

Cornell University Technical Report

Final Results: National Survey on Education and Training Material Repositories, December 2022

The survey was designed, run, and summarized by the following authors. Survey results were collected from October 10, 2022 to November 20, 2022. 112 surveys returned.

Questions 14, 15, and a few select responses were not included due to personal data.

Katharine Cahill, Ohio Supercomputer Center

David Joiner, Kean University

Scott Lathrop, NCSA and Shodor Education Foundation

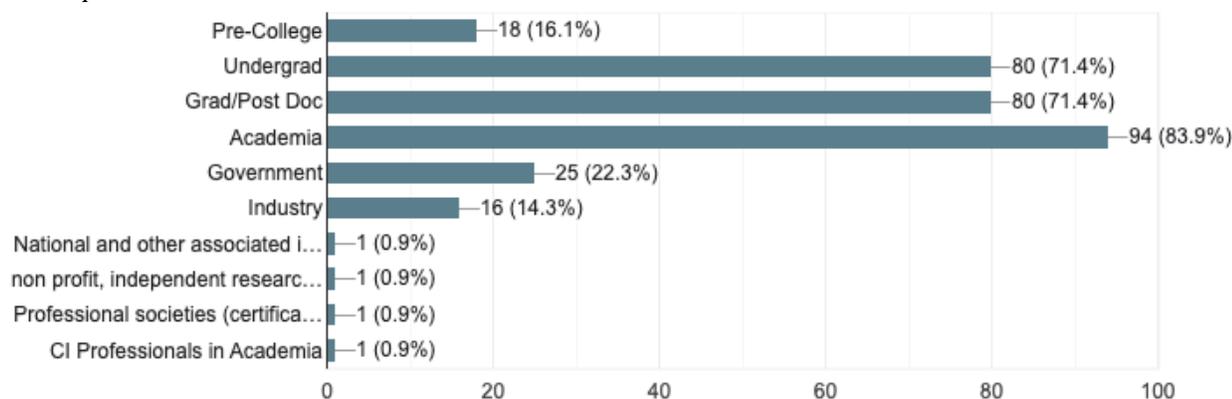
Susan Mehringer, Cornell University Center for Advanced Computing

JP Navarro, Argonne National Laboratory, University of Chicago

Aaron Weeden, Shodor Education Foundation

1. Which communities do you support or participate in? (check all that apply)

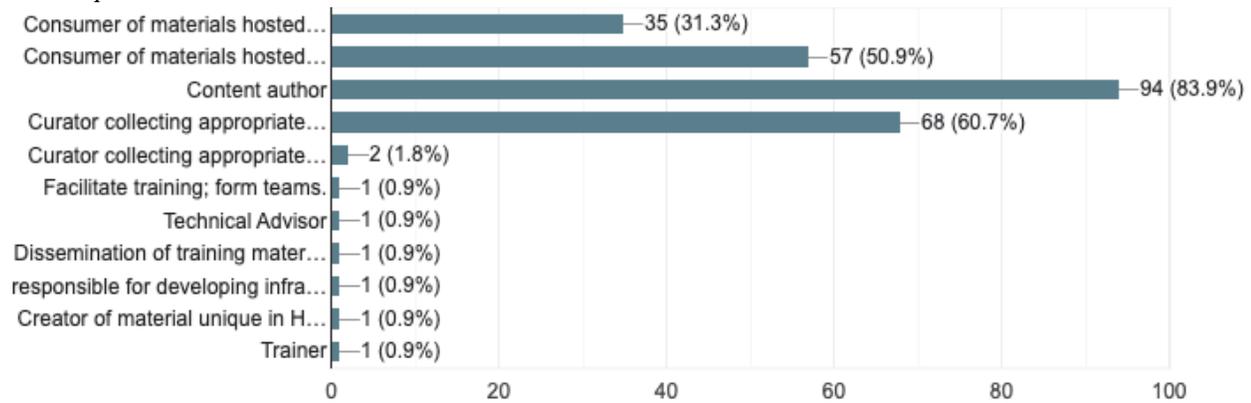
112 responses



94	Academia
80	Undergrad
80	Grad/Post Doc
25	Government
18	Pre-College
16	Industry
1	Other: National and other associated institutes
1	Other: non profit, independent research, research and development
1	Other: Professional societies (certifications based)
1	Other: CI Professionals in Academia

2. When it comes to training and education materials, I consider myself to be (check all that apply):

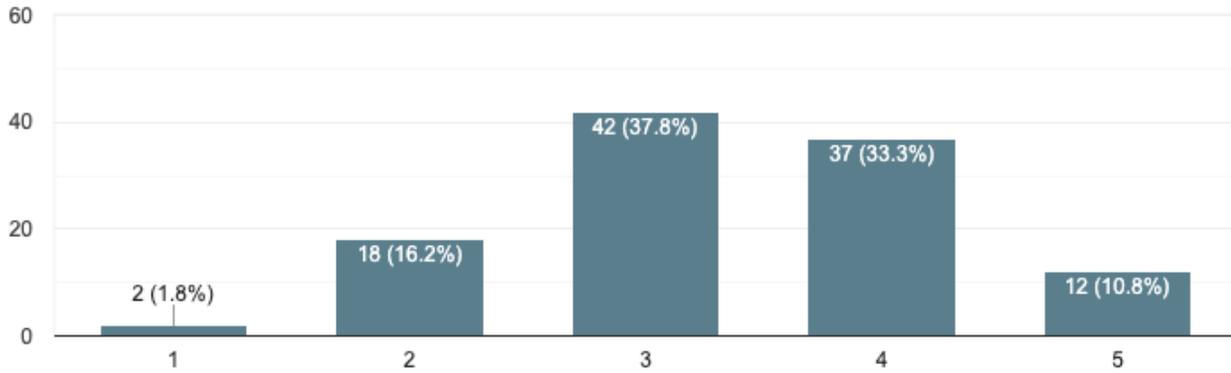
112 responses



94	Content author
68	Curator collecting appropriate materials for my community
57	Consumer of materials hosted by other organizations
35	Consumer of materials hosted by my organization
2	Curator collecting appropriate materials for your community
1	Other: Facilitate training; form teams.
1	Other: Technical Advisor
1	Other: Dissemination of training materials to various communities
1	Other: responsible for developing infrastructure for hosting training material
1	Other: Creator of material unique in HPC (e.g., LLVM and binary instrumentation-based floating-point exception checking of GPUs)
1	Other: Trainer

3. How easy is it for you to find appropriate training and education materials on specific topics, for yourself? (1 = Very difficult, 5 = Very easy)

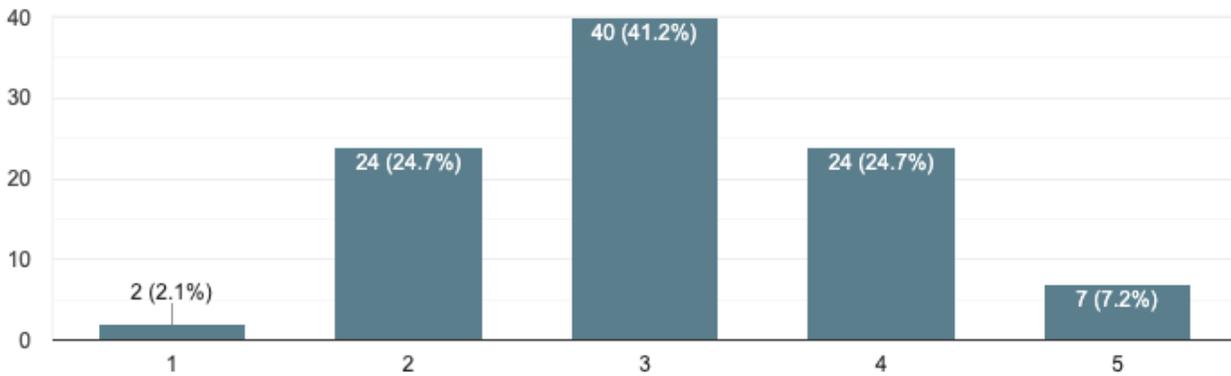
111 responses



42	3
37	4
18	2
12	5
2	1

4. How easy is it for you to find appropriate training and education materials on specific topics, for group(s) you support? (1 = Very difficult, 5 = Very easy)

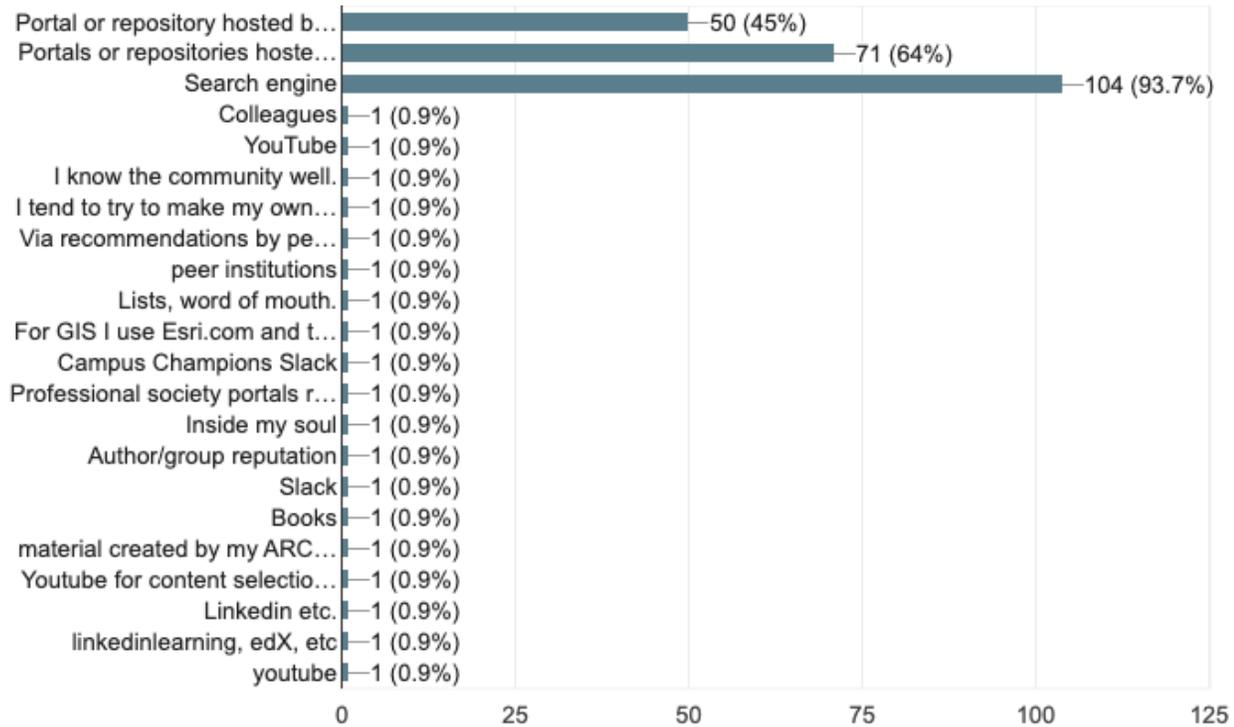
111 responses



40	3
24	4
24	2
7	5
2	1

5. Where do you look for material?

111 responses



104	Search engine
71	Portals or repositories hosted by other organizations
50	Portal or repository hosted by my organization
1	Other: Colleagues
1	Other: YouTube
1	Other: I know the community well.
1	Other: I tend to try to make my own materials
1	Other: Via recommendations by peers from appropriate departments/institutions
1	Other: peer institutions
1	Other: Lists, word of mouth.
1	Other: For GIS I use Esri.com and their online training materials
1	Other: Campus Champions Slack
1	Other: Professional society portals resources pages
1	Other: Inside my soul
1	Other: Author/group reputation
1	Other: Slack
1	Other: Books
1	Other: material created by my ARC staff
1	Other: Youtube for content selection and some amount individual work to curate the information gathered.
1	Other: Linkedin etc.

1	Other: linkedinlearning, edX, etc
1	Other: youtube

6. Which portals, repositories, search engines or other resources do you use or find helpful?

93 responses

Google
Google, Google Scholar
I usually start with a Google search and the follow what pops up.
xsede online training, cornell virtual workshop, NCSA's virtual training, github now hosts quite a few course materials from trainings (e.g., carpentries) to university full courses
Since COVID, there's a ton of stuff on YouTube, but it needs to be better organized. It would be nice to create a community access portal that scrapes sources and lets people rate what they find. If it has a low rating, we can skip over or block it. That is a project I hoped to mobilize, but haven't got around to it. If others agree it would be useful, let's form a project.
google to find github repos
Github, DuckDuckGo, StackOverflow, JuliaBlog, R-bloggers, PyQuants, multiple Meetups, kaggle
3dem and ccpem mailing lists
all helpful
American Accounting Association, Association of Certified Fraud Examiners, ISACA, The Institute of Internal Auditors
Campus Champions Slack
CARCC, Virtual Residency, Campus Data Sci Hub, RCI, Data Librarian Network
Cornell University HPC training, NSF XSEDE training
CSERD
CSinParallel, CDER, Google
discussions in meetings (like CaRCC, TECC)
general purpose search engines
Github, DuckDuckGo, StackOverflow, JuliaBlog, R-bloggers, PyQuants, multiple Meetups, kaggle
Google

Google
Google and Academia website
Google incl. Google Scholar
Google is great for finding materials for self-study. Google scholar, databases in libraries
Google Scholar, manufacturers etc. sites.
Google search, SDSC web site, web sites of other HPC centers, bookmarks I collected over the years google to find github repos
Google, Bing
google, CVW
Google, first and foremost, Reddit, mailing list archives
Google, Google Scholar
Google, Google Scholar
Google, Software Carpentry
Google, University Sites etc.
Google.
HPC University, The Carpentries, Carpentries Incubator, Cornell Virtual Workshop
http://www.hpcuniversity.org/
https://rpedia.stanford.edu/ (our user documentation website); https://www.sherlock.stanford.edu/docs/ (Stanford-wide docs for HPC system) https://software-carpentry.org/ (for general linux; git things)
https://www.esri.com/training/
https://www.eurocc-access.eu/ , Google, https://julialang.org/learning/ , https://events.prace-ri.eu/
I am not aware of any specific search engines or portals for this, hence I just use Google.
I try to find relevant choices (university courses, MOOCs, NCC training databases/ repositories) by generic search engine or personal suggestions, then check them out
I usually start with a Google search and the follow what pops up.
I usually use search engines (Google) to find material that is relevant, often this is in web pages hosted by other HPC centers.
IAC NIST CISA DISA RDA Gateways Data.gov data.world others IEEE
Instructional media
internal wiki, software carpentry, Google etc.

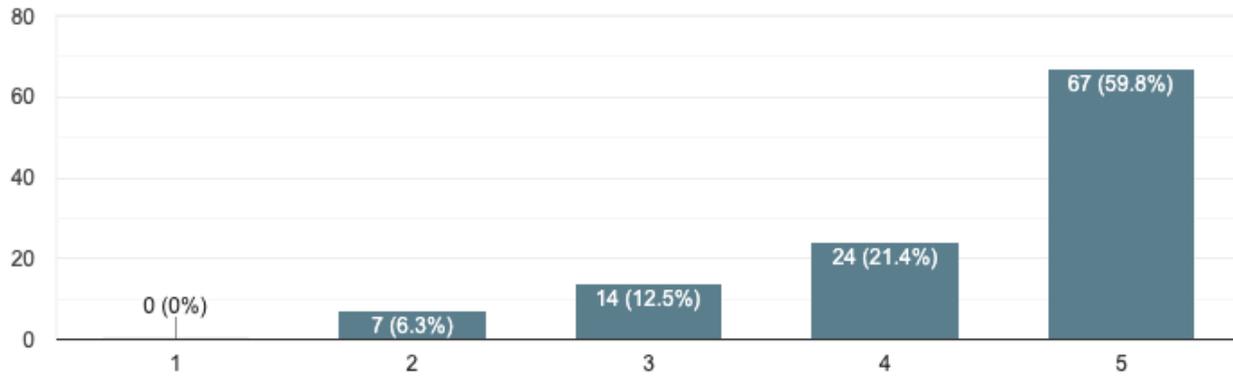
LSU@HPC training sessions, docs.rc.fas.harvard.edu, TACC, Google
MIT Opencourseware, YouTube, repositories of lecture slide from topical summer schools Mostly google
None
Nvidia repository; LinkedIn Learning; Google search
Open source communities, XSEDE/ACCESS portal, and other university programs sites.
Peer institutions (Purdue, SDSC, and others)
Presentations from groups like CaRCC
Prof2Prof, Google
QUBES hub, Google, Bioconductor, Github, and others
schedmd, Harvard FASRC, Cornell CVW, my own GoogleDocs
SDSC HPC User Training, followed by Google search, then look for ORNL, vendors, and papers/publications
Shodor, large super computing centers, old XSEDE materials
Since COVID, there's a ton of stuff on YouTube, but it needs to be better organized. It would be nice to create a community access portal that scrapes sources and lets people rate what they find. If it has a low rating, we can skip over or block it. That is a project I hoped to mobilize, but haven't got around to it. If others agree it would be useful, let's form a project.
Software Carpentry
Stackoverflow, mailing list archives, software carpentry, github, published documentation/user manuals, user guides from universities/national labs.
StackOverFlow, TLDP, official documentation
The Carpentries
The Carpentries, Campus Champions, twitter
The Carpentries, NanoHub, Coursera, Kaggle
Thriftbooks
Too many to list
Totally depends on what I'm looking for at the time and way range of topics is excessively large
Traditional Internet search engines, news for new platforms or programs, Stack Overflow, Individual documentation of specific platforms
Training at other HPC centers
UVM IT Wiki, UVM IT Webpages / Knowledge Base, Google
Virtual Residency, Supercomputing in Plain English, Linux Clusters Institute, XSEDE (now ACCESS)
xsede online training, cornell virtual workshop, NCSA's virthal training, github now hosts quite a few course materials from trainings (e.g., carpentries) to university full courses
XSEDE was good. Campus Champions and CaRCC lists. Lots of slack orgs.
Youtube, GeeksOfGeeks, Git

Training at other HPC centers
Google, Bing
Google incl. Google Scholar
Open source communities, XSEDE/ACCESS portal, and other university programs sites.
Cornell University HPC training, NSF XSEDE training
Instructional media
Prof2Prof, Google
Google, University Sites etc.
internal wiki, software carpentry, Google etc.
Mostly google
The Carpentries, Campus Champions, twitter
Too many to list
HPC University, The Carpentries, Carpentries Incubator, Cornell Virtual Workshop
StackOverFlow, TLDP, official documentation
Specific software tutorials
Online repositories
UCLA IDRE, Google
OLCF, NERSC, and ALCF training/workshop webpages
DReSA is a new Australasian portal that we've created. Sounds like a global movement!
Google, Reddit
Google, HPC communities
Google
Google and then internal repos
linkedinlearning, edX, Google
Youtube
Google, Stack Overflow, (in-house USGS portal)

7. How important is, or would be, having easy access to repositories of training and education materials from multiple organizations to your community?

(1 = Not important, 5 = Very important)

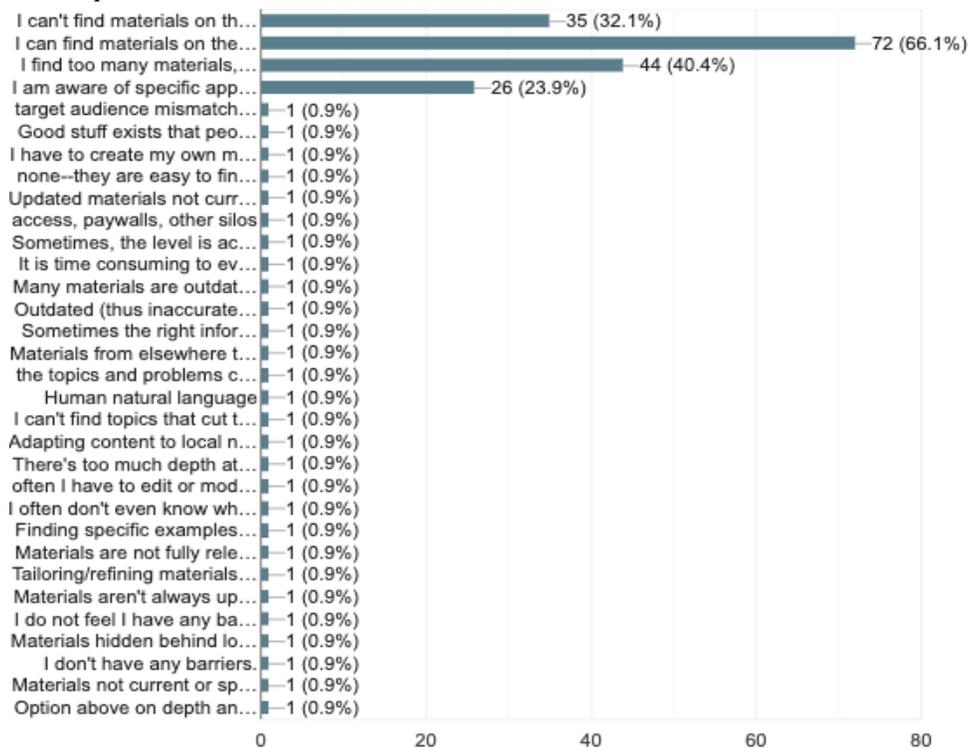
112 responses



67	5
24	4
14	3
7	2
0	1

8. What barriers have you encountered when searching for materials?

109 responses



72	I can find materials on the topic, but not at the depth or level I need
44	I find too many materials, and I can't effectively sort through them all
35	I can't find materials on the topic I need
26	I am aware of specific appropriate materials, but search engines don't list them in the top results
1	Other: target audience mismatch (materials assume engineering background that learners do not have), outdated (this is a huge problem with OERs, they are abandoned without support).
1	Other: Good stuff exists that people don't know about. Poor marketing/communications.
1	Other: I have to create my own materials, or have others create them, because they don't exist.
1	Other: none--they are easy to find, but sometimes not in a reusable format (e.g. markdown, quarto, Libreoffice, etc.)
1	Other: Updated materials not currently available or poorly produced
1	Other: access, paywalls, other silos
1	Other: Sometimes, the level is actually too complex.
1	Other: It is time consuming to evaluate if they materials are of high quality.
1	Other: Many materials are outdated. This is the primary problem I see

1	Other: Outdated (thus inaccurate) materials clogging search results
1	Other: Sometimes the right information seems not to exist (yet)!
1	Other: Materials from elsewhere typically not system-specific enough
1	Other: the topics and problems change too quickly for the time I have to invest in being a grant writer, system administrator, researcher, and research facilitator.
1	Other: Human natural language
1	Other: I can't find topics that cut to the heart of the matter, unlike in, say, ML. Also a regular E-newsletter from those editing/curating such material does not exist.
1	Other: Adapting content to local needs
1	Other: There's too much depth at times. SchedMD's sbatch page, for instance.
1	Other: often I have to edit or modify training materials to make them relevant to the systems I support
1	Other: I often don't even know what to look for
1	Other: Finding specific examples for a specific workflow
1	Other: Materials are not fully relevant to our audience- made for specific centers or systems, or for users with sudo.
1	Other: Tailoring/refining materials suitably for the targeted audiences
1	Other: Materials aren't always up to date or are site specific
1	Other: I do not feel I have any barriers
1	Other: Materials hidden behind logins or paywalls
1	Other: I don't have any barriers.
1	Other: Materials not current or specific enough
1	Other: Option above on depth and level wouldn't select

9. Which barrier, if removed, would be most helpful for finding appropriate materials?

108 responses

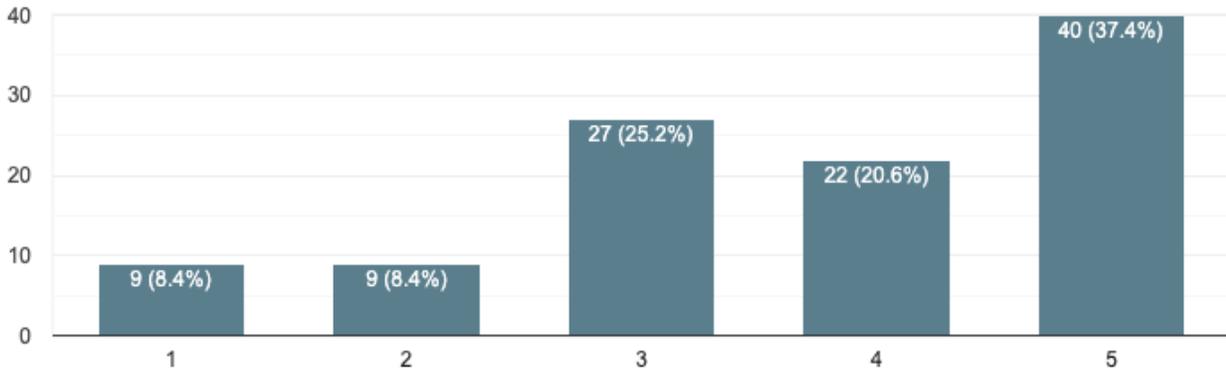


46	I can find materials on the topic, but not at the depth or level I need
29	I find too many materials, and I can't effectively sort through them all
12	I can't find materials on the topic I need
10	I am aware of specific appropriate materials, but search engines don't list them in the top results
1	Other: Again, rating (by our intimate community of practice) what is there would be very useful.
1	Other: All of the above too many barriers and bottlenecks to all
1	Other: as above.
1	Other: I don't have any barriers.
1	Other: in the reward system (mostly time) for HPC staff and researcher faculty
1	Other: Issue regular E-newsletters and have curated updates issued regularly. Organize a community meeting happening once a month
1	Other: Logic, reasoning, processes etc
1	Other: Materials are not fully relevant to our audience- made for specific centers or systems, or for users with sudo.
1	Other: Materials hidden behind logins or paywalls
1	Other: Please see response to Question 8
1	Other: Sharing with communities with similar needs
1	Other: Materials are not fully relevant to our audience- made for specific centers or systems, or for users with sudo.
1	Other: as above.
1	Other: Please see response to Question 8

1	Other: Materials hidden behind logins or paywalls
1	Other: I don't have any barriers.

10. Does your organization want to make it easier for the public to find your training and education materials? (1 = Not at all, 5 = Very much)

107 responses



40	5
27	3
22	4
9	2
9	1

11. What do you consider to be the biggest challenge(s) in sharing your materials?

99 responses

Annotation, indexing and search
Appropriate platform, on-premises CMS system is nearly useless to us
Authors may be hesitant to share, sharing in a format that others can adapt to their site and needs
Awareness that they are available
Central repository I can contribute to that will be recognized as important by my management
Changing institutional priorities / policies
Cluster-specific content that would not work elsewhere
common repository, awareness of other groups' efforts to avoid reinventing the wheel Communicating to the right audience directly
Communication
Concern about sharing the teaching content of my institution, even when I have prepared it myself. I've been paid to carry out the role of creating it, therefore, I feel I should not be sharing it with others for free.
Content needs to be refreshed and formatted, topics better organized
Contribution/acknowledgement
Contribution/acknowledgement
Copyright, trust, and collaboration or lack thereof
cost and staffing
cost of producing and customising to user needs
distributed campus
editing and publishing
Ensuring they meet the needs of those consuming them.
Findability
Finding an audience who needs them
Finding an audience who needs them
Finding the time to clean them up and get them ready to audience outside our institution..
finding: understanding the audience level that the material is created for, finding materials that aren't tightly coupled to a specific center. Sharing: lack of knowledge about how best to share, use SEO so that they are found

finding: understanding the audience level that the material is created for, finding materials that aren't tightly coupled to a specific center. Sharing: lack of knowledge about how best to share, use SEO so that they are found
Formatting them in a way that useful and accessible to others.
Getting the word out that they are available
Give an accurate content description, and choose the right level (introductory, intermediate, or advanced).
Happenstance. The audience is small.
Having good quality materials that are maintained that merit sharing.
Having to maintain general and specific training materials for public and our consumption.
Host web site
Hosting them in an intuitive way for both us and the consumer.
I am not sure. Platform would be my guess, but that is not based on anything.
I write for a specific audience, with probably more assumptions than are helpful
Intellectual Property agreements
it's not in the reward system at all (nearly all administrators have no idea what HPC is, why open source is important, and how the purchases are justified, either strategically or operationally
it's not in the reward system at all (nearly all administrators have no idea what HPC is, why open source is important, and how the purchases are justified, either strategically or operationally
it's very specific to the org I support so does not really apply to other institutions Lack of curated repositories, lack of inter-institutional community
Lack of recognition/incentives from institution.
Lack of time / staff to prepare them for convenient sharing (change format, remove personal / institutional quirks, etc.)
Letting others know that it is available
Letting others know that it is available
Letting people know it exists.
maintaining website - lack of staff
Make it easy to find. Putting it on the internet is not enough but I don't know how to go further Make the material easily findable for the right audience
Making materials which are suitable for synchronous or asynchronous consumption. Much of ours are developed for our environment and wouldn't be useful to others.
Much of ours are developed for our environment and wouldn't be useful to others.

Much of what I work on/train is so new that its barely in production and even then its still far from appropriately mature to not have issues for the variety of individuals I train. Usually a combo of operating system or hardware.
My organization's attitude: per default, all the material I produce belongs to my university.
My own time that is skewed toward research publications and not a recognizable reward structure for doing so in education.
National leaders should make an effort to alert appropriate consumers about these materials.
need a hub for depository
No good common place to reliably publish & list these materials over a long period of time. No place to share them
None, really. We publish them and communicate that to our userbase.
None. I host my content on a publicly accessible webpage.
Not sure how much my materials are overlapping with others and what are my unique contributions.
Nothing
organization
Organized outreach to other organizations.
People finding our content
Plagiarism
Procuring funding to keep developing materials, and then writing reports to the funding agency. Project sustainability and maintenance
Reaching out to the right audiences
Relevance to other institutions - just because it works for us doesn't mean it translates well elsewhere
Search engines not listing them in top results; GitHub repos not being found; our video archives not being found
Showing up in web searches. Simplicity of contribution
Sometimes it is hard to find relevant application examples. For sharing materials, if it not a top google hit, it is not discovered by the community.
Standardized Catalog for others to use.
Technical expertise, time availability, career gains
the approval
The biggest challenges would be making the materials known to others and accessible. the path/procedure for sharing is not clear.
The proliferation of search engines that don't often prioritize relevant content. In the same vein, and sadly, EBSCO and ProQuest databases are guilty of this too!

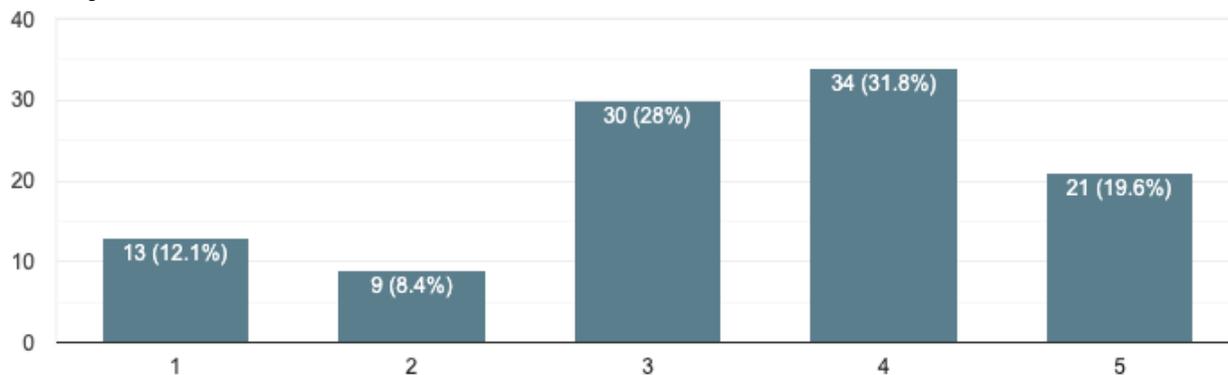
the time investment to package and share them, lack of support to keep them up-to-date (once they are shared they linger), advertising them so others actually make use of them.
the time investment to package and share them, lack of support to keep them up-to-date (once they are shared they linger), advertising them so others actually make use of them.
The time it takes
They are tailored to our systems vs being general or broad
time to put in the effort
time to refine my material into classes, time to make good videos or tutorials, and time
time, access, funding
Too busy.
Undiversified teaching method
University wanting to keep material as IP.
Users of my materials might not understand the teaching approach.
Very site specific
Visibility, and finding the right technical solution to make it useful to my community
Willingness to try new approaches. Resources for proof-of-concept models.
Working in a classified environment.
Make it easy to find. Putting it on the internet is not enough but I don't know how to go further
My organization's attitude: per default, all the material I produce belongs to my university.
Procuring funding to keep developing materials, and then writing reports to the funding agency.
Project sustainability and maintenance
Having good quality materials that are maintained that merit sharing.
Simplicity of contribution
Letting people know it exists.
None, really. We publish them and communicate that to our userbase.
Happenstance. The audience is small.
Ensuring they meet the needs of those consuming them.
National leaders should make an effort to alert appropriate consumers about these materials.

Copyright, trust, and collaboration or lack thereof
People finding our content
Authors may be hesitant to share, sharing in a format that others can adapt to their site and needs
Organized outreach to other organizations.
Finding the time to clean them up and get them ready to audience outside our institution..
Showing up in web searches.
Concern about sharing the teaching content of my institution, even when I have prepared it myself. I've been paid to carry out the role of creating it, therefore, I feel I should not be sharing it with others for free.
Willingness to try new approaches. Resources for proof-of-concept models.
the path/procedure for sharing is not clear.
maintaining website - lack of staff
The biggest challenges would be making the materials known to others and accessible.
Undiversified teaching method
No place to share them
Not sure how much my materials are overlapping with others and what are my unique contributions.
Getting the word out that they are available
common repository, awareness of other groups' efforts to avoid reinventing the wheel
Findability
Too busy.
Intellectual Property agreements
organization
The proliferation of search engines that don't often prioritize relevant content. In the same vein, and sadly, EBSCO and ProQuest databases are guilty of this too!
Search engines not listing them in top results; GitHub repos not being found; our video archives not being found
No good common place to reliably publish & list these materials over a long period of time.
University wanting to keep material as IP.
Awareness
Where to put it so that it can be found and used

Materials are very organization/usage-specific
Time and support to refine materials for public consumption
time
Raising awareness
Organizing them
Knowing how to start and at what level of intricacy
Time to do it
Find the right place for the right audience
I am new to the organization and am unsure why materials are currently accessible in-house only.

12. 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?

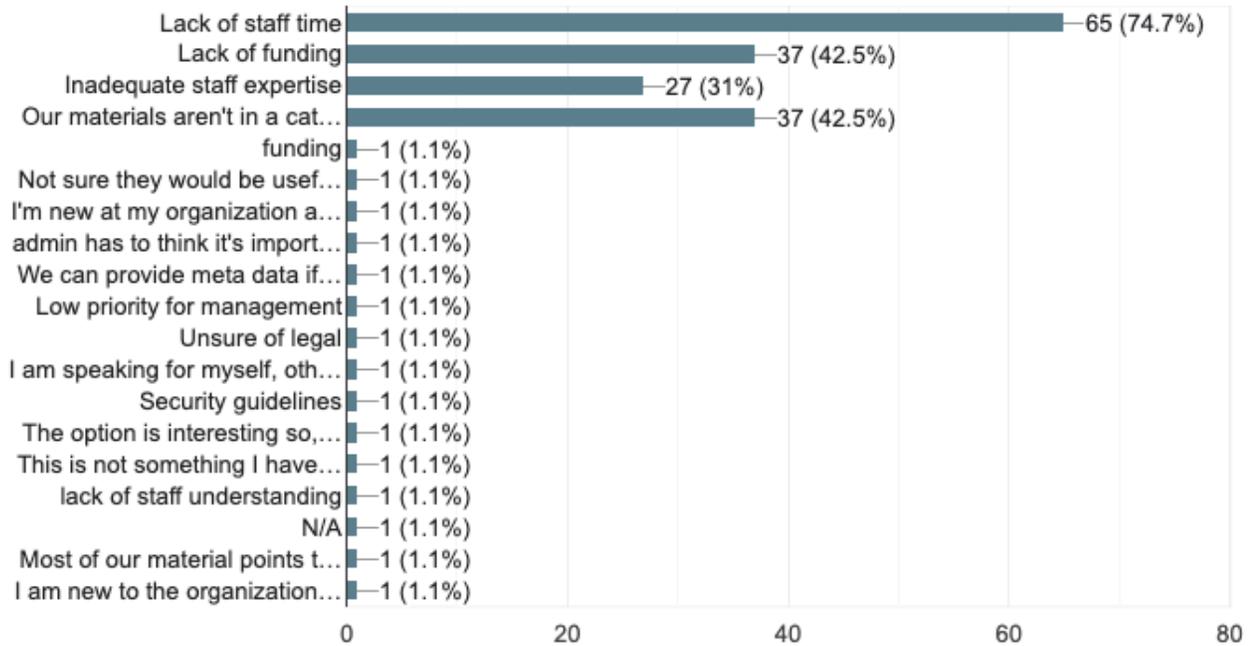
107 responses



34	4
30	3
21	5
13	1
9	2

13. If your organization is not willing and able to provide metadata about your materials in a standard format, what are your roadblocks?

87 responses



65	Lack of staff time
37	Lack of funding
37	Our materials aren't in a catalog
27	Inadequate staff expertise
1	Other: funding
1	Other: Not sure they would be useful or meaningful. We use Software Carpentry training - that has been socialized by [our institution], in general.
1	Other: I'm new at my organization and am not sure if I can share materials. Previous institutions, no roadblocks and materials are shared on github
1	Other: admin has to think it's important (it would help if Educause thought it was important too--and not just parrot what the EdTech and commercial vendors say)
1	Other: We can provide meta data if necessary
1	Other: Low priority for management
1	Other: Unsure of legal
1	Other: I am speaking for myself, others both produce more material and may have more time (=funding) to support such efforts.
1	Other: Security guidelines

1	Other: The option is interesting so, yet to explore the option with my org.
1	Other: This is not something I have explored.
1	Other: lack of staff understanding
1	Other: N/A
1	Other: Most of our material points to other outside material
1	Other: I am new to the organization and am unsure why materials are currently accessible in-house only.