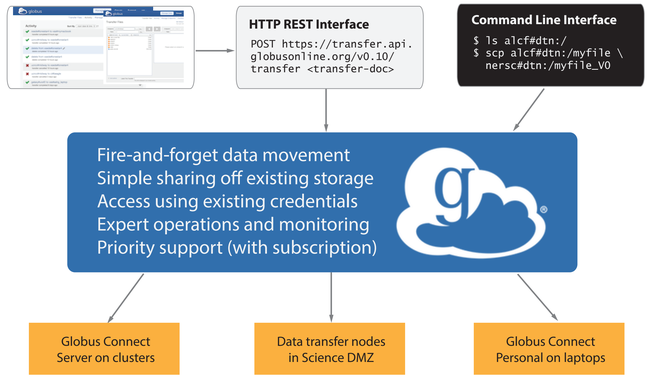
# Globus: Research Data Management-as-a-Service

globus.org

**Objective**: To provide frictionless access to sophisticated cyberinfrastructure for managing large datasets throughout their lifecycle, at a modest cost to ensure sustainability.

## Background

Globus started as an experiment to explore the feasibility of using SaaS approaches to radically simplify research data transfer. Globus is now used by dozens of research institutions and is the recommended file transfer service for UChicago RCC, NCSA, XSEDE, NERSC, and many other cyberinfrastructure providers nationwide. It implements methods for managing the secure transfer, sharing, and publication of large data sets.




*A user view of Globus*

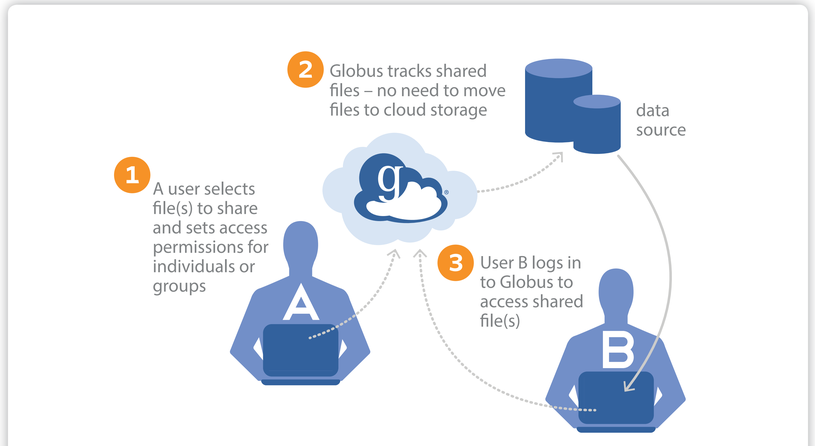
## Project Impact

- Deliver advanced capabilities for moving, sharing, annotating, and publishing reasearch datasets.
- Accelerate discovery and expand scope of inquiry
  - Enable PIs to spend more time researching
  - Significantly increase data processing rates
- Expand usage of advanced computation across entire US research and education system.
- Expand use of Software-as-a-Service approaches across all areas of research.

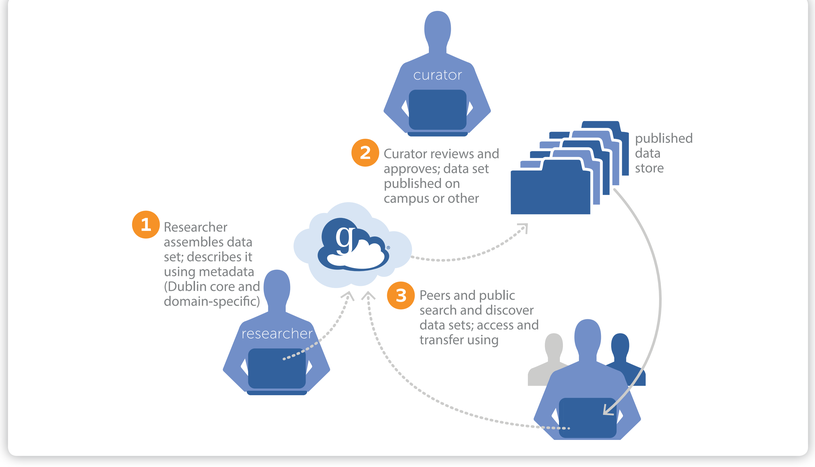
***CONTRIBUTORS:*** B. Allen, R. Ananthakrishnan, R. Aydt, J. Bester, K. Chard, V. Cuplinskas, P. Dave, I. Foster, R. Kettimuthu, J. Kordas, L. Lacinski, M. Lidman, L. Liming, M. Link, S. Martin, B. McCollam, C. Pickett, D. Powers, J. Pruyne, G. Rohder, S. Rosen, D. Shifflett, T. Sutton, S. Tuecke, V. Vasiliadis, J. Williams

## Transfer files reliably, securely, and fast.


## Share files directly from existing storage.



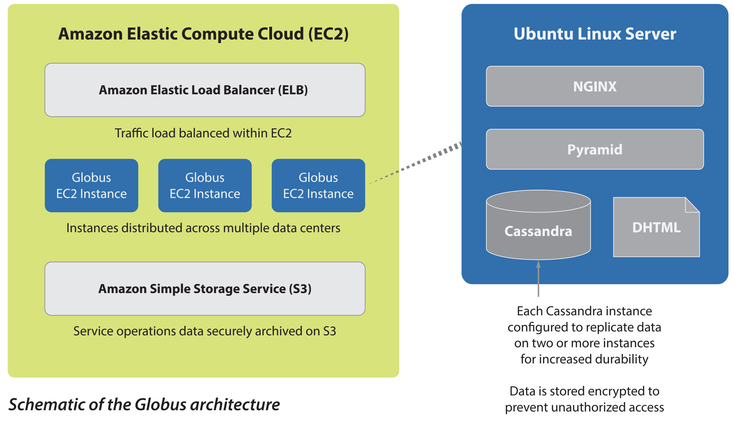

## Publish files with metadata for discovery.




## Globus at the UChicago Research Computing Center

- Transfer/sharing available to all with an RCC account
- Endpoint name: **ucrcc#midway**
- Sign up: **globus.org/signup**
- Enable laptop: **globus.org/globus-connect-personal**

## Globus operations hosted on Amazon Web Services




***Schematic of the Globus architecture***

> *"With Globus, I transferred 1.7T data in 9 hours without needing to check in or babysit the transfer. I could never do this type of transfer with scp or rsync because it would take too long and possibly crash in the middle."*
>
> - Yun Luo, University of Chicago

> *"Globus moved 100,000 files totaling 98 TB from Argonne to NERSC and Oak Ridge in a couple of days, with no human involvement."*
>
> - Researcher at Argonne National Laboratory

## Selected Publications, Presentations, Awards

- Radical Simplification of Data Movement via SaaS, Comm. ACM, January 2012
- Tutorials, BOFs, papers presented at SC12 through SC15
- Multiple workshops, papers presented at Bio IT World, 2012 through 2015
- Tutorials, papers presented at XSEDE 2012 through 2015
- 2012 R&D 100 Award, November 2012

## Globus by the Numbers

**10,000** Active endpoints

**20 billion** Files transferred

**30,000** Registered users

**110** Petabytes moved

2012 R&D 100 50th ANNIVERSARY

NSF

THE UNIVERSITY OF CHICAGO