Scaling HPC Education

Susan Mehringer, Cornell University Center for Advanced Computing, shm7@cornell.edu

Mary P. Thomas,San Diego Supercomputer Center, mpthomas@ucsd.edu

Kate Cahill, New Jersey Institute of Technology, katharine.cahill@njit.edu

Charlie Dey, Texas Advanced Computing Center, Charlie@tacc.utexas.edu

David Joiner, Kean University, djoiner@kean.edu

Richard Knepper, Cornell University Center for Advanced Computing, rich.knepper@cornell.edu

John-Paul Navarro, Univ. of Chicago/Argonne National Lab, navarro@anl.gov

Jeaime H. Powell, Texas Advanced Computing Center, jpowell@tacc.utexas.edu

Scaling HPC Education:

Introducing the HPC ED pilot project, a federated training repository that is designed to allow resource providers, campus portals, schools, and other institutions to both incorporate training from multiple sources into their own familiar interfaces and to publish their local training materials.

BPHTE23: S. Mehringer, M. P. Thomas, K. Cahill, D. Joiner, R. Knepper, and J. H. Powell, "Scaling HPC Education,"

J. Comput. Sci. Educ., 2023.

Goal – Improved material sharing & discovery

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.
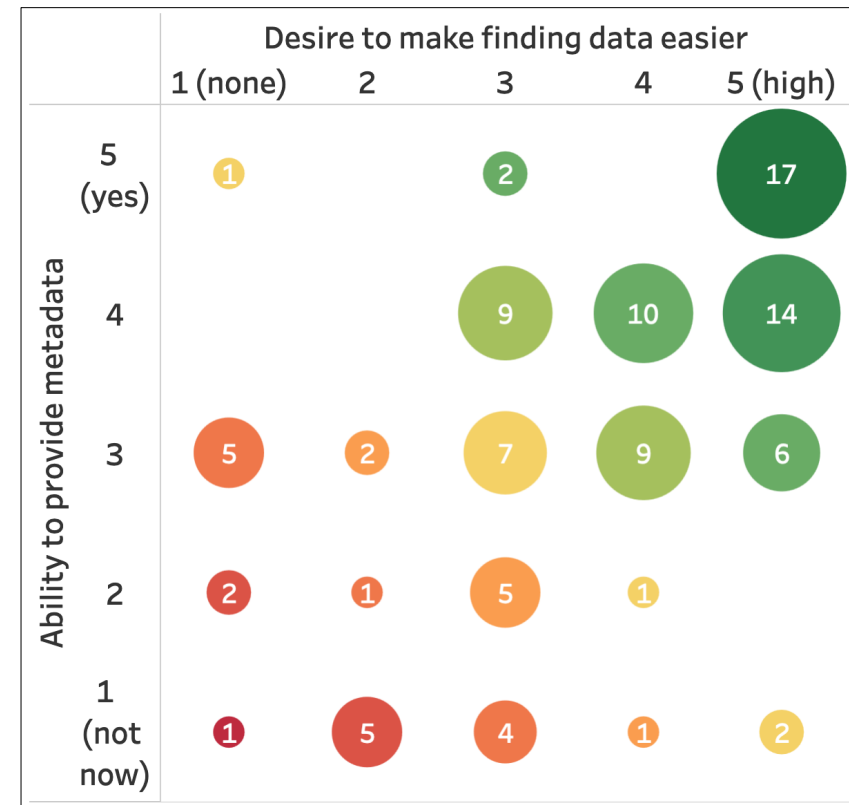
- Organizations and authors will retain physical and legal ownership of their materials by sharing only catalog information.

- 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

In 2022, a survey was conducted to learn if there would be a benefit to improving how HPC training materials are shared and discovered.
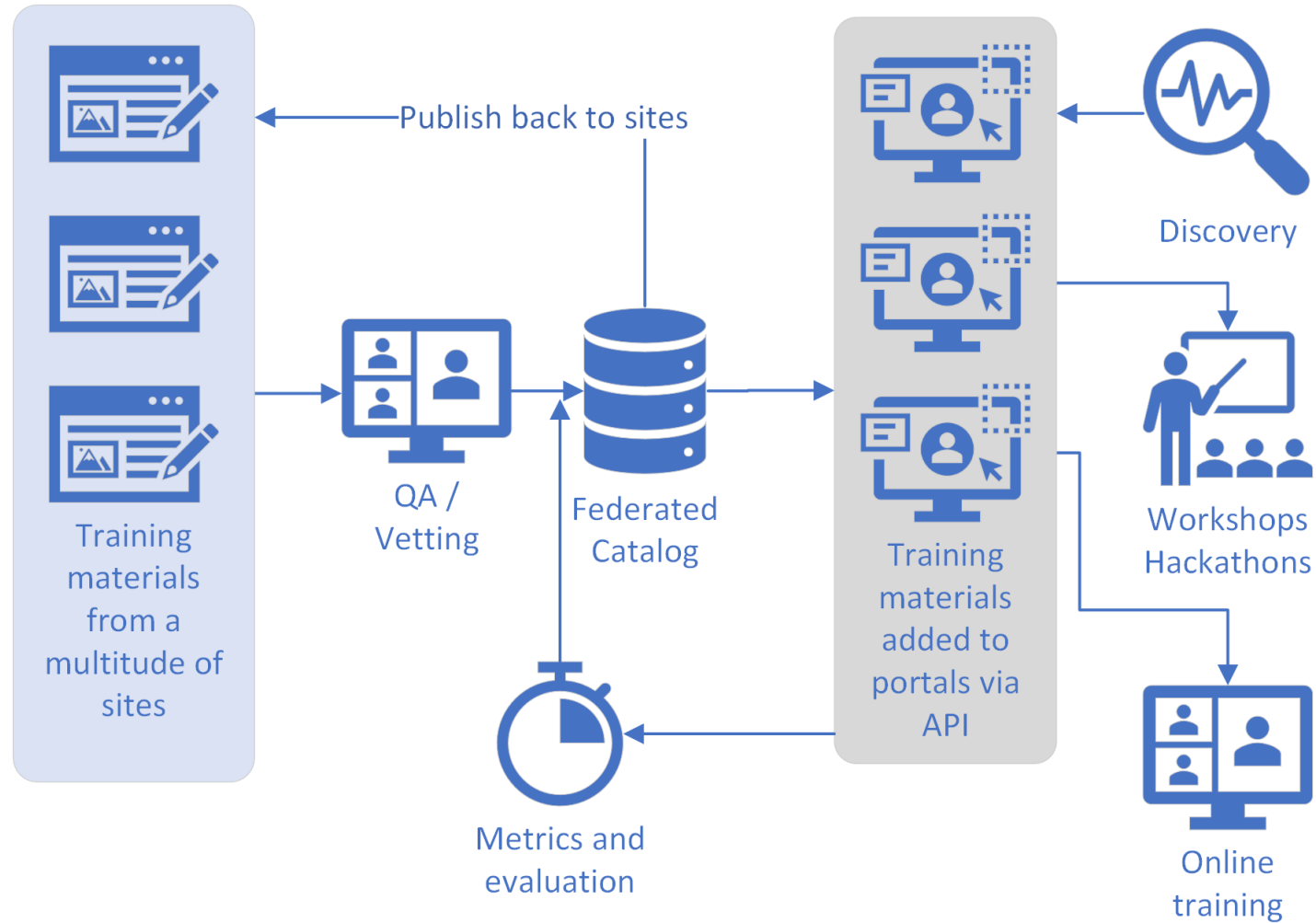
Barriers encountered	Responses
I can't find materials on the topic I need	35
I can find materials on the topic, but not at the depth or level I need	72
I find too many materials, and I can't effectively sort through them all	44
I am aware of specific appropriate materials, but search engines don't list them in the top results	26
Other	28
Total	93

Desire to make finding data easier

Ability to provide metadata	1 (none)	2	3	4	5 (high)
5 (yes)	1		2		17
4			9	10	14
3	5	2	7	9	6
2	2	1	5	1	
1 (not now)	1	5	4	1	2

Results: https://www.cac.cornell.edu/about/pubs/Survey2022.pdf

S. Mehringer, K. Cahill, S. Lathrop, C. Dey, M. Thomas, and J. H. Powell, "Assessing Shared Material Usage in the High Performance Computing (HPC) Education and Training Community," in Sixth Workshop on Strategies for Enhancing HPC Education and Training (SEHET23), 2023.

Federated Architecture

Training materials from a multitude of sites

QA / Vetting

Federated Catalog

Publish back to sites

Metrics and evaluation

Training materials added to portals via API

Discovery

Workshops Hackathons

Online training

Metadata, Taxonomy, and Ontologies

Common training metadata enables sharing and discovery by defining the standard minimum information that describes training materials.

- 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.

- The "HPC Ontologies and Metadata" Affinity group was created in Sept, 2023.

- 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 mailing list, send an email to hpc-ed@googlegroups.com.