

Globus Search:

Find and explore distributed datasets

Andy Boughton
(abought@globus.org)
APECx Demos – December 2024



THE UNIVERSITY OF
CHICAGO



Argonne
NATIONAL LABORATORY





Find Data Anywhere

- **Search any type of data, stored across multiple institutions**
- **Advanced query features**
- **No infrastructure setup required**
- **Integrates with other Globus services (like auth)**

Case Study

ESGF (Climate Science)

>5M datasets

25M+ files

>100GB metadata

>10 PB of data

Complex search facets

Global collaboration





Globus Search: Part of an integrated platform



Learn more:

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

<https://docs.globus.org/api/search/overview/>



Today's Demos:

Notes: <https://github.com/abought/apecx-demos/search>

Live site: <https://abought.github.io/apecx-demo-static-search-portal/>

Interactive Demo

<https://abought.github.io/apecx-demo-static-search-portal/>





Audience Questions

- **What kinds of data would you expect to search?**
 - Articles (free-text or facet search)
 - Software (AI/ML models, etc)
 - Biological datasets (like molecular structures)
- **What would help people to make decisions about which search results to look at?**
 - Preview review select asset types (graph, table, PDB viewer, etc)
 - Display of “value added” results on top of raw inputs
 - Click to run follow-up analysis
- **Focus on common data types first: MU deliverables etc**



Further Discussion

- **What features should we explore next?**
 - Review of suggested user stories (Lee)
 - Identify key building blocks needed to get started
- **How will work on the portal proceed?**
 - Timelines
 - Personnel