



OAC-2320977

# CyberTraining: Pilot: HPC ED: Building a Federated Repository and Increasing Access through Cybertraining

10/1/2023 – 9/30/2024



PI Rich Knepper, Cornell University; co-PIs Mary Thomas, SDSC; David Joiner, Kean U.; JP Navarro, U. Chicago/Argonne; Jeaime H. Powell, Texas Advanced Computing Center

## Introduction

Cybertraining materials abound, but they can be difficult to find, and often have little information about the quality or relevance of offerings. Using existing software technology, the HPC ED Pilot project team will build a platform for the community to better share and find training materials through a federated repository.

- Organizations and authors will retain physical and legal ownership of their materials by sharing only catalog information.
- Organizations can expand local portals to use the best and most appropriate materials from both local and remote sources.
- Learners can take advantage of materials that are reviewed and described more clearly.
- Learners and organizations will quickly and accurately find materials via standardized descriptive tagging.

The pilot will demonstrate that resource providers, campus portals, schools, and other institutions can both incorporate training from multiple sources into their own familiar interface and publish their local training materials to a much wider audience.

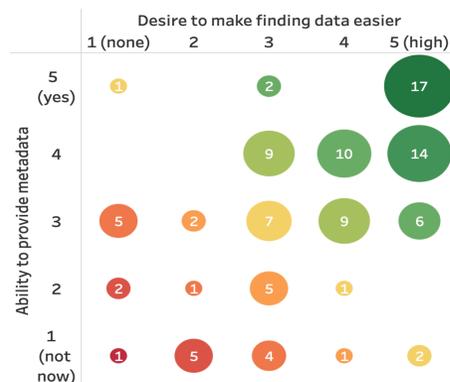
## Community Survey Results

A survey of Training Developers, Curators, and Consumers was conducted to learn if there would be a benefit to improving how HPC training materials are shared and discovered. Results showed that there are many barriers that make finding materials difficult; the most cited was not being able to find materials at the right depth or level.

Table 1: Results from the question: What are the barriers that you have encountered when searching for materials?

| Barriers encountered  | Responses |
|---|-----------|
| I can't find materials on the topic I need  | 35        |
| I can find materials on the topic, but not at the depth or level I need                             | 72        |
| I find too many materials, and I can't effectively sort through them all                            | 44        |
| I am aware of specific appropriate materials, but search engines don't list them in the top results | 26        |
| Other   | 28        |
| <b>Total</b>  | <b>93</b> |

Figure 1: Results from two survey questions: Does your organization want to make it easier for the public to find your training and education materials?



Would your organization be willing and able to share your training and education materials in a public catalog by providing metadata about your materials in a standard format?

The results imply that the community sees a need to improve the discovery of materials and many have an interest in and the ability to contribute to a solution.

## Federated Architecture

In the workflow below, Training Developers/Curators will use our secure API to submit a request to publish their materials and events into the Federated Catalog. Once approved, the information is entered into the catalog via a second API. The catalog of resources has the potential to encompass thousands of training resources and events that will be made available through an API that allows sites to add resources to their local portals and share resources with the catalog.

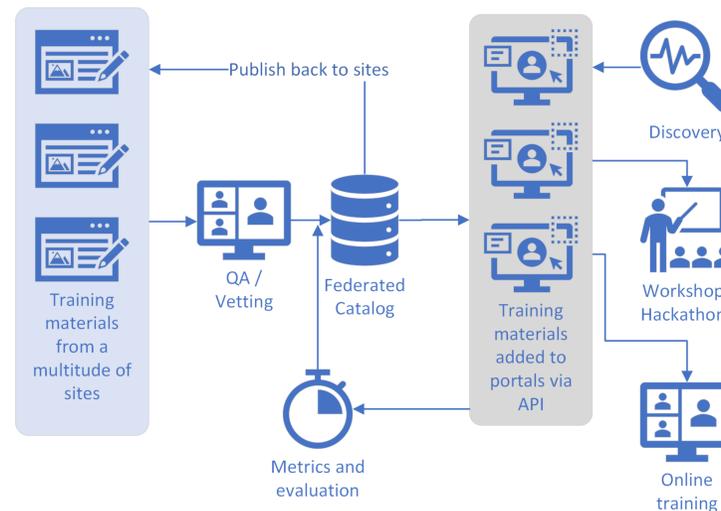


Figure 2: HPC ED Federated Architecture Workflow

## Metadata, Taxonomy, and Ontologies

We will work with existing community efforts to build a set of standardized HPC training metadata; this is critical for publishing and discovering training information effectively.

- Where possible, we will identify and use existing, common metadata sets, taxonomies, and ontologies.
- Where needed, we will identify and add new terms to these existing ontologies and work with existing communities to update them or to develop the HPC/CI training materials ontology.

Initial metadata will describe the training material, its access methods, educational characteristics, and the publisher and material source.

## Sharing/Publishing Materials

- Using our API, any authorized organization can automatically publish standard metadata from their local catalog to the Federated Catalog.
- For developers that have materials but do not have a local catalog, we plan to coordinate with the ACCESS Support project which already offers a way to publish materials into their reference material catalog.
- We will support FAIR principles (Findable, Accessible, Interoperable, and Reusable) for the sharing of training materials.

## Discovery of Materials

A federated training discovery/search RESTful API with advanced search capabilities will be used to (1) enable individuals to perform precise searches on specific metadata values, (2) find materials from a particular organization, author, or targeted skill level, (3) perform more advanced key terms-based searches that rely on our related HPC training material ontology and taxonomy work, and (4) enhance an organization's portal or website by including Federated Catalog items.

## Quality Assurance

We will assure the accuracy of the metadata of items shared through the federation by (1) verifying the status and nature of materials, (2) validating accuracy, (3) accrediting that metadata accurately reflects the catalog item, (4) including community input via a 5-star rating system, and (5) monitoring the existence and uptime of links.

Review information will be shared with contributors, providing a value add. Some processes will be fully automated, others assisted by AI, and others by human labor as well as crowd-sourced input.

## Project Training and Early Adopters

Early adopters will be alpha clients and define/refine methodologies, processes, and initial user interface templates. Letters of commitment have been received from projects that are enhancing coursework at MSI institutions, training Cyberinfrastructure Professionals (CIP), and collaborating with various ACCESS projects.

We will engage with MSIs, non-research/research institutions, and industry via workshops, hackathons, and BoFs to:

- Share and discover catalog materials and test procedures;
- Contribute to building a set of standardized HPC training materials;
- Foster community engagement with the Federated Catalog.

## Building/Sustaining Community

• A working group was formed within the ACM SIGHPC Education Chapter to discuss metadata standards for sharing materials.

• A BOF was held at PEARC23 to gather community input on what is needed to make existing materials more findable, accessible, interoperable, and reusable. An affinity group with ACCESS was recently created as a result of the BOF.

• We are sharing regular updates through the HPC ED Google Group mailing list and newsletter. To join the HPC ED Google Group and receive project updates, email hpc.edu.train@gmail.com.

