



Computation Institute

Research Data Management and Analysis using Globus Platform

Ravi K Madduri, University of Chicago and Argonne National Laboratory

madduri@uchicago.edu

@madduri



THE UNIVERSITY OF
CHICAGO



globus.org/genomics

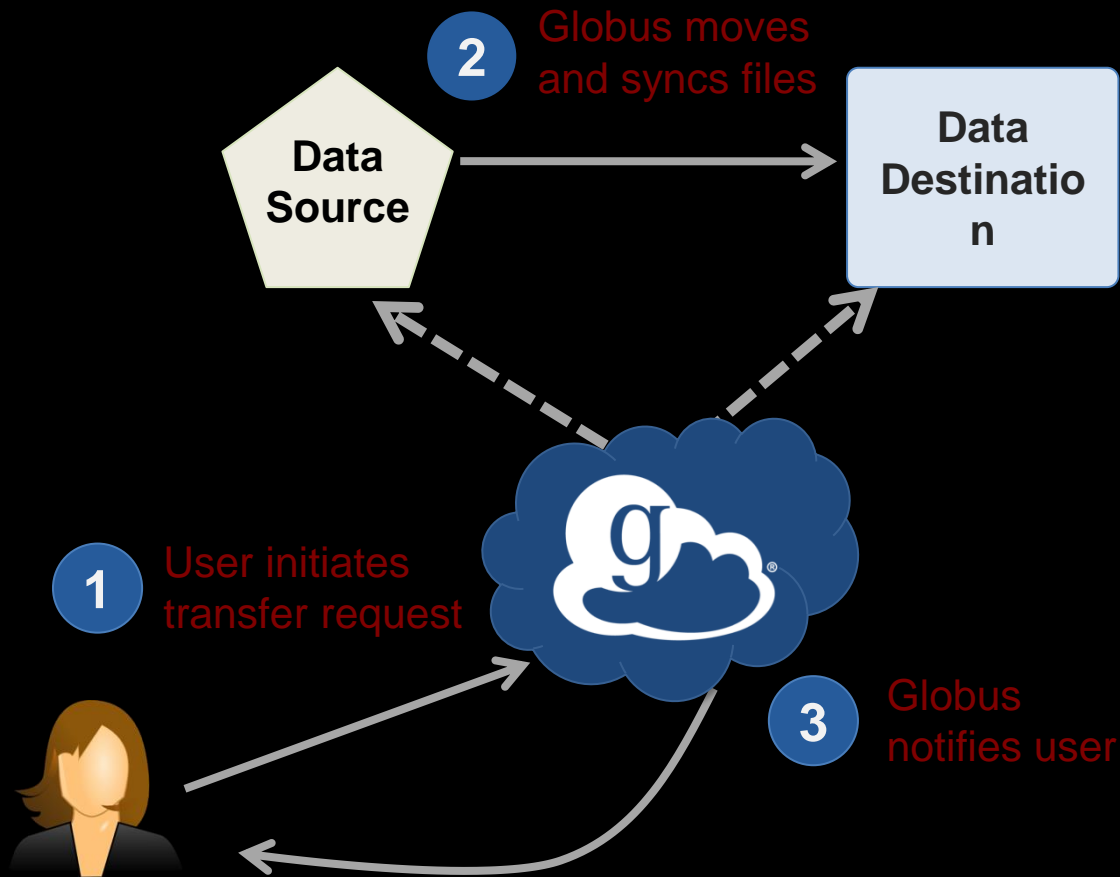


Our vision for a 21st century discovery infrastructure

Provide **more** capability for
more people at **lower cost** by
delivering “**Science as a Service**”

www.globus.org

Globus: Fast, reliable data transfer



Globus: Federated identity

InCommon.




CILogon



Shibboleth.



 THE UNIVERSITY OF CHICAGO

Sign In

You are logging in to: classes.uchicago.edu
A Web-Single-Signon protected site

CNetID:

☐ Hospital Employee?


Password:


[Forgot your password?](#)

Signing in allows you to access multiple University of Chicago web applications while entering your CNetID and password only once. To end your session, simply close your browser.

Questions? Contact the IT Services Service Desk by phone at 2-5800 (773-702-5800), via email at itservices@uchicago.edu, or get walk-in help at the TECHB@R on the first floor of Regenstein Library during reference desk hours <http://hours.lib.uchicago.edu/>.

Alumni account holders may contact alumni-support@uchicago.edu or call 1-877-292-3945 between 9 AM and 3 PM CST with any questions.

Authentication powered by Shibboleth™ 

 Cornell University

CUWebLogin

ApplicantID:

Password:

[What is this?](#)
[I forgot my password!](#)
[I don't have a NetID, now what?](#)

MyProxy

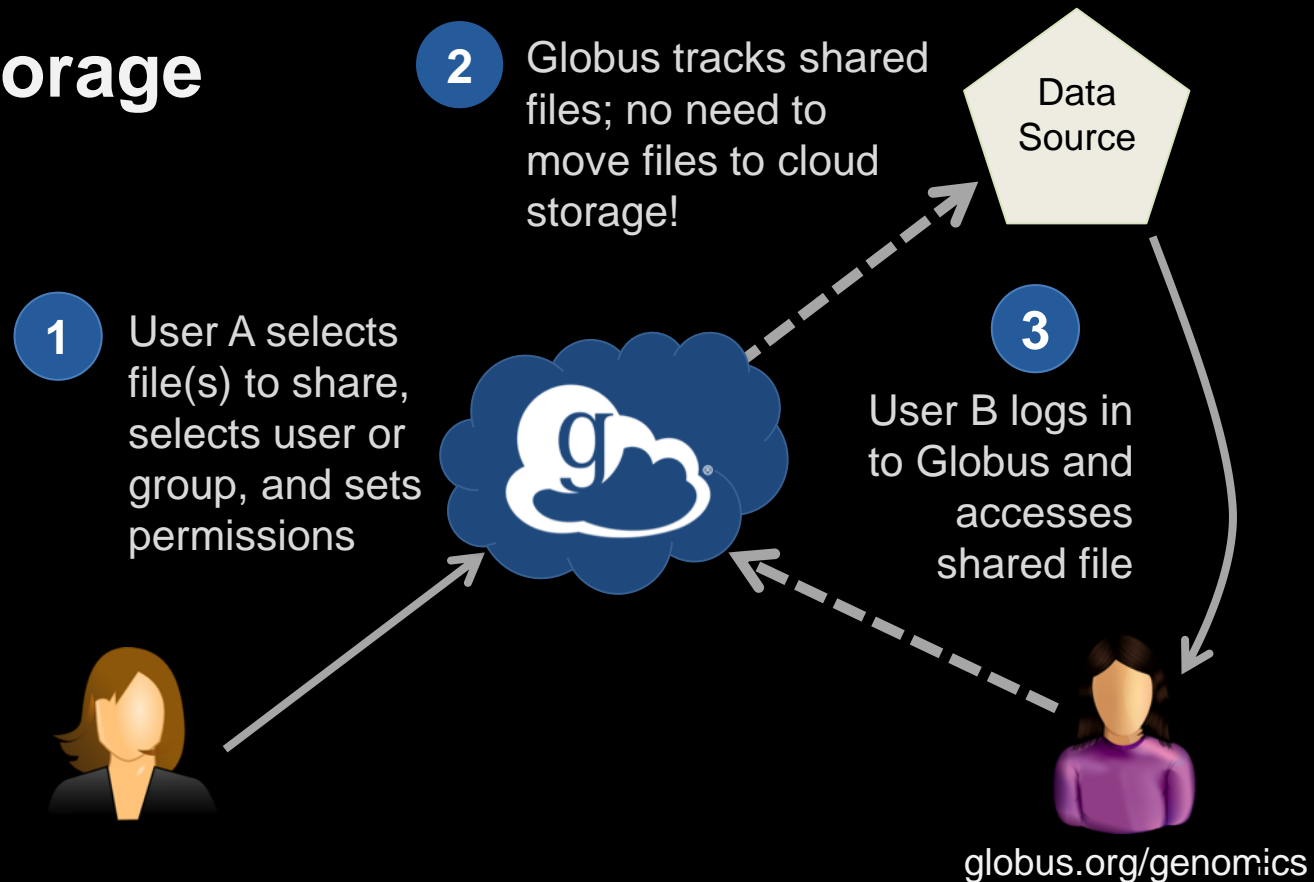
NEWT



globus.org/genomics

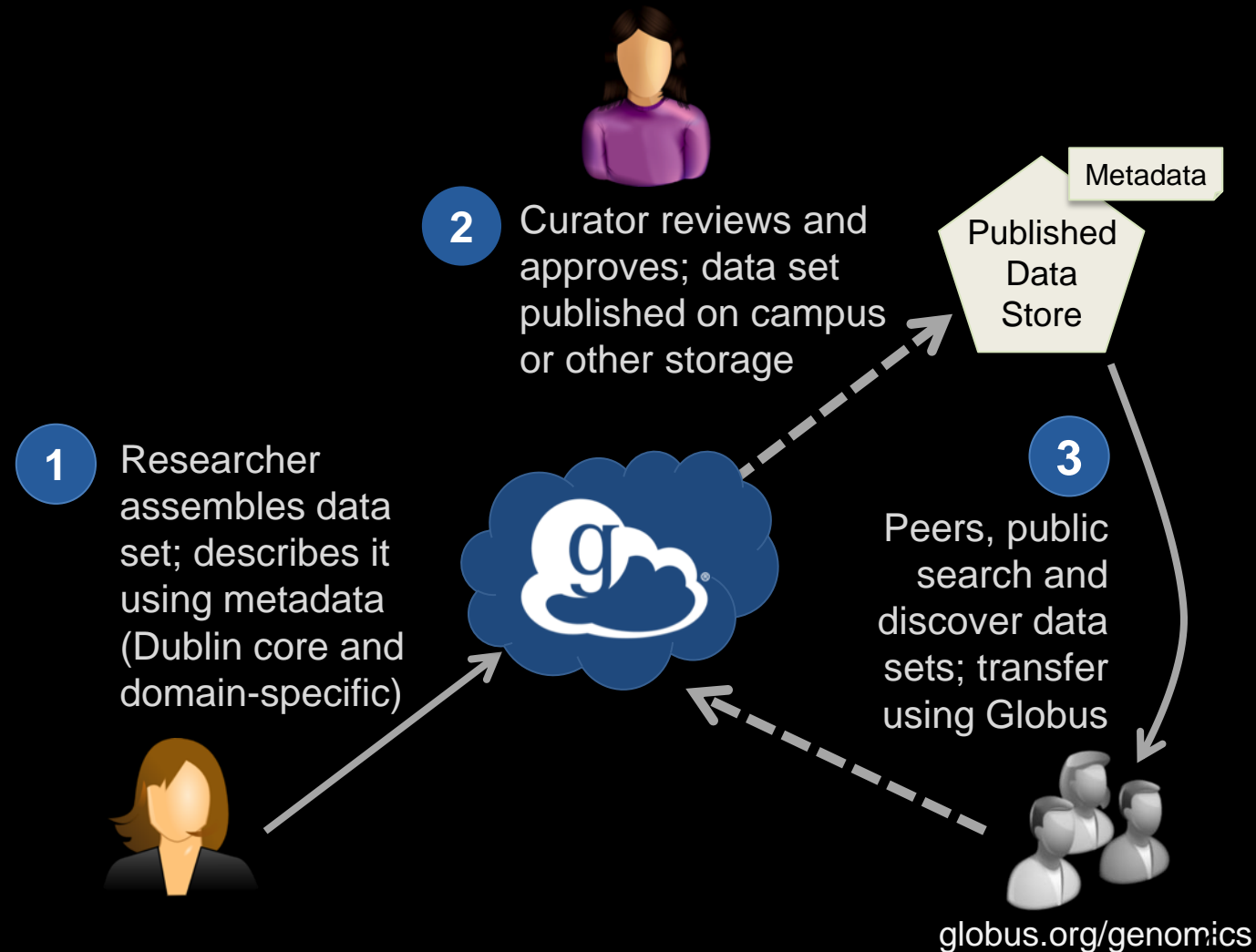
Simple, secure *sharing* off existing storage systems

- Easily share large data with any user or group
- No cloud storage required

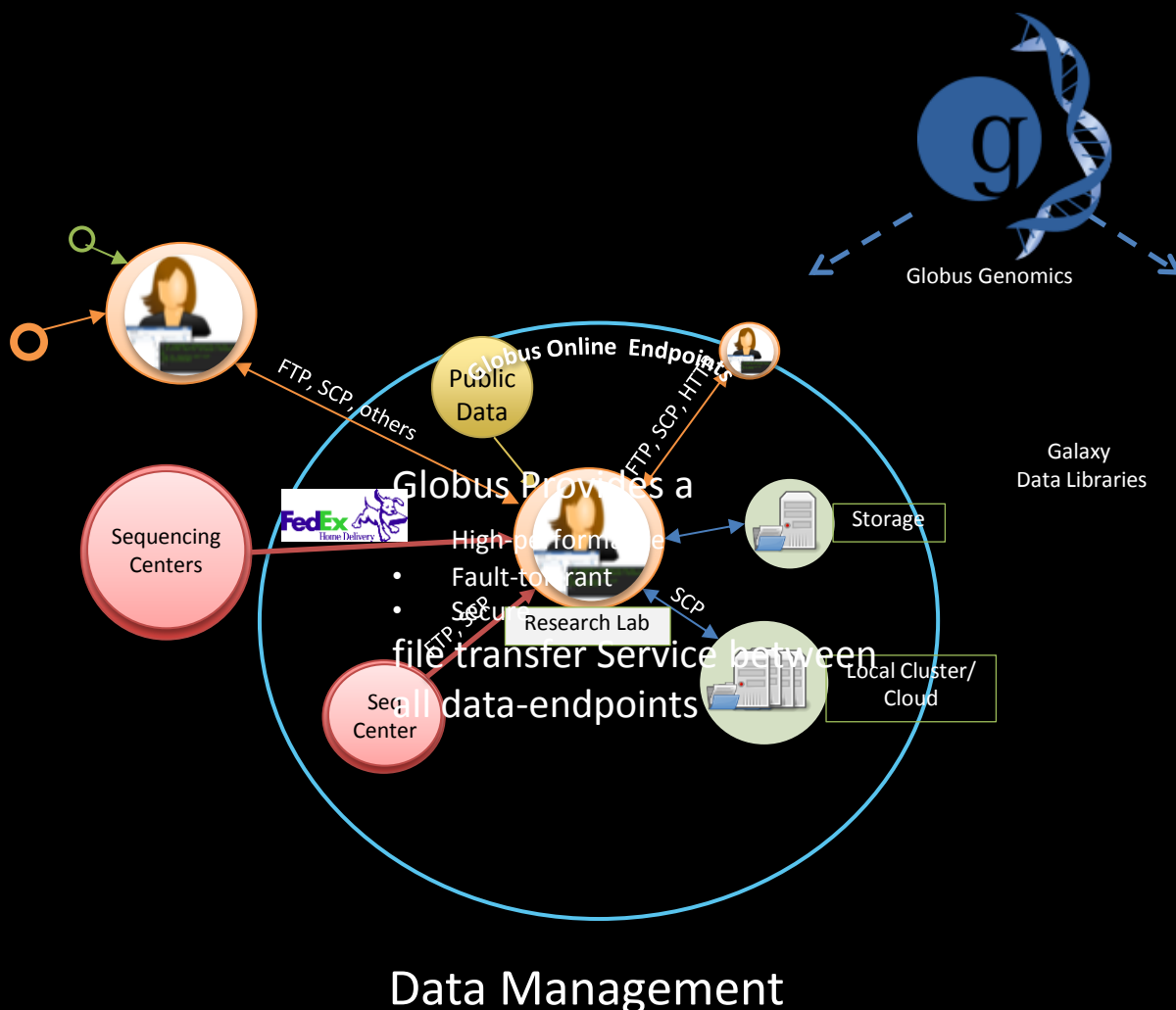


Curated *publication* of data, with relevant metadata for *discovery*

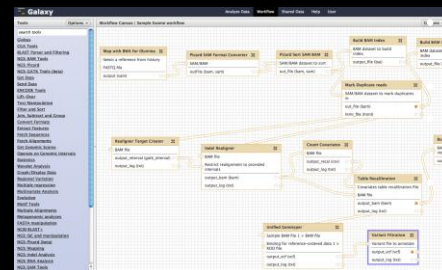
- Identify
- Describe
- Curate
- Verify
- Access
- Preserve



Globus Genomics



Galaxy Based Workflow Management System



Globus Integrated within Galaxy Web-based UI Drag-Drop workflow creations Easily modify Workflows with new tools



Analytical tools are automatically run on the scalable compute resources when possible

Globus Genomics on Amazon EC2


Data Analysis

It's about the user experience...

flickr ...for your photos

Gmail ...for your e-mail
by Google

NETFLIX ...for your entertainment

 globus ...for your research data

Globus Platform-as-a-Service



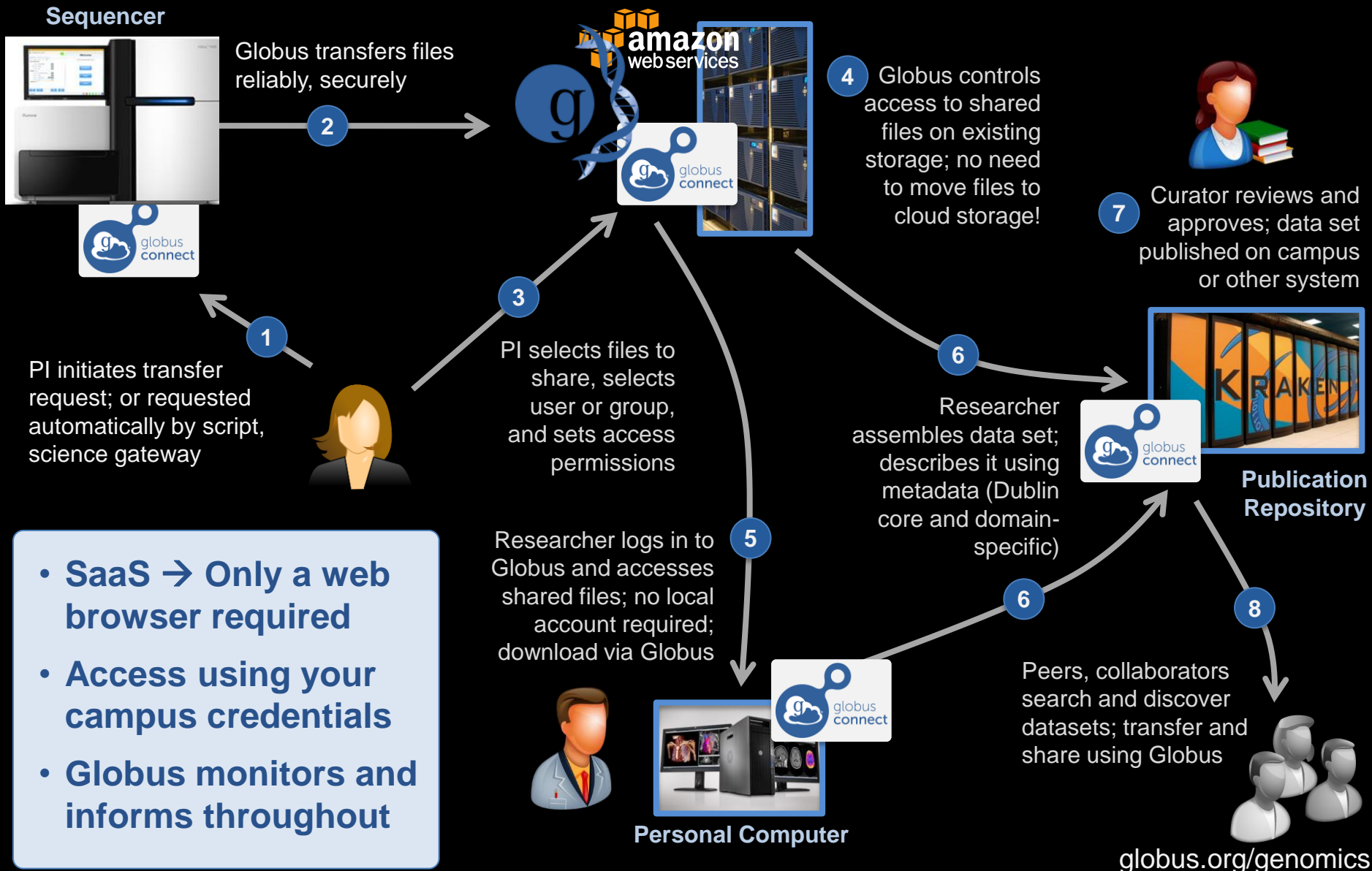


How would it look if GeneLab
were powered by Globus
Services?

What are the benefits? What are
some of the open questions?

Managing the research data lifecycle with Globus

Globus Genomics



Some Examples



Services using Globus Platform

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Home About News Dev

The new Systems Biology Knowledgebase is a collaborative effort designed to accelerate community source access to new knowledge.

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

Reliable, high-performance, secure computing.

Earth System Grid

Home Data Account About Contact Us Logout

EuPathDB
Eukaryotic Pathogen Database Resources

Version 2.0
10 Sep 14

Gene ID: PF3D7_1133400 Gene Text Search: synth

About EuPathDB | Help | Login | Register | Contact Us

Home New Search My Strategies My Basket (0) Tools Data Summary Downloads Community

Data Summary

News and Tweets

- 10 September 2014
CryptoDB 7.0 News
- 10 September 2014
PlasmoDB 12.0 Released
- 10 September 2014
AmoebaDB 4.1 Released

Tweets

EuPathDB @eupathdb
Gut Microbiota Elicits a Protective Immune Response against Malaria Transmission @CellPress

Identify Genes by:

Expand All | Collapse All

- Text IDs, Organism
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern

Identify Other Data Types:

Expand All | Collapse All

- Isolates
- Genomic Sequences
- Genomic Segments (DNA Motif)
- SNPs
- ESTs
- ORFs

Tools:

BLAST
Identify Sequence Similarities
Results Analysis
Analyze Your Strategy Results
Sequence Retrieval
Retrieve Specific Sequences using IDs and coordinates
Pathogen Portal

EuPathDB Project

AmoebaDB CryptoDB GiardiaDB MicrosporidiaDB PiroplasmaDB PlasmoDB ToxoDB TrichDB TriTrypDB

org/genomics



Key Issues

- Lowering the barrier to entry and Democratizing research
- Reproducibility/Transparency/Accessibility
- Benefits of Crowd Sourcing – GalaxyZoo
- Sustainability – TCO
- Ease of Use (UX)
- API

Our work is supported by:



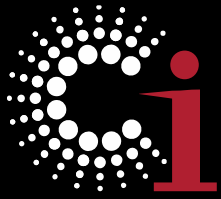
U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

Argonne
NATIONAL LABORATORY





Thank you!

@madduri