect Multi-User
    - Can host shared endpoints for Plus subscribers
  - Personal endpoints with Globus Connect
  - Access to shared endpoints created by others
- **Plus: $7/month (or $70/year)**
  - Create and manage shared endpoints (from any sharable or personal endpoint)
  - Peer-to-peer (Globus Connect to Globus Connect)
  - Support for web and command line interfaces

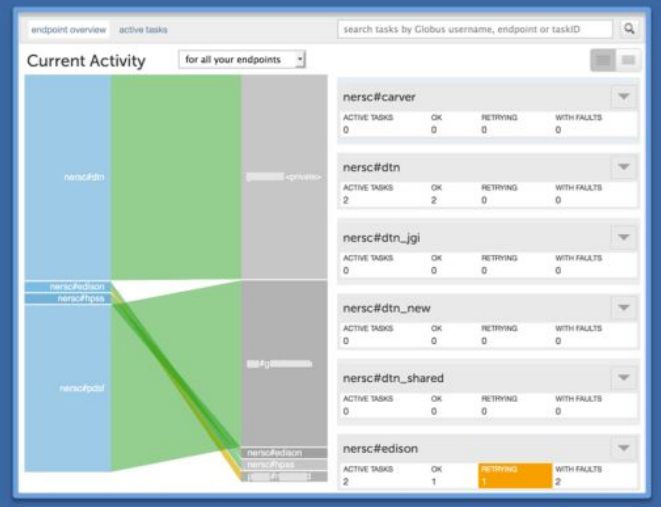
Award Abstract # 1240726
EAGER-ESOCS: Exploring Sustainability Options for Cyberinfrastructure Se

NSF Org:	OAC Office of Advanced Cyberinfrastructure (OAC)
Recipient:	UNIVERSITY OF CHICAGO
Initial Amendment Date:	June 13, 2012
Latest Amendment Date:	June 13, 2012
Award Number:	1240726
Award Instrument:	Standard Grant
Program Manager:	Rudolf Eigenmann OAC Office of Advanced Cyberinfrastructure (OAC) CSE Directorate for Computer and Information Science and Engineering
Start Date:	July 1, 2012

Institutional subscriptions

Subscriptions

- **Globus Provider Plan**
  - Managed endpoints
  - Host shared endpoints
  - Management console
  - Data publication
  - Amazon S3 endpoints
  - Usage reporting
  - Priority support
  - Application integration
- **Branded Web Site**
- **Alternate Identity Provider (InCommon is standard)**
- **Mass Storage System optimization**



globus.org/provider-plans

39

2015

globus | I Want To... | Pricing | Resources | Support | About

Features (click ⓘ for description)	Basic	Starter	Standard	High Assurance ⓘ	HIPAA BAA ⓘ
File transfer ⓘ	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
Managed endpoints ⓘ	None	1	Unlimited	Unlimited	Unlimited
Management console ⓘ	–	✓	✓	✓	✓
Usage reports ⓘ	–	✓	✓	✓	✓
Support for Globus Connect, Web, CLI ⓘ	–	✓	✓	✓	✓
File sharing ⓘ	–	✓	✓	✓	✓
Globus Plus users ⓘ	–	–	✓	✓	✓
Data publication ⓘ	–	–	✓	✓	✓
Application integration support ⓘ	–	–	✓	✓	✓
HTTPS support ⓘ	–	–	✓	✓	✓
Session/Device Isolation ⓘ	–	–	–	✓	✓
Additional authentication assurance ⓘ	–	–	–	✓	✓
Comprehensive audit logging ⓘ	–	–	–	✓	✓
Business Associate Agreement ⓘ	–	–	–	–	✓
Support service level ⓘ	–	Monday-Friday, 9am-5pm Central; 1-business day response			
Named support contacts ⓘ	–	1	5		
Pricing	Free	Contact us for subscription pricing or request details on pricing for Globus with protected data support			

2019

Transparent and equitable subscription pricing

Pricing based on institutional research budget

- Proxy for scope/scale of research
- Transparency and equity across diverse institutions

Add-ons: Storage connectors

Partnerships: Internet2 Net+

How are we doing so far?

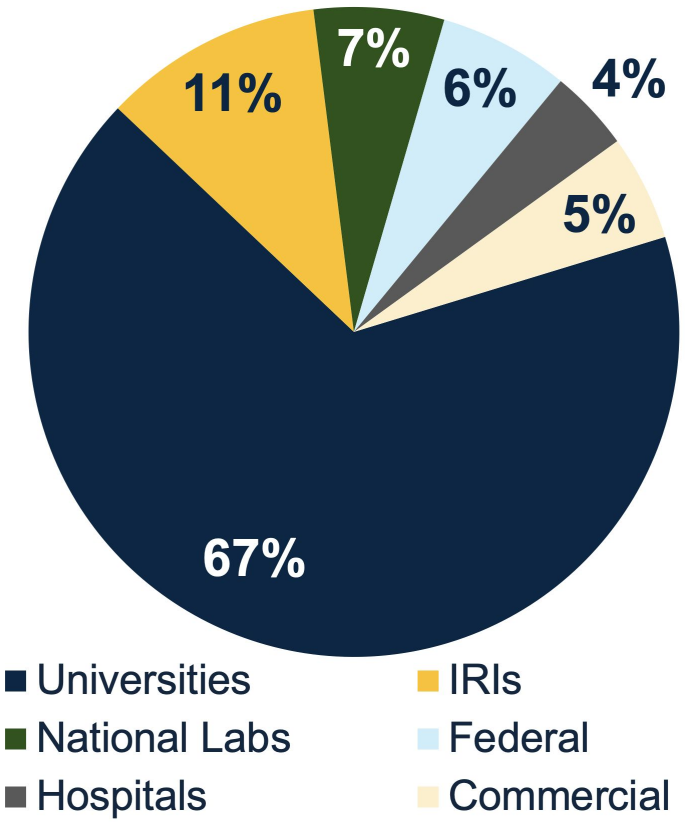
250+ Subscribers

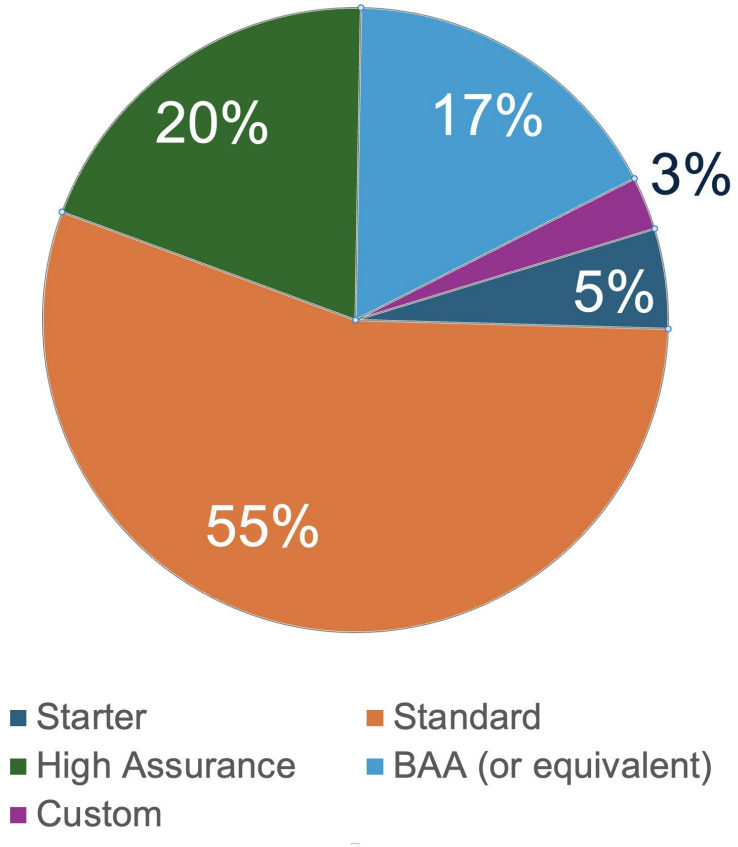
~70% of R1 universities

~60% of active endpoints are associated with subscribers

~80% national laboratories use Globus

National deployments in Australia, New Zealand, Canada





Globus department at the University of Chicago

Established in 2017 as a translational unit, with an externally funded financial model

Globus Labs

Research mission and grant funded

Globus Platform

Delivers services for research community and funded by subscriptions

Globus Labs

Best Paper

TaPS: A Performance Evaluation Suite for Ta

J. Gregory Pauloski, Valerie Hayot-Sasson, Sicheng Zhou, Ian Foster and Kyle Chard

2012 R&D 100 50th ANNIVER

2018 R&D 100 WINNER

IEEE COMPUTER SOCIETY

2023 Best Paper Award

IEEE Internet Computing

ACM Gordon Bell Prize

Recognizing Outstanding Achievement in High Performance Computing

DOE CSGF

AWARD RECIPIENT

Ian T Foster

ACM-IEEE CS Ken Kennedy Award (2022)

ACM Gordon Bell Prize (2022)

ACM Fellows (2009)

Ian Foster Recognized with ACM-IEEE CS Kennedy Award

HPCwire 2024

Google Scholar

Globus Labs

University of Chicago

Verified email at uchicago.edu - Homepage

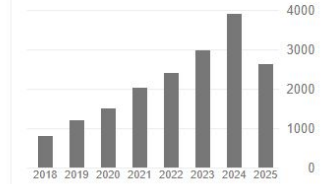
distributed computing language models AI4Science

FOLLOW

TITLE	CITED BY	YEAR
RADAR-Radio Afterglow Detection and AI-driven Response: A Federated Framework for Gravitational Wave Event Follow-Up P Patel, A Corsi, EA Huerta, K Merfeld, V Tiki, Z Li, T Bicer, K Chard, ... arXiv preprint arXiv:2507.14827		2025
Flight: A FaaS-based framework for complex and hierarchical federated learning N Hudson, V Hayot-Sasson, Y Babuji, M Baughman, JG Pauloski, ... Future Generation Computer Systems, 107998	2	2025
Will It Blend? Mixing Numerical and Machine-Learned Physics Quantities for Accurate on-the-Fly Surrogate Modeling M Tynes, K Chard, I Foster, L Ward International Conference on Computational Science, 270-284		2025
Bayesian Optimization Hackathon for Chemistry and Materials S Baird, M Ansari, Z Afzal, Q Ai, A Al-Feghali, M Alain, M Altamirano, ...		2025
PBench: Workload Synthesizer with Real Statistics for Cloud Analytics Benchmarking Y Zhou, C Liu, B Urgaonkar, Z Wang, M Mueller, C Zhang, S Zhang, ... arXiv preprint arXiv:2506.16379		2025
WRATH: Workload Resilience Across Task Hierarchies in Task-based Parallel Programming Frameworks S Zhou, Z Li, V Hayot-Sasson, H Pan, M Gonthier, JG Pauloski, R Chard, ...	1	2025

Cited by VIEW ALL

	All	Since 2020
Citations	19322	15573
h-index	62	54
i10-index	315	276



Co-authors

Ian T. Foster, University of Chicago and Argon...

Kyle Chard, University of Chicago and Argon...

Logan Ward, NVIDIA (formally Argonne Nation...

Ben Blaiszik, University of Chicago and Argon...

NSF REU Site: AI and Science Summer Lab: advancing AI-enabled scientific discovery across the physical and biological sciences

Sustainability must be tailored to product: Parsl

Parsl is a Python library designed to support the definition and execution of parallel programs across diverse computing resources

Open-source software with 100+ contributors and millions of software downloads

Open governance model and fiscally sponsored NumFocus project

Parsl

NUMFOCUS
OPEN CODE = BETTER SCIENCE

Thank you, Globus team

Ian Foster

Steve Tuecke

Rachana Ananthakrishnan

Vas Vasiliadis

Kristin Kolesiak

Thank you, sponsors!

U.S. DEPARTMENT OF
ENERGY

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

THE UNIVERSITY OF
CHICAGO

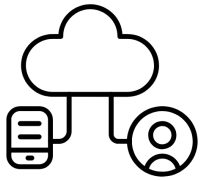
Argonne
NATIONAL LABORATORY

powered by
amazon
web services

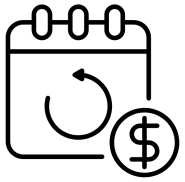
Thank you, subscribers!

Lessons learned operating Globus

A carefully designed research IT platform, delivered and operated at enterprise grade, can advance discovery by reducing friction



A hybrid cloud deployment model democratizes access, improves user experience, and reduces costs



A freemium/subscription-based support model can enable sustainable operation of such services

University-based non-profit with research and platform model enables delivery of innovative capabilities at production scale