



# Overview and Demonstration

April 16, 2024

Contact Us: [hpc.edu.train@gmail.com](mailto:hpc.edu.train@gmail.com)

CyberTraining: Pilot: HPC ED: Building a Federated Repository and Increasing Access through Cybertraining is supported by NSF grant [OAC-2320977](https://www.nsf.gov/awardsearch/showAward?AWD_NUM=2320977)

# Topics

---

- HPC-ED Overview
- Share & Discover Prerequisites
- Share & Discover Walkthrough
- Case Studies
- Q & A
- Help & Feedback

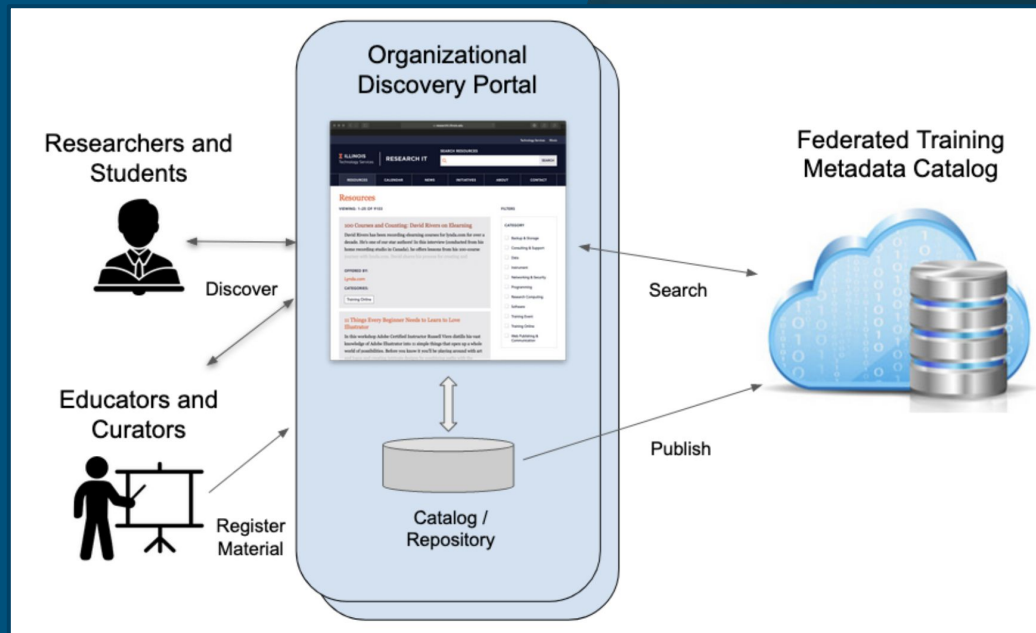


**HPC ED**

[HPC-ED.GITHUB.IO](https://hpc-ed.github.io)

# HPC-ED Overview

- 1 year pilot project
- Project Goals:
  - Make existing HPC Training and Education materials more findable
  - Create a federated catalog that can be easily searched
  - Facilitate community use of project tools and provide feedback
- Project site: <https://hpc-ed.github.io/>
- See documentation here: <https://github.com/HPC-ED/HPC-ED.github.io/wiki>





# Prerequisites

1. Plan: Do you want to Share and/or Discover? One resource or many?
2. Where is your metadata? Do you have read/write access? This impacts how you will share/search/build
3. Review the documentation: <https://github.com/HPC-ED/HPC-ED.github.io/wiki/>  
Software prerequisites
  - a. Install the Globus CLI as detailed at <https://docs.globus.org/cli/>
  - b. If you plan to publish from a Python program use the [Globus Python SDK](#)

## Share

- Review the metadata in the example JSON file and the metadata document
- Provider\_ID (Ownership) - Assigned by request; send email to [hpc-ed-train@gmail.com](mailto:hpc-ed-train@gmail.com).

## Search/Discover

- Plan: which parameters are important to you? How will your community search and view the results?
- Try the developer's portal: <https://search-pilot.operations.access-ci.org/hpc-ed-v1>
- Review the five search options: <https://github.com/HPC-ED/HPC-ED.github.io/wiki/Searching-Metadata>

# JSON Example



```
{
  "ingest_type": "GMetaEntry",
  "ingest_data": {
    "subject": "urn:ogf.org:glue2:access-ci.org:resource:cider:infrastructure.organizations:898:localid_456",
    "visible_to": ["public"],
    "id": "std",
    "content": {
      "Abstract": "Are you interested in adding programming skills to your professional toolkit?",
      "Authors": ["Jane Doe", "John Doe"],
      "Cost": null,
      "Duration": 240,
      "Expertise_Level": ["All"],
      "Keywords": ["linda"],
      "Language": "en",
      "Learning_Outcome": ["Proficient"],
      "Learning_Resource_Type": "recorded lesson",
      "License": null,
      "Provider_ID": "urn:ogf.org:glue2:access-ci.org:resource:cider:infrastructure.organizations:898",
      "Rating": 4.4,
      "Resource_URL_Type": "URL",
      "Start_Datetime": "2024-01-26T12:20:51.668905",
      "Target_Group": ["Research groups"],
      "Title": "C++ Essential Training for Career Changers GMetaEntry",
      "URL":
        "https://www.lynda.com/C-tutorials/C-Essential-Training-Career-Changers/572157-2.html",
      "Version_Date": "2023-12-14T05:03:28.952096Z"
    }
  }
}
```

#### Examples:

<https://github.com/HPC-ED/HPC-ED.github.io/wiki/Publishing-Metadata>

#### Metadata:

<https://github.com/HPC-ED/HPC-ED.github.io/wiki/Metadata-Description>



# Share and Discover Walkthrough

1. Federated Training Metadata Catalog architecture:
  - a. Built using the Globus Search service - <https://www.globus.org/platform/services>
  - b. Globus Search is AWS hosted highly scalable, reliable, and free to all Globus subscribers
2. Publishing Metadata: <https://github.com/HPC-ED/HPC-ED.github.io/wiki/Publishing-Metadata>
  - a. Pre-requisites
  - b. Publishing formats
  - c. Publishing methods and tools
  - d. Checking publishing status with the CLI
  - e. The developer site: <https://search-pilot.operations.access-ci.org/>
3. Searching Metadata: <https://github.com/HPC-ED/HPC-ED.github.io/wiki/Searching-Metadata>
  - a. Using the CLI or the Python library
  - b. Using the Globus SDK for Javascript to build a Web Application
  - c. Embedding someone else's Javascript Web Application





# Case Study 1: JOCSE

## An exercise in Jekyll Screen Scraping

- Jekyll site hosted on GitHub containing journal volumes, with individual articles
- Site organized into markdown files in an "almost-YAML" format
- Metadata exists in a uniform format, so automated processing is possible, but not directly stored in a database
- Used a python script to parse into a dictionary, using GMetaList ingest type

```
---
abstract: "This paper presents a new mathematics elective for an unc
history, often highlighted by extreme abstraction and difficulty.
This changed in the 1960s when Bruno Buchberger created an algorithm
audiences:
  - "Educator"
  - "Learner/Student"
  - "Professional/Practitioner"
  - "Researcher"
authors:
  - "Adam E. Parker"
education-levels:
  - "Higher Education"
  - "Graduate/Professional"
end-page: 7
keywords:
  - "Computational Science"
  - "Course content"
  - "Grobner Bases"
  - "Methods of instruction"
  - "Tools for teaching"
start-page: 2
subject:
  ingest_data: {
    "gmeta": [
      {
        "content": {
          "Abstract": "This paper presents a new mathematics elective for an
          "Authors": [
            "Adam E. Parker"
          ],
          "Cost": "no",
          "Keywords": [
            "Computational Science",
            "Course content",
            "Grobner Bases",
            "Methods of instruction",
            "Tools for teaching"
          ],
          "Language": "en",
          "Learning_Resource_Type": "Journal article",
          "License": "null",
          "Provider_ID": "urn:ogf.org:glue2:access-ci.org:resource:cider:inf
          "Resource_URL_Type": "DOI",
          "Target_Group": [
            "Research Community",
            "Educators",
            "Students"
          ],
          "Title": "Computational Algebraic Geometry as a Computational Scie
          "URL": "doi:10.22369/issn.2153-4136/1/1/1",
          "Version_Date": "2010-12-21T22:08:00-05:00",
          "id": "std"
        },
        "subject": "urn:ogf.org:glue2:access-ci.org:resource:cider:infrastructure
        "visible_to": [
          "public"
        ]
      }
    ]
  }
```



# Case Study 2: TACC

**What:** Shared ~5 online tutorial videos

**How:** Wrote a Google Sheets extension to generate individual JSON code for use through the Globus CLI.

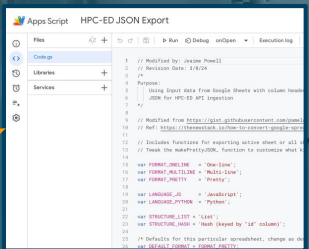
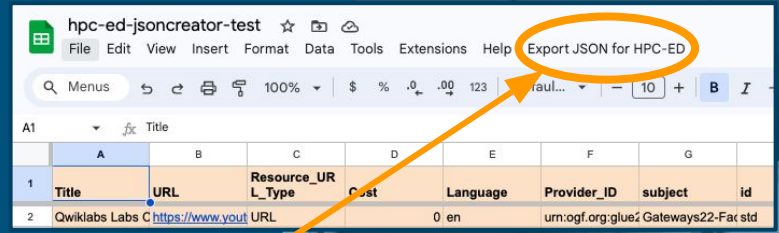
**Why:** Thought was an option to populate a Google Sheet using Google Forms and then export the resulting JSON as a GUI method to add objects to the catalog.

### What went wrong:

- Could not locate the “**Provider\_ID**”
- Could not find a ready made Google Sheets to JSON exporter that generated the proper format so I modified other code
  - *Modified Code:* <https://gist.github.com/jeaimehp/2656730f3cd021b59aa845e6f8d483e9>
  - *Original Code:* <https://gist.github.com/pamelafox/1878143/>
  - *Ref:* <https://thenewstack.io/how-to-convert-google-spreadsheet-to-json-formatted-text/>
- Could not update entries once published

### Tips:

- (Seems obvious but...) Ensure the data you enter is not only accurate but also check spelling and grammar 🙏
- GUIs are great for a limited number of entries
- Google Sheets Extensions are written in JavaScript and deployed through “Apps Script”
- Ensure your Google Sheet headings match the JSON object key item names





# Case Study 3: SDSC

- What: Used command line tools used to upload 60 training items and to search for training items
- How/What went right:
  - Wrote Python script to convert existing Training Catalog JSON files to HPC-ED formatted JSON files
  - Using "Publishing Metadata" instructions in wiki from command line: can do single or batch (multiple)
  - Updating with cron jobs
- What went wrong/challenges:
  - All of our material is located at the web hosting system → needed permissions to run cron jobs
  - Setting up the right globus environment on web server system (75 line script)
  - Determining search criteria is a challenge
  - Debugging typos in json file.
- Tips:
  - Testing on local host is "easy", but you need Globus Tools installed
  - Harder to do on web hosting or other machines: need root? Custom versions of python or other ENV vars?
  - Be prepared for different challenges on the system where you have your training material stored
  - Avoid downloading all entries: it takes a long time and storage

SDSC SAN DIEGO SUPERCOMPUTER CENTER  
COMPUTING WITHOUT BOUNDARIES

UC San Diego

HOME • EDUCATION & TRAINING • INTERACTIVE VIDEOS

Interactive Video Tutorials

Here we provide some sample privileges of an interactive video tool designed to allow the user to navigate extended video recordings with greater ease, hopefully leading to better learning.

Search Title:

Filter By: All Years

Clear Search

**COMPLEXES: HPC Hardware Overview**

All users of advanced CI can benefit from basic understanding of the hardware to determine which factors affect application performance. This session provides an overview including CPUs (processors, cores, hyperthreading, instruction sets), the anatomy of a compute node (sockets, memory, at accelerators, accelerators) and cluster architecture (login and compute nodes, interconnection using Linux tools, networking tools.)

```

"202404-HPC-Hardware-Overview": {
  "name": "202404-HPC-Hardware-Overview",
  "title": "COMPLEXES: HPC Hardware Overview",
  "start": "2024-04-01",
  "end": "1712259800",
  "desc": {
    "long": "<span style='font-weight: 400;'>All users of advanced CI can benefit from a basic understanding of the hardware to determine which factors affect application performance. In this session we provide an overview including CPUs (processors, cores, hyperthreading, instruction sets), the anatomy of a compute node (sockets, memory, at accelerators, accelerators) and cluster architecture (login and compute nodes, interconnection using Linux tools, networking tools.)</span></p>
  }
}
  
```

at 202404-HPC-Hardware-Overview.json

```

"subject": "QIMACS",
"ingest_date": {
  "subject": "urn:ogf.org:glue2:access-ci.org:resource:cider:infrastructure.org:organizations:856:202404-HPC-Hardware-Overview",
  "subject": "QIMACS"
}
  
```

ACCESS Operations Search Pilot

Search Pilot / HPC Training Material (HPC-ED) - Alpha catalog v1 (match all) / COMPLEXES: HPC Hardware Overview

← Back to Search

**COMPLEXES: HPC Hardware Overview**

Training Material Metadata

Title	COMPLEXES: HPC Hardware Overview
Abstract	A brief introduction into what makes up a HPC system and how users should use programming required.
Authors	Nicole Wolter
Expertise Level	[All]
Learning Outcome	None
Learning Resource Type	recorded lesson
Target Group	[Research groups]
Keywords	[Expanded: "HPC Training", "TSCC", "Voyager"]



# Case Study 4: CVW

What: Shared ~40 online tutorials

How: Wrote a C# program to pull metadata from local JSON files and build a multi-resource input JSON file

What went wrong:

- Local metadata needed to be augmented and cleaned up
- Quotation marks in the abstract text
- Failed to properly replace example ID with my ID

Tips:

- Study the JSON fields; which will stay the same for all of your resources?
- Test your json file for syntax errors with a simple tool e.g. <https://jsonlint.com/>
- Test ingesting one before ingesting many
- Read the documentation
- After issuing a Globus command, note the Task ID, then issue `globus search task show <Task ID>`
- After uploading, check the developer's portal: <https://search-pilot.operations.access-ci.org/hpc-ed-v1>
- Plan a procedure you can execute easily as needed



# Q & A

---



# Help and Feedback

- **Office Hours:** April 30, 2024 at 4:00 PM Eastern
- **Slack Channel:** Use <https://tinyurl.com/Slack-HPC-ED>  
Join the TACC-Learn then the cybertraining-hpc-ed-testers channel (dm any of us for help).
- **Email:** [hpc.edu.train@gmail.com](mailto:hpc.edu.train@gmail.com)
- Complete this [interest survey](https://tinyurl.com/InterestSurveyHPC-ED) -  
<https://tinyurl.com/InterestSurveyHPC-ED>.  
It would be very helpful to us to track progress of the pilot.



Rich Knepper  
Kate Cahill  
Charlie Dey

Brian Guilfoos  
David Joiner  
Susan Mehringer

JP Navarro  
Je'aime Powell  
Mary Thomas

**THANK YOU!**

**HPGED**  
HPC-ED.GITHUB.IO