

# Globus Data: Move, share, & automate

Andy Boughton  
([abought@globus.org](mailto:abought@globus.org))  
APECx Demos – February 2025



THE UNIVERSITY OF  
CHICAGO



Argonne  
NATIONAL LABORATORY





# Globus Transfer: Part of an integrated platform



## Learn more:

<https://www.globus.org/platform>

<https://www.globus.org/data-transfer>



## Today's notes:

<https://github.com/globus/apecx-demos/tree/main/transfer-data>



## Try it yourself

Get an endpoint: <https://www.globus.org/globus-connect-personal>

Manage files: <https://app.globus.org/>



# Modern data is decentralized



**How do we...**  
**...move, share, describe,**  
**discover, reproduce?**





# Why Globus Transfer?

- **Reliable transfers: Fire-and-forget**
  - Automatic retries if a host goes down
  - Resume where you left off
- **Designed for big data**
  - Robust error-checking for huge files
  - Take full advantage of fast networks
- **Advanced sharing features**





# Works with many storage systems





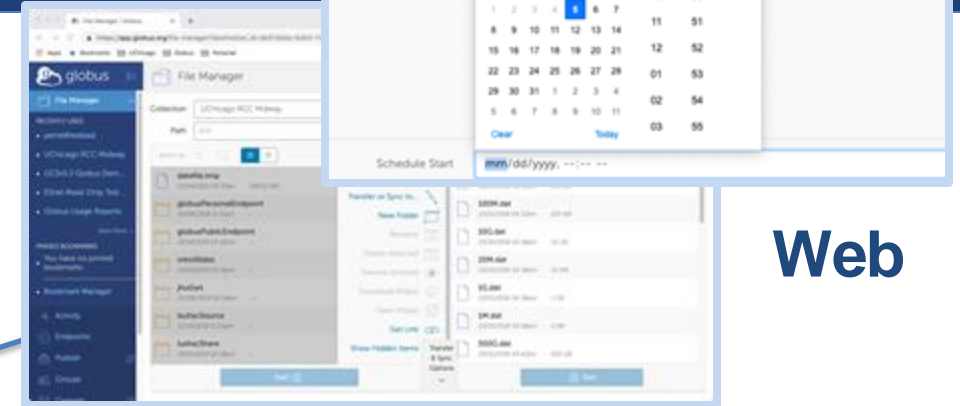
# Many ways to interact

```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

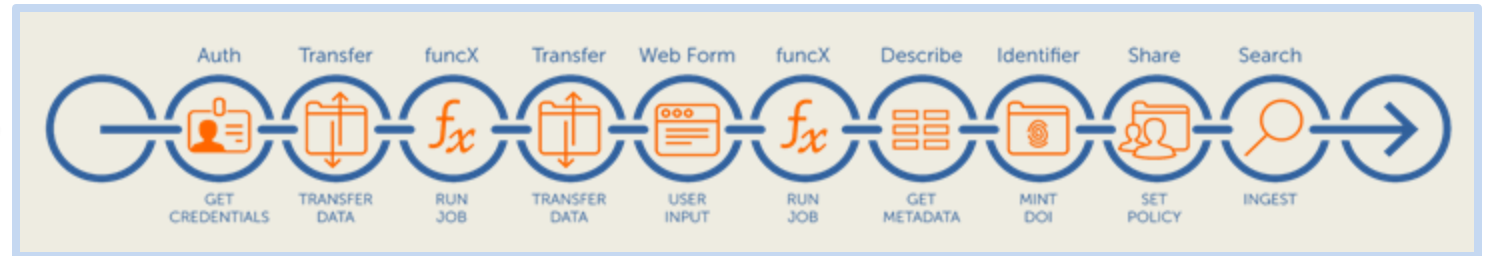
Options:
  -v, --verbose          Control level of output
  -h, --help             Show this message and exit.
  -F, --format [json|text] Output format for stdout. Defaults to text
  --map-http-status TEXT Map HTTP statuses to any of these exit codes:
                        0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
```

**CLI**



**Web**



**Flows**

```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
```

**Rest  
API**



# Globus High Assurance features

- **Granular permissions and access controls**
- **Can meet PII, PHI, CUI, or HIPAA BAA requirements**
- **Additional authentication rules**
  - Authenticate with a *specific identity*
  - Can require MFA and re-authentication
- **Options to ensure trust**
  - Local audit logging
  - Enforced encryption of all user data in transit
  - Choose which features are allowed





# Demo of file management capabilities

<https://app.globus.org/file-manager>







# Do your systems use Globus?

- **Find institutional subscriptions:**  
<https://app.globus.org/settings/subscriptions/search>
- **Find shared file storage systems:**  
<https://app.globus.org/collections>

\* Some listings are private. When in doubt, ask your systems administrator!

Powerful features

**Rethinking** what storage can do



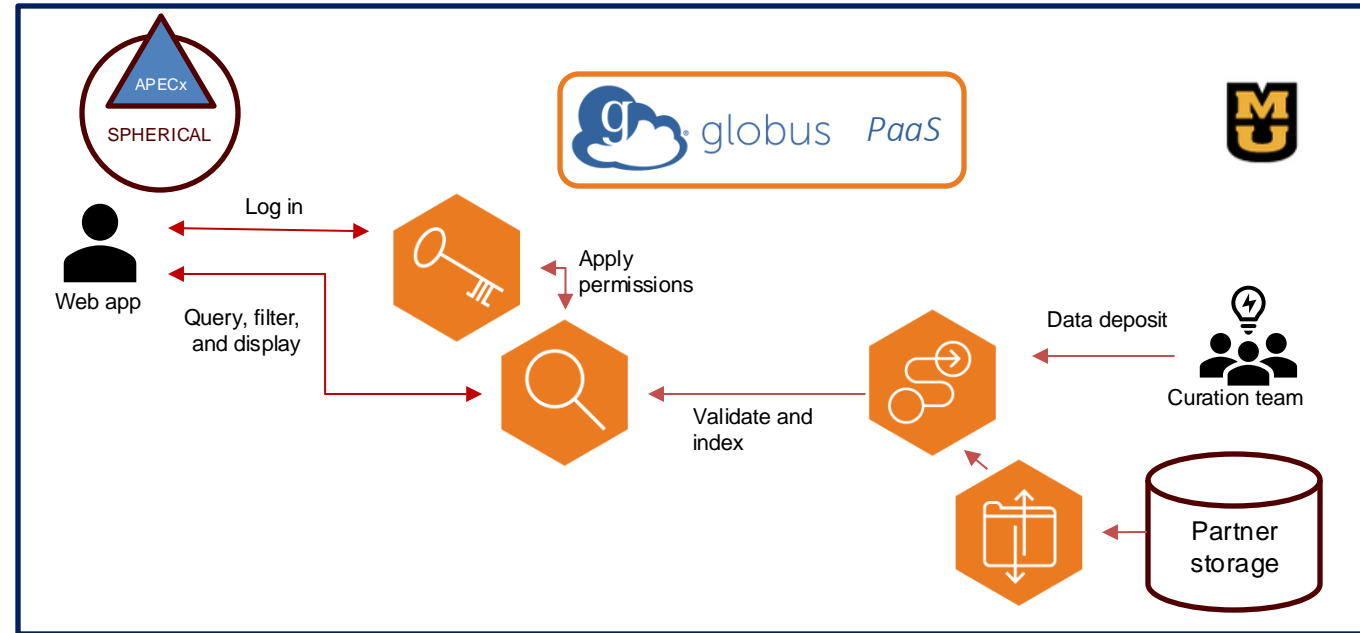
# Globus Platform: Features work together



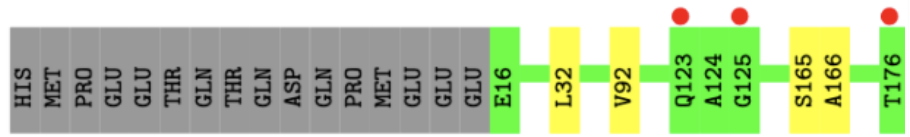


# Example: Data submission

- **Transfer is not just about files**
  - Can combine with a compute workflow
- **Globus Compute is scalable**
  - Web app can use compute nodes with existing HPC

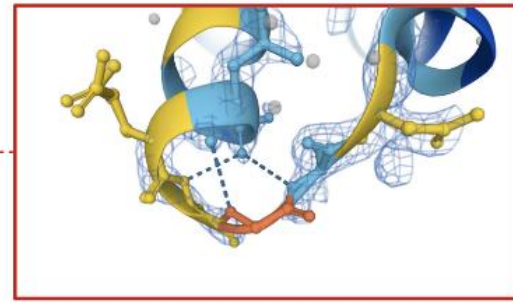
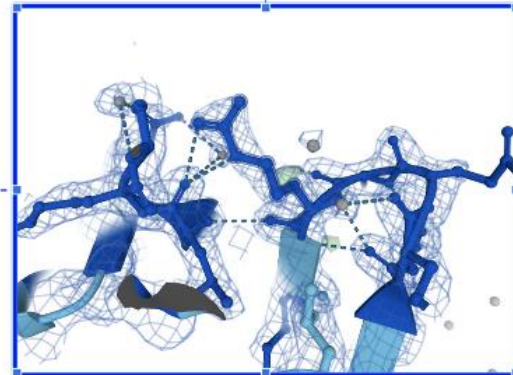
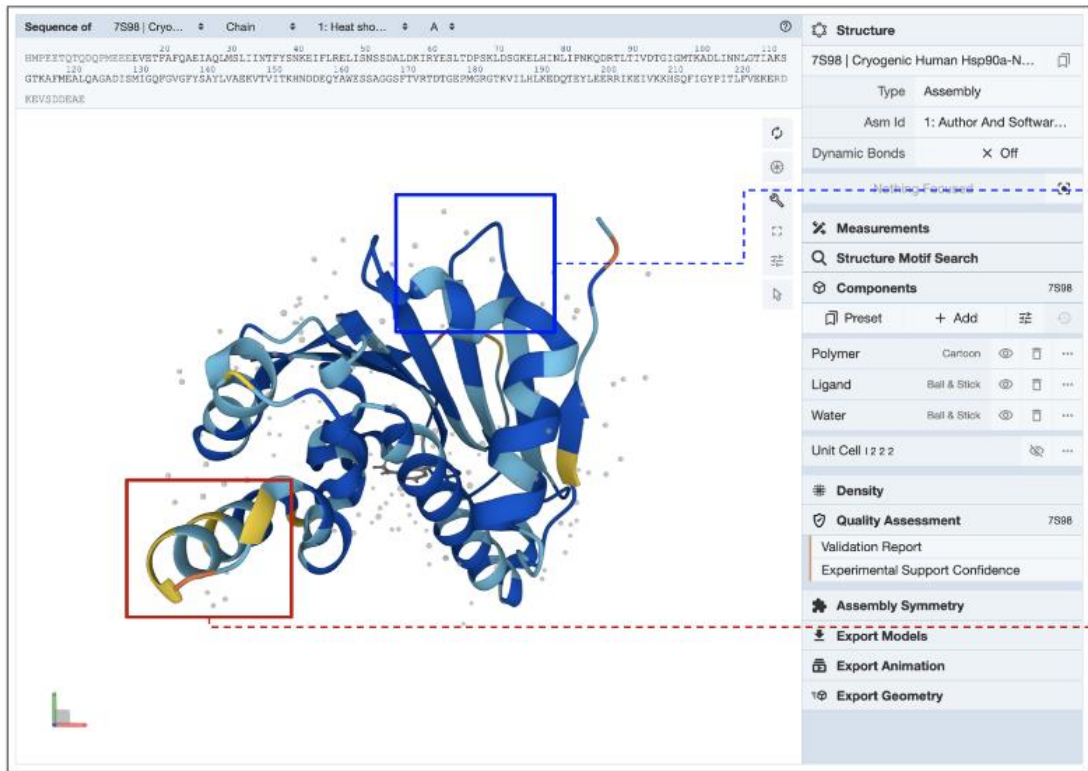
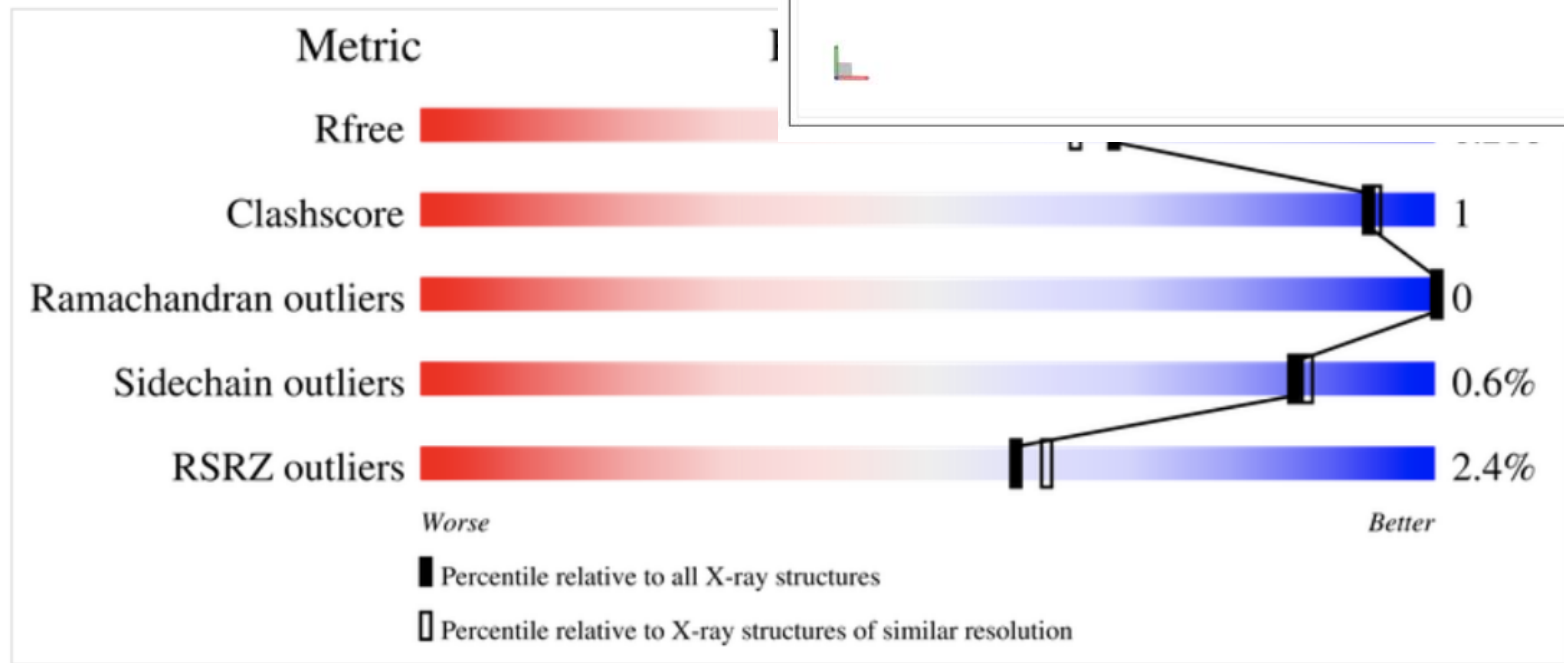


● Molecule 1: Heat shock protein HSP 90



Residues missing from model

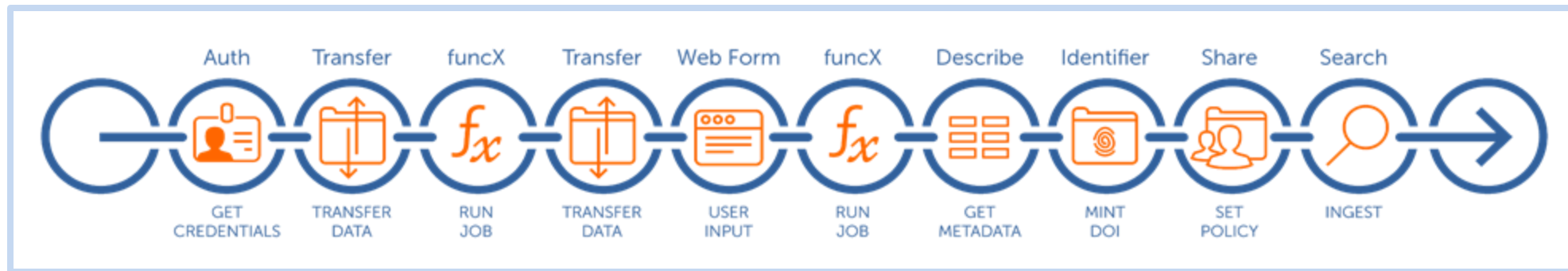
wwPDB Validation





# Workflows: Coordinate transfer steps

- Run compute (data validation) on submission
- Transfer in two stages (user submission to restricted destination)
- Index documents in **search** as part of upload process



## Atlanta International Airport Climate Data

### Tags

airport

### Region

south

### Sample

DATE	LATITUDE	LONGITUDE
19510101	33.62972	-84.44
19510102	33.62972	-84.44
19510103	33.62972	-84.44
19510104	33.62972	-84.44
19510105	33.62972	-84.44
19510106	33.62972	-84.44
19510107	33.62972	-84.44
19510108	33.62972	-84.44
19510109	33.62972	-84.44
19510110	33.62972	-84.44
19510111	33.62972	-84.44
19510112	33.62972	-84.44

Open in New Tab

Sample : 1951 Scatter Plot using Plotly

```
{ 4 items
  "@datatype": string "GMetaResult"
  "@version": string "2019-08-27"
  "subject": string "5352507d-1293-4827-8ba9-c2d4a0..."
  "entries": [ 1 item
    0: { 3 items
      "content": { 7 items
        "path": string "/portal/catalog/dataset_atl"
        "name": string "Atlanta International Airport ..."
        "id": string "5352507d-1293-4827-8ba9-c2d4a0..."
        "region": string "south"
        "sample": string "https://g-felc1.fd635.8443.dat..."
        "tags": [ 1 item
          0: string "airport"
        ]
        "globus": { 1 item
          "transfer": { 3 items
            "type": string "directory"
            "path": string "/portal/catalog/dataset_atl"
            "collection": string "a6f165fa-aee2-4fe5-95f3-97429c..."
          }
        }
      }
      "entry_id": null
      "matched_principal_sets": [] 0 items
    }
  ]
}
```

- Search record can specify where to find dataset
  - Download folder
  - View example file
- UI can render sample files in data collection via web UI





# Leverage existing investments

File Manager

Collection:

Path:

Bookmark

select all up one folder refresh list filter view

Jetstream2 home directory- max size 10GB

NAME	LAST MODIFIED	SIZE
testbench.png	11/25/202...	331.33 K

Share

Transfer or Sync to...

New Folder

Rename

Delete Selected

Download

Open

Upload

Get Link

Show Hidden Items

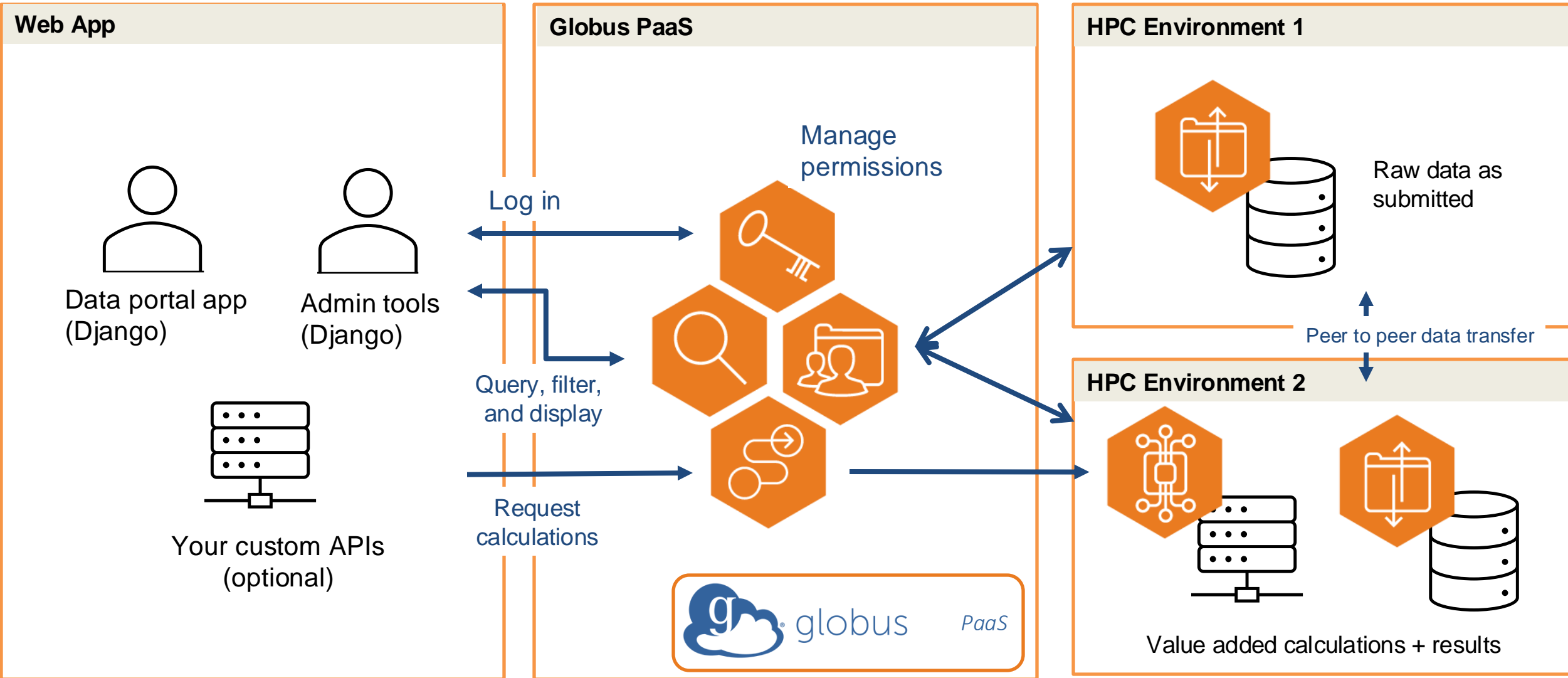
Manage Consent

Can use existing web tools to offer advanced features fast

Globus auth provides a consistent identity across many tools

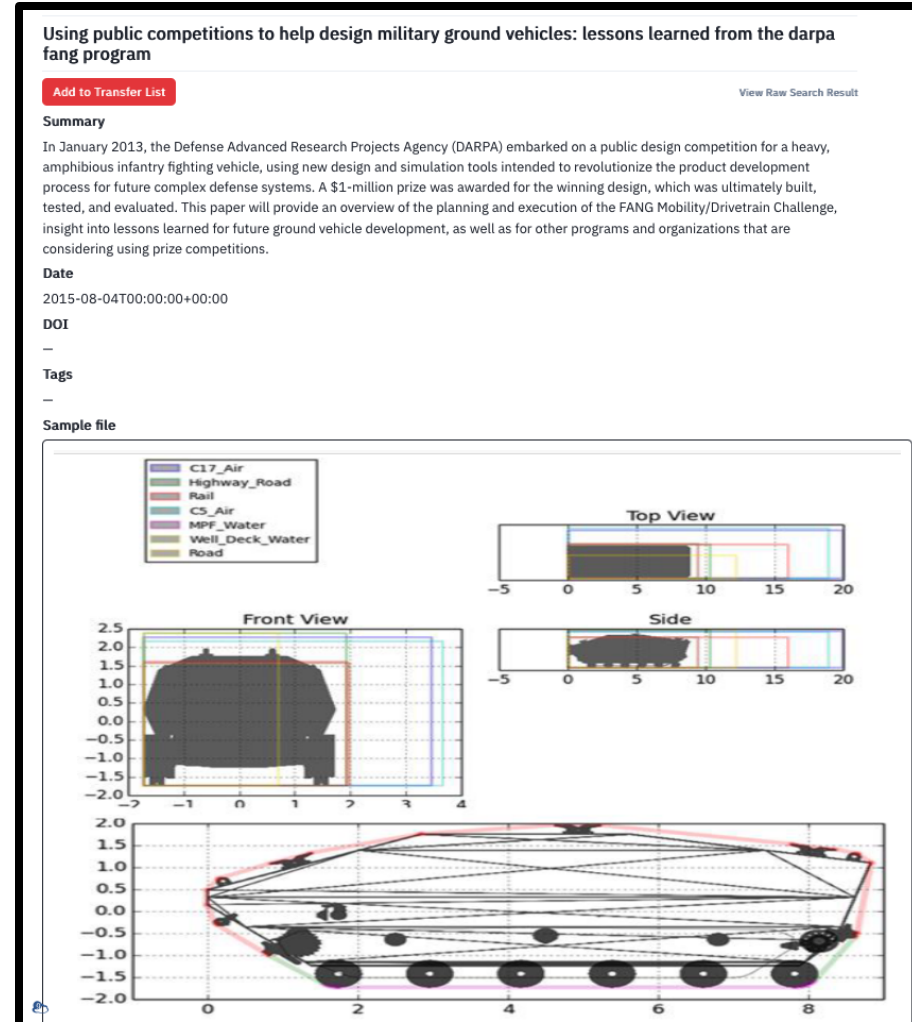


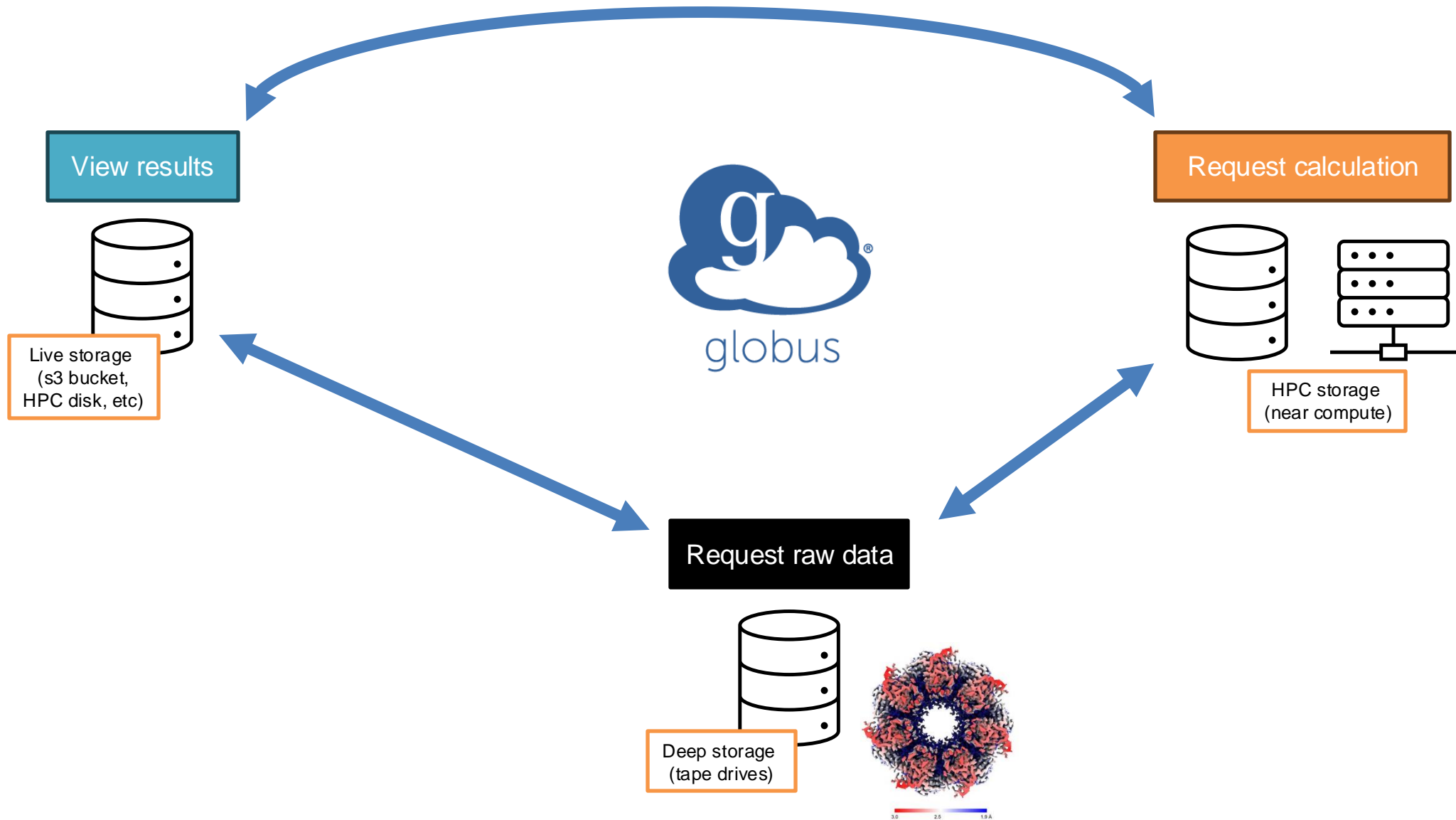
# Example: How users interact with the Portal



# Web sharing: Another kind of transfer

- **Globus connect server enables file access over HTTPS**
  - Adds feature on top of existing collections
  - Great for sharing files via web data portal
- **Follows the same security controls as other Globus services**
  - One central identity system, many ways to access files
  - Avoids the common problem of many systems with inconsistent access rules







# Consistent access controls

Shared With

Mapped Collection: alcf#



Add Permissions — Share With

USER OR GROUP	CREATED	EXPIRATION	READ	WRITE
Permissions granted by role				
[Redacted]	-	never expires		
[Redacted]	-	never expires		
[Redacted]	-	never expires		
Andy Boughton (andrew.boughton@globus.org)	-	never expires		
Path: /				
[Redacted]	2/4/2025, 11:28 AM	never expires		

Guest collections share PART of a dataset



APECX DATA AT ARGONNE LCF

Add Permissions - Share With

Path

/



Browse

Share With

- ☒ user - share with specific individuals
- ☐ group - make data accessible to members of a group
- ☐ all users - make data accessible to all logged in users of Globus
- ☐ public (anonymous) - make data accessible to everyone

Username or Email

enter a username or sear

Add

Permissions

☒ read

☐ write

Add Permission

Cancel

# Partnerships enable enhanced access

- **Argonne ALCF provides compute and storage to support big data projects**
  - We have 100TB to start
  - ALCF Eagle provides two 200PB Lustre stores: room to grow!
- **Globus is one of the only allowed ways to access ALCF without SSH**
  - Can be used with Globus application service accounts



# Discussion questions

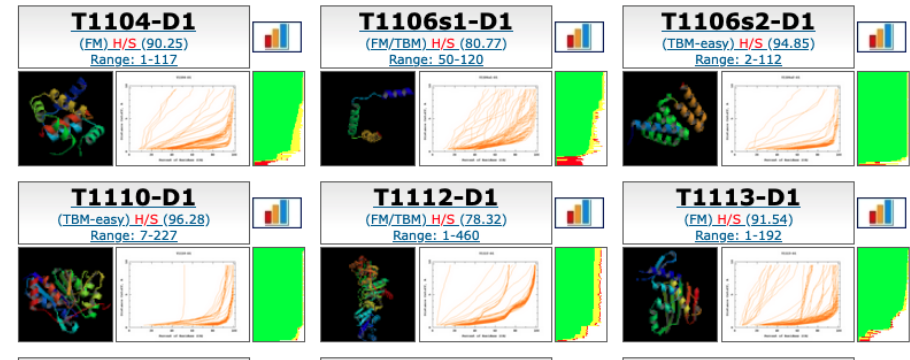
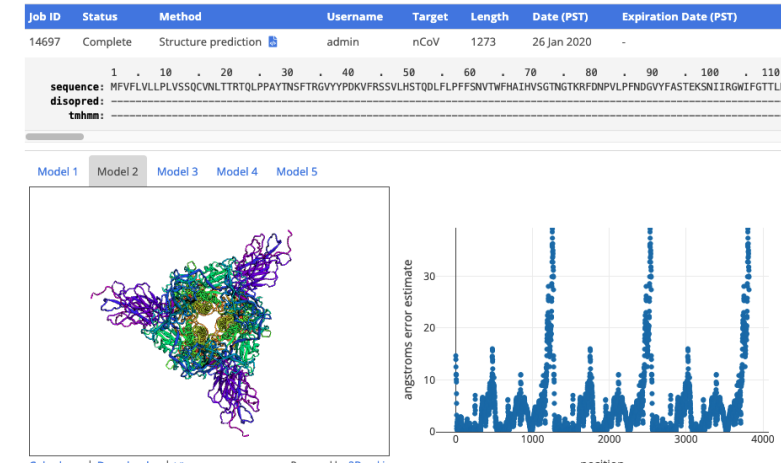






# Audience Questions: Function

- What sorts of insights do you hope to gain browsing?
  - List separate datasets (like Figshare, Zenodo, PDB)
  - Compare related predictions (like CASP)
  - One prediction per dataset, or many?
  - Results, or interpretation?





# Audience Questions: Data

- **Are there defined QC/validation workflows for the most common data types?**
  - What resources do they need to run?
- **Any special data requirements?**
  - Sensitive classifications (PHI, PII, CUI, etc)
  - HIPAA / BAA / high assurance



# Audience questions: General

- **What do we want the portal to *do*?**
  - We have the technology
  - You have the science