

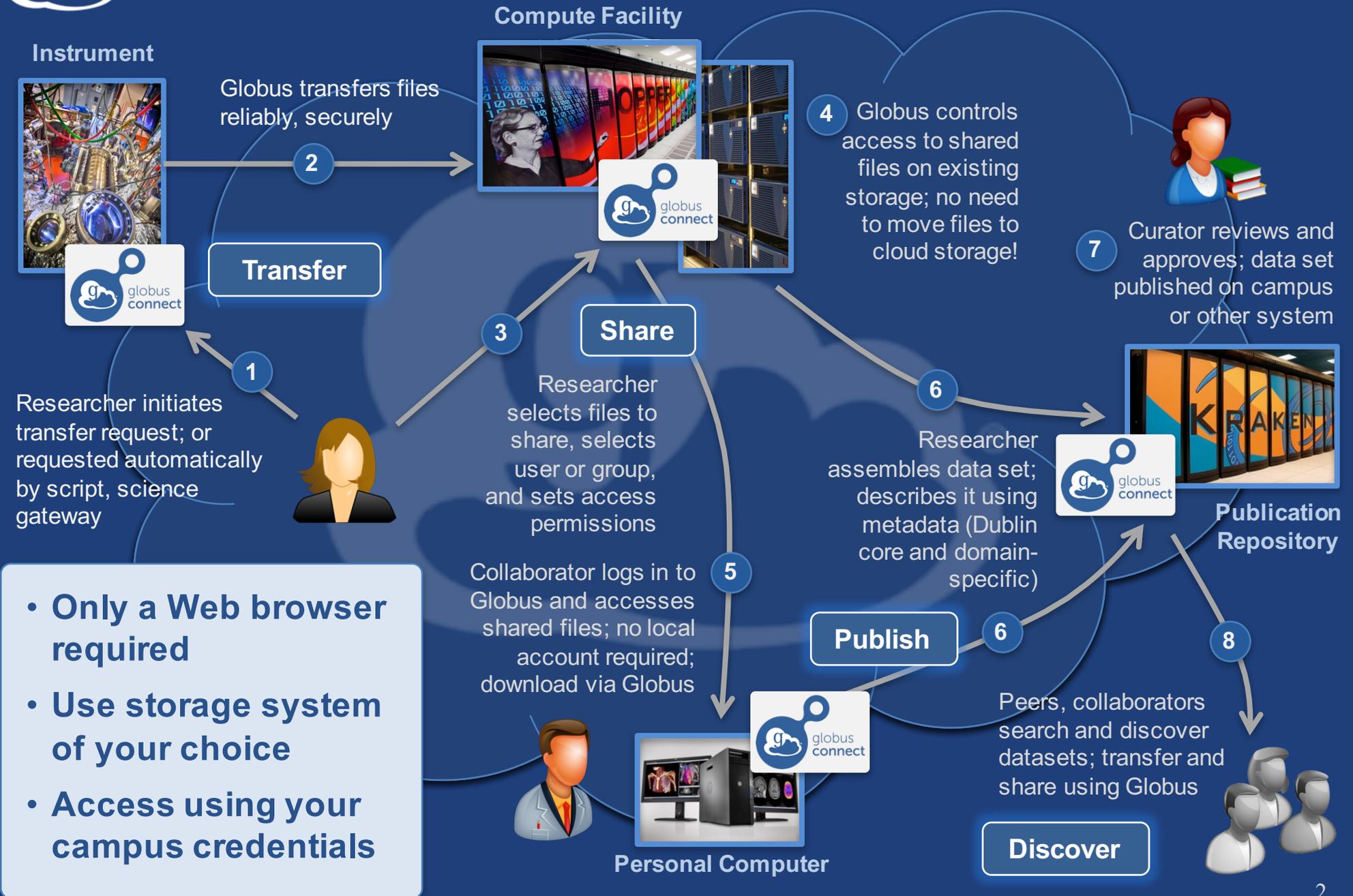
Research Data Management with Globus

Steve Tuecke
tuecke@globus.org





Globus and the research data lifecycle





Who uses it and for what?

- **Researchers**
 - Data sharing, transfer, archiving, upload, download, ...
- **Research data portal operators**
 - Data upload, download
- **Operators of experimental facilities**
 - Data delivery, data replication
- **Campus computing administrators**
 - Research acceleration, resource management



Globus by the numbers

4

major services

160 PB
transferred

25 billion
files processed

38,000
registered users

13

national labs
use Globus

2,500
active endpoints

450+
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



Petrel



Lab Data Service Goals

A uniform data fabric across the lab

... with seamless access to large data...

**... for use in computation, collaboration
and distribution ...**

**... that is project focused and self
managed ...**

... and is described and discoverable



Petrel: data store for project



1 Project Request



2 Provision space,
user sets access
control



PETREL
Data Management and Sharing Pilot

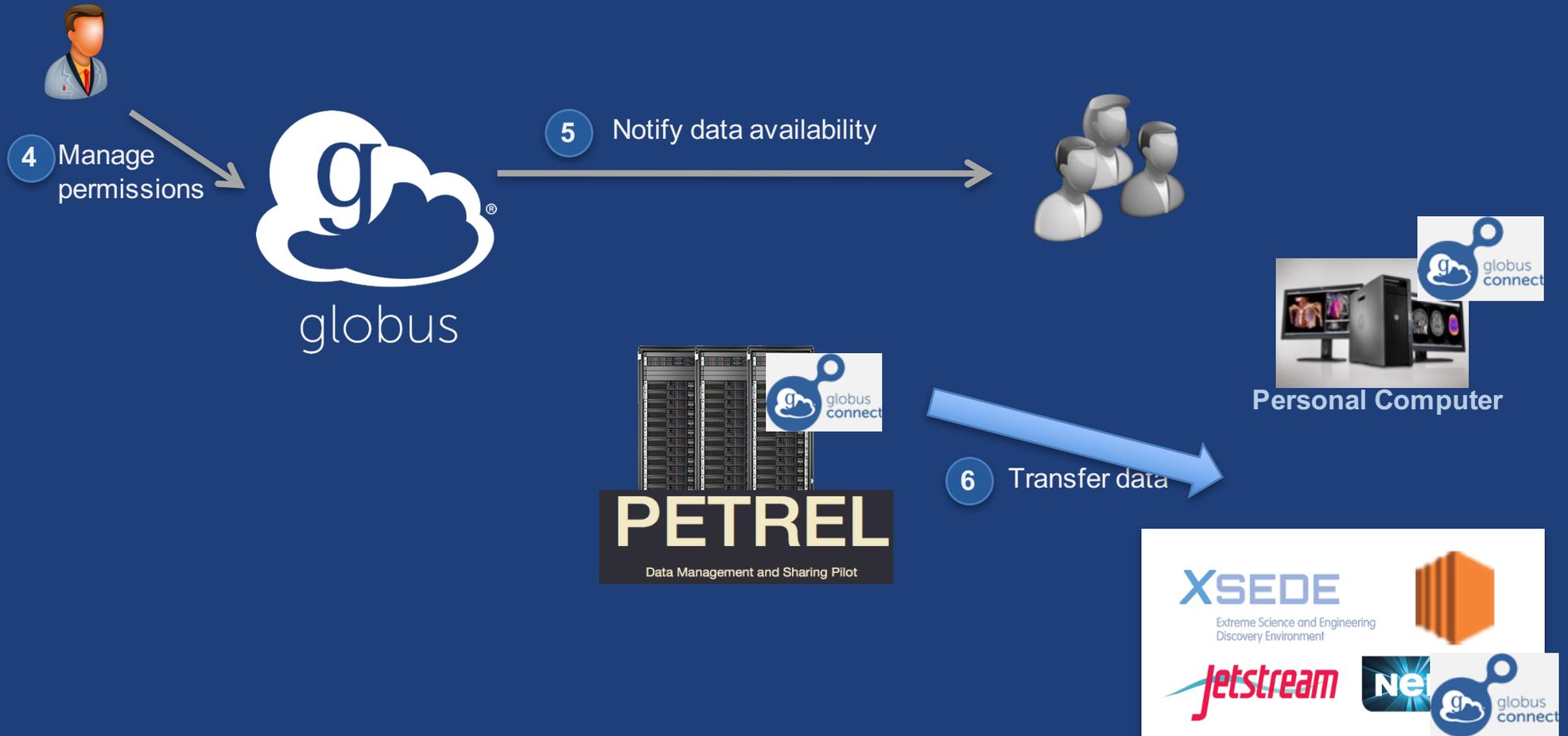
3

Transfer data





Petrel: share with collaborators





Using Petrel

- **Documentation and request for allocation**
 - petrel.alcf.anl.gov
- **Project based allocation**
 - 100 TB
 - Self-managed by PI
 - Low administrator overheads
- **Current Usage**
 - 17 projects
 - Largest distribution group
 - 88 collaborators
 - Largest project store
 - 205 TB



Demo



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



powered by
amazon
web services