

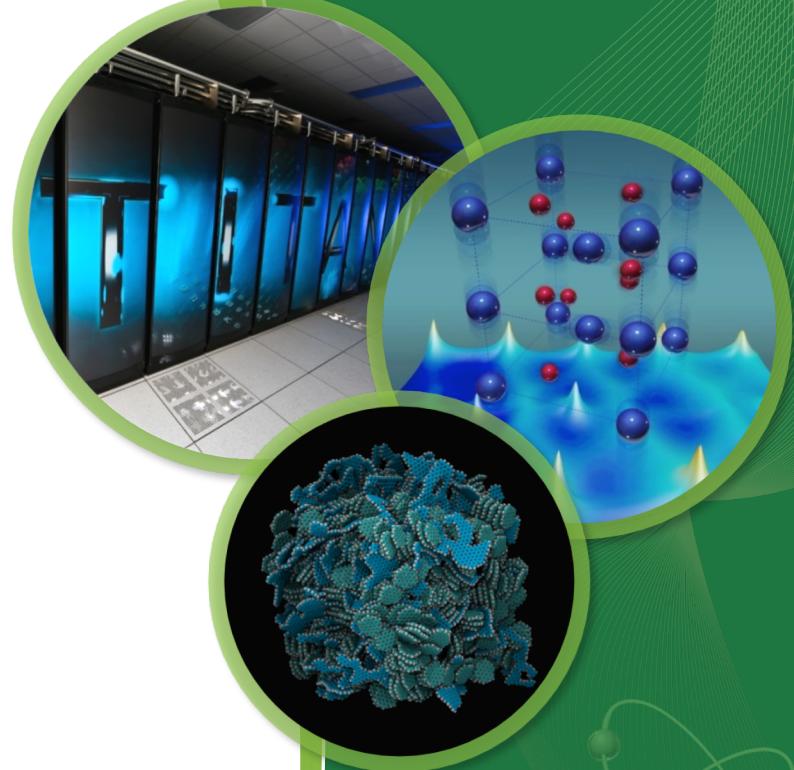
# A Prototype for a Scientific Data Catalog at ORNL

Whitney Nelson

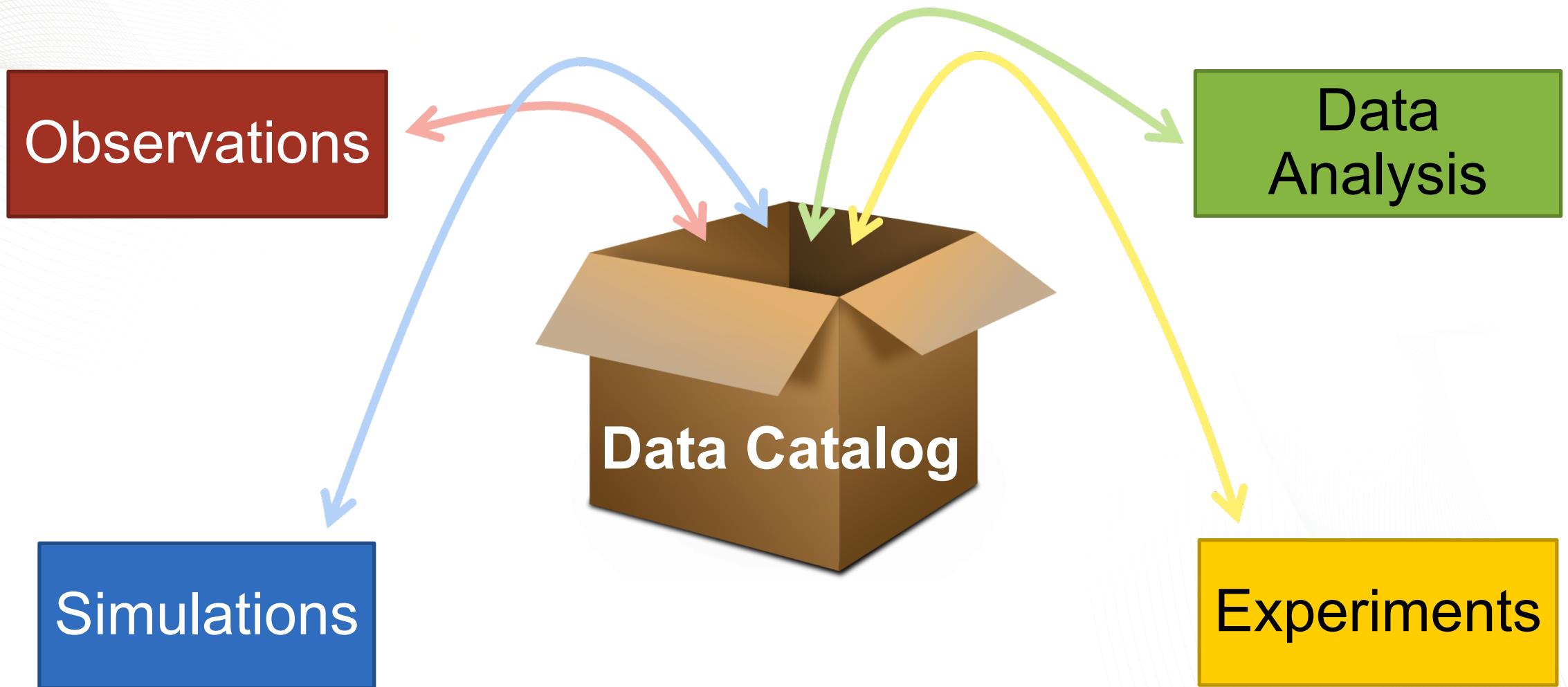
Mentors: Dr. Arjun Shankar, Dr. Suhas Somath

Advanced Data and Workflow Group

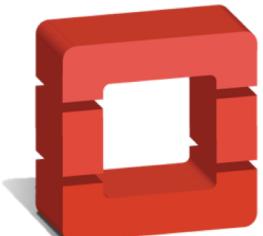
National Center for Computational Sciences



# Motivation: Enabling Scientific Discovery



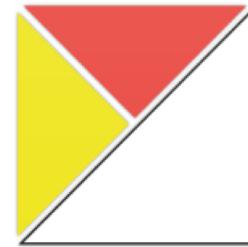
## Infrastructure



**openstack™**  
CLOUD SOFTWARE



## Software



**ckan**

**Apache**  
HTTP SERVER  
A colorful feather icon.

**NGINX**

**Solr**



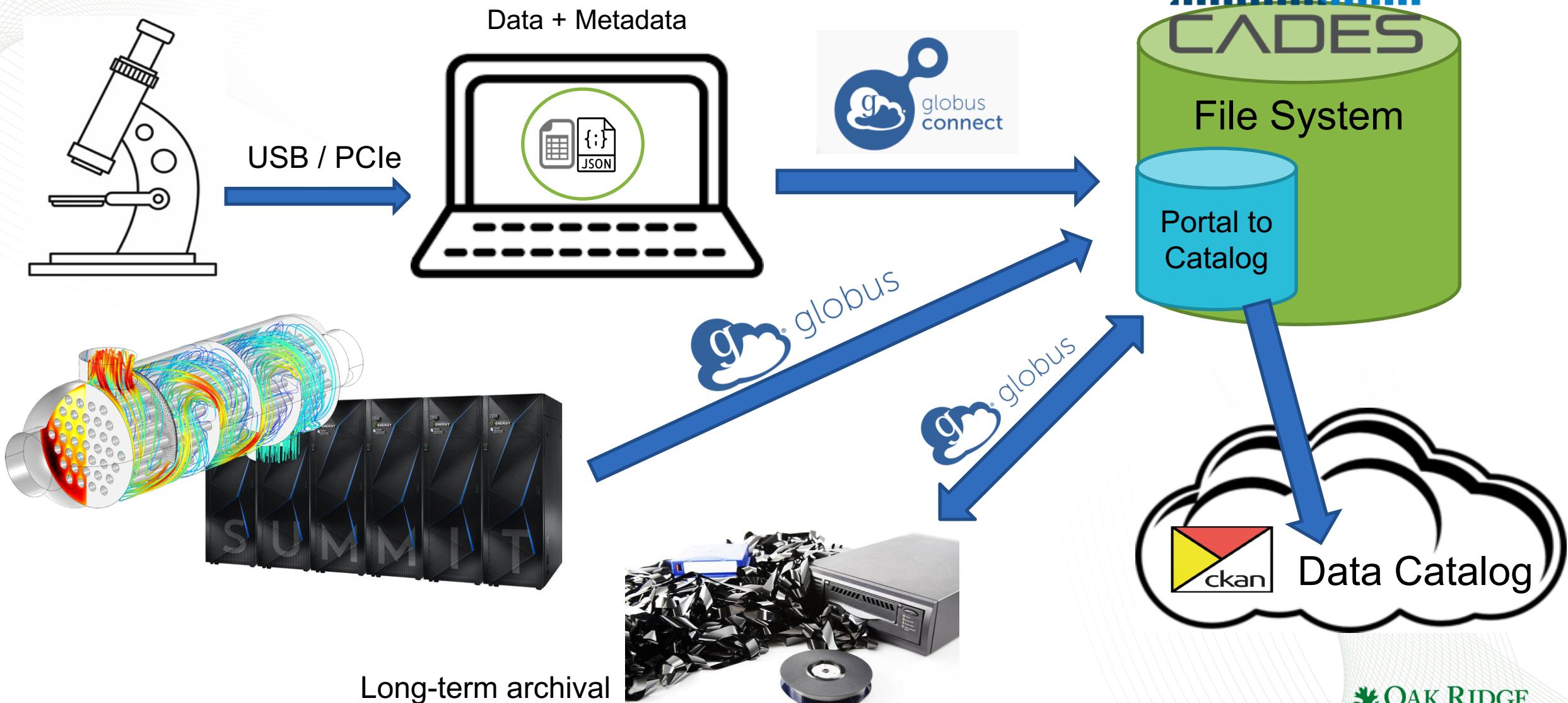
**PostgreSQL**

## Data Transfer Tool



**globus online**

# ORNL Data Catalog Solution



# Results

- Utilized CADES infrastructure and CKAN software to build data catalog framework
- Established best practices for transferring large datasets with Globus tool
- Enabled rich and relevant metadata uploads alongside datasets
- Created reference documentation
  - Users - uploading / downloading from data catalog
  - System administrators - for maintaining and expanding the data catalog

# Future Work

- Utilize CADES Scalify for storing large data files
- Extract metadata from external repositories
- Replace slow http request with Globus
- Enable https protocol for secure connections