



PEARC 2026 BoF

Stronger Together

Collaboration & Coordination between HPC Training Collections

Organizers: Susan Mehringer, Katharine Cahill, Christopher Cameron, David Joiner,
Charlie Dey, Brian Guilfoos, Rich Knepper, JP Navarro

Session Format & Goals

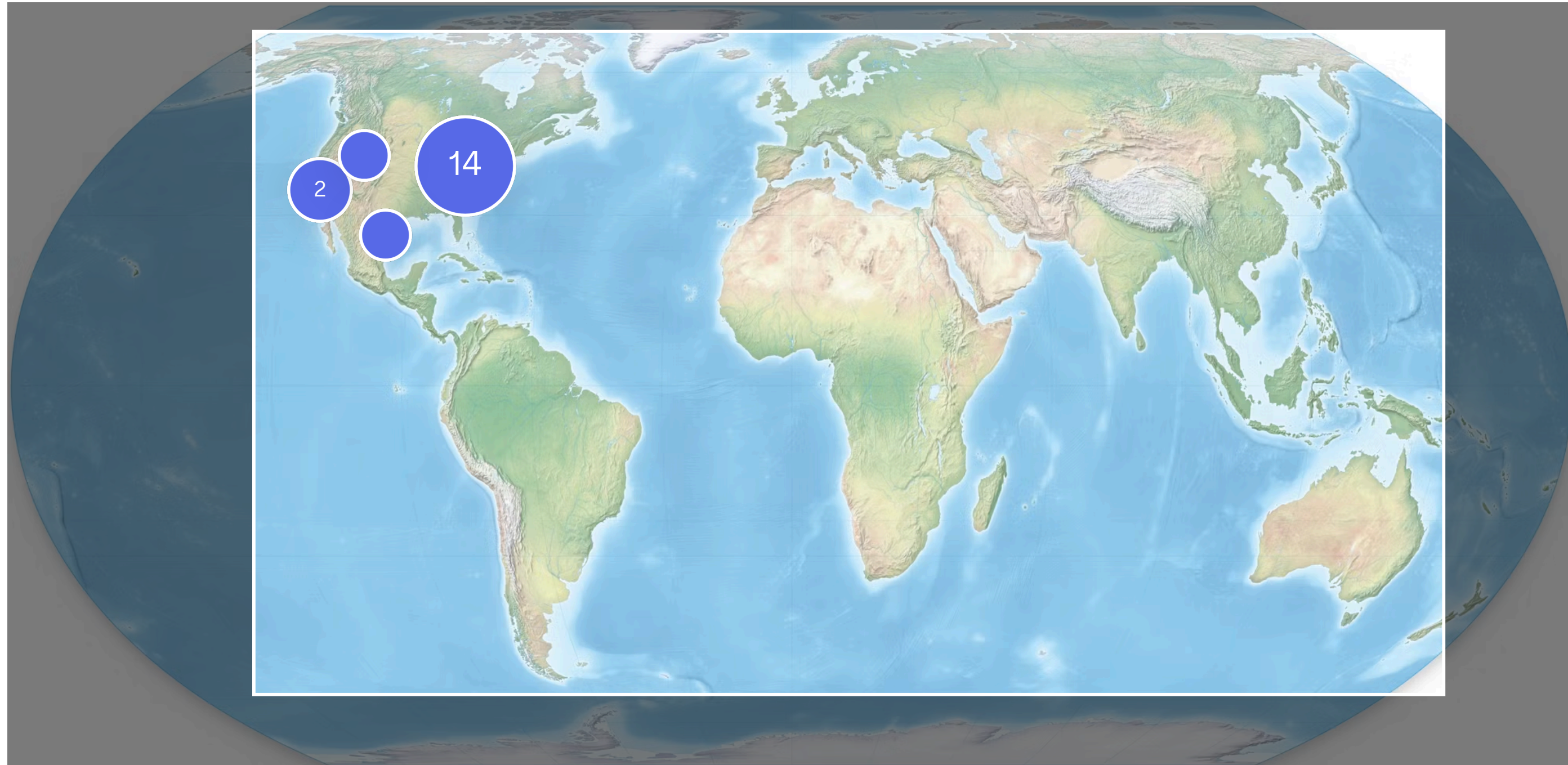
Coordinating the HPC Catalog Ecosystem

The HPC landscape spans centers, labs, universities, and corporations. By learning from each other, we can:

- Improve our individual training collections.
- Avoid work redundancy on core topics.
- Synthesize our catalogs for better discovery.



Drop a pin where you live:



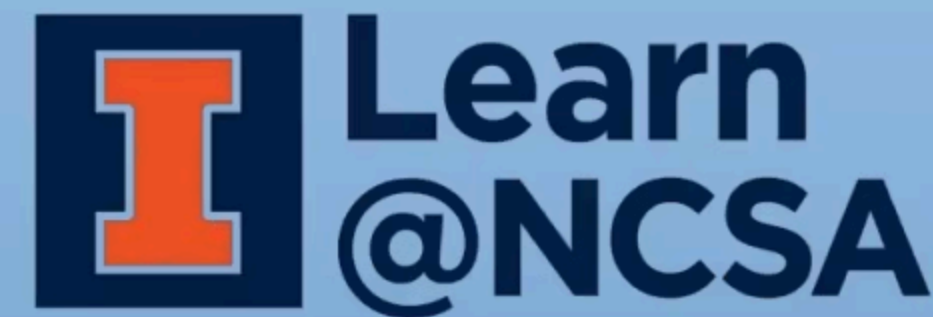
Part 1: Overview & Survey

LEAD: BRIAN

Mapping the Catalog Landscape

We surveyed collection providers to try and identify patterns and common challenges.

A low N limits the conclusions we can draw, however there are some observations



Mentimeter



Part 1: Overview & Survey

LEAD: BRIAN

Observations

We surveyed collection providers to try and identify patterns and common challenges.

Strong interest in collaboration, especially in sharing materials and co-developing materials.

- General HPC skills are universal, but collections otherwise provide limited specializations
- Most are not currently backed by databases
- No survey respondents reported storing materials intended to be reused in other training contexts (e.g., lesson plans for instructors to leverage)
- Agentic search leads AI use in collections.



Part 2: Consistent Metadata

LEAD: CHRISTOPHER

Consistent metadata enables universal search crawlers and avoids duplicated discovery efforts.

Standardizing Subsets

What core fields must every training resource map to for federation?

Required: what it's called; where it is; what kind of link is it; what language; does it cost; who provides it; index name for Globus

Recommended: what it's about; who made it; can I share it; what to search for; what kind of material; expertise needed; who it's for; how long it lasts

Tools & Automation

How do we tag resources seamlessly without manual fatigue?

Manual entry: spreadsheet tool — direct entry & bulk upload

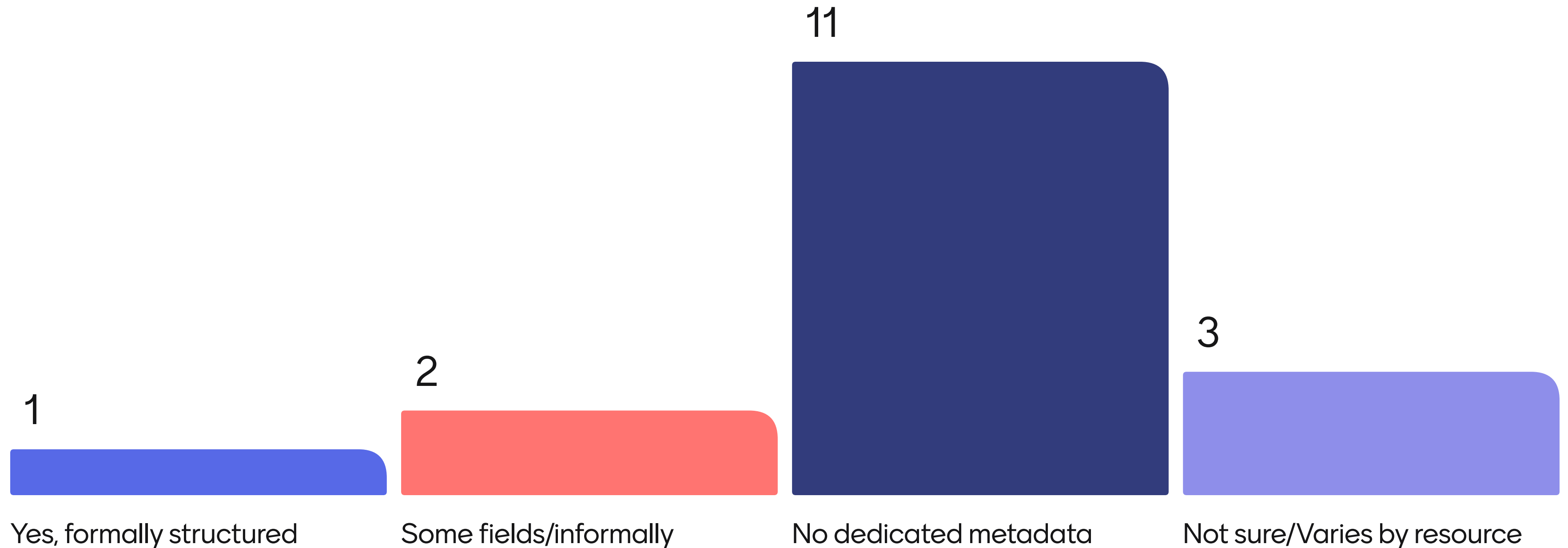
Scripted upload: maintainer scripts — format metadata as JSON

Link discovery: crawlers (e.g. Crawl4AI) — crawl sites, output markdown link lists

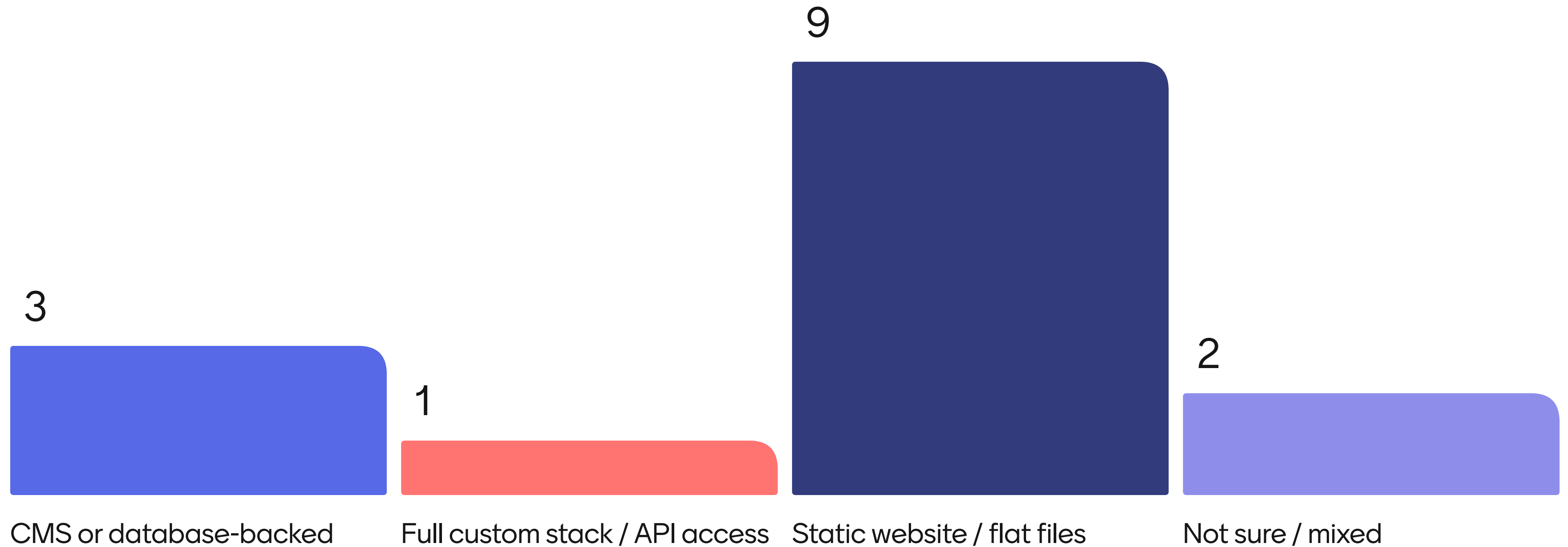
AI-assisted review: current: LLMs review human-entered metadata, suggest fixes; future: LLMs draft metadata from database/crawl data



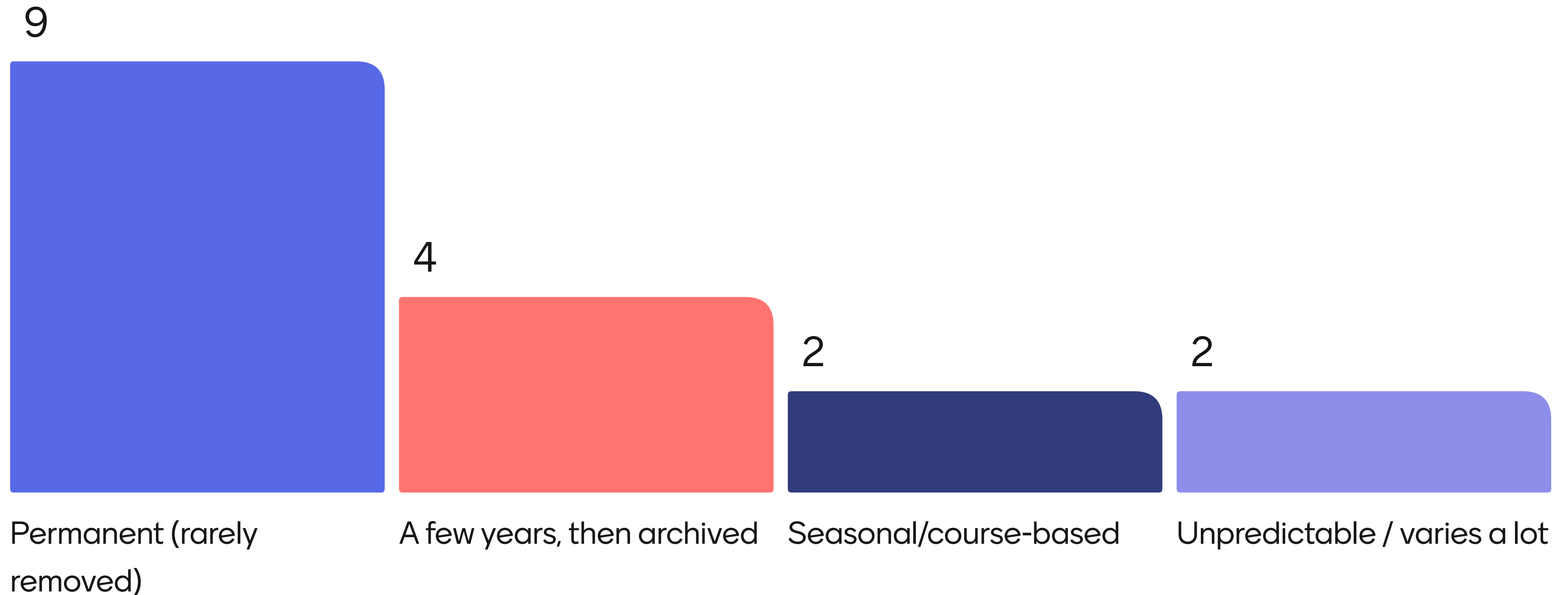
Do your content items already have metadata aligned with a standard like RDA or Dublin Core?



How is your training content currently stored?



How long do your resources typically stay live/available?



Part 3: Referencing Outside

LEAD: KATE

Direct Linking vs Link Rot

Conceptual basics (like "Intro to SSH" or "Git Basics") are universally reusable, but sites hesitate to link externally.

- > Do you actively route users to other centers' tutorials?
- > What is your backup plan when target links break?



What has been your experience with linking to other center's tutorials?

Rarely, and for general intro materials.

Worry about link rot means me mostly link to live sessions from PSC

Concern about long-term availability.

been pretty good/painless, but we are also selective with what we link. We also do not have any automated checking of links, which would be useful to add...

More intro/basic materials are pulled externally, and links need to be maintained

We primarily link to external sites for extended information and tutorials we don't cover. The challenge is now ensuring all linked sites are accessible and meet current accessibility guidelines

Regularly link to other's materials when talking about their resources.

I happily link to other people's training materials and prefer to use that instead of creating my own because I'm super lazy. Shout out to Git Immersion and the University of Helsinki!

What has been your experience with linking to other center's tutorials?

When appropriate linking to high quality external sources has worked well.

Our institution requires a "not affiliated with or endorsed by" disclaimer to be included for external links.

We will look at knowledge bases for other centers and see how we can improve our own. we rarely link pages to other centers for tutorials

Yes. We recommend a number of different resources from other institutions and centers. We also have a LinkedIn Learning path of recommend courses (for general, into courses)

Good, but users don't always follow the links from outside resources.

Part 4: Material Collaboration



LEAD: CHARLIE

Instead of 50 centers separately writing "Intro to Linux" slides,
how do we co-develop reusable modules?



Current Efforts

Collaborative writing via GitHub or shared repositories.

What collaborative environments
do you currently utilize?

How many people outside of your
organization do you collaborate
with?

Is it an "official" capacity? or more
casual in nature?



What collaborative environments do you currently utilize?

Slack

github

Gitlab

google docs

Discord

Git, Slack

Github

GitHub (but only within
our team)

What collaborative environments do you currently utilize?

Slack, email, github, overleaf

internal - Teams.
external (users) - Teams, Slack.

Sphinx + GitHub Pages + Slack

We use git for training catalog development.

Slack, GitHub, Gdrive, Not ebookLM

GitHub

github pages, software carpentries Internal is google chat

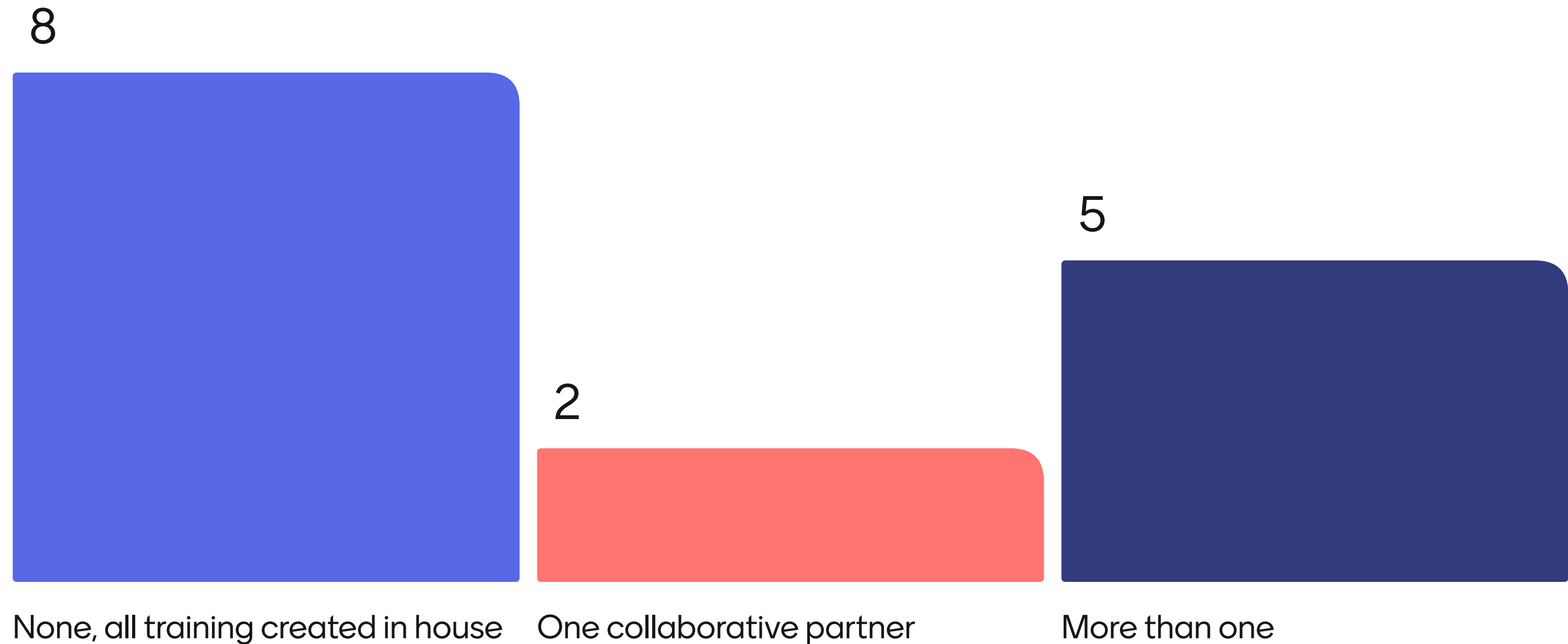
Github LearnR - share course materials as an R package Liascript - in progress attempt to share markdown tutorials

What collaborative environments do you currently utilize?

not me but others use
Moodle somehow?

PowerPoint/SharePoint
Online

How many people outside of your organization do you collaborate with?



If you have a training material collaboration, is it official or casual?

Not structured, more casual collaboration

1.



Official collaboration: project or partnership

2.



Part 4: Material Collaboration



LEAD: CHARLIE

Instead of 50 centers separately writing "Intro to Linux" slides,
how do we co-develop reusable modules?



Preparing for the future?

Emerging Topics

What topics are you developing training on right now?

What topics do you have planned?

How do you keep up?



What topics are you developing training on right now?

DevOps

Skypilot

Good Clinical Practice

AI. Redeveloping basic
HPC skills content

GPUs. Specific
resources. AI.

Basic intro on HPC,
Jupyter/big data, ML,
Neural net, crypto, parallel
programming (MPI)

Practical use of AI
agents

SDLC/EPLC

What topics are you developing training on right now?

Hands-on exercises to help beginning users get practice with running jobs

Train-the-trainer pedagogy

AI Data Readiness AI workflow optimization
Asynchronous training

Using the resources efficiently, safely using AI on clusters

AI agents for science

Storage research, permission and directory structure

Walkthroughs of HPC environments

Reworking intro to hpc materials

What topics are you developing training on right now?

System Development Life Cycle (SDLC)/Enterprise Performance Life Cycle (EPLC)

Agile/Scrum/Kanban

Shorter form content of our existing training materials. Novice users don't always want to watch our 2hr+ recorded session

Infrastructure as Code

What topics are you planning/considering to develop in the future?

good hpc practices

How to write hands-on lessons

When to use HPC on prem vs. in the cloud

How to properly use AI agent on the HPC systems

How to use or not use AI agents on the cluster

Training the trainer

Research Computing conceptual fundamentals. (The why, not how)

Institutional policy-compliant AI usage

What topics are you planning/considering to develop in the future?

Guidelines for AI agent
use on our systems

estimating resource
requirements for hpc
jobs

version control for
environments/job scripts

Intermediate HPC as a
microlearning series

parallelization

Proper use of AI -
breadth view to deep
dive

How do you keep up?

with difficulty

Barely

I don't. No time to even maintain what we already have!

Google Cloud Skills Boost

Service Now Tickets

Tech contents on YouTube

Reading other research computing centers' documentation

Rely on SMEs on specific topics to inform us on changes/updates

How do you keep up?

Driven by client demand

news.ycombinator.com
(but it can be janky)

GitHub trending repos

conference, webinars,
papers

Tech Blogs, news
articles, ...

PEARC

awesome lists

TechEX/ACAMP

How do you keep up?

InCommon IAM Online

Part 5: Communications & Outreach



LEAD: SUSAN



What information to include?

- Title/Description/Outcome
- Format
- Target Audience
- Date/Time
- Requirements/Cost/Prereqs



Who are we trying to reach?

- Each other
 - for collaboration
- Local or wider audience
- Current or prospective users

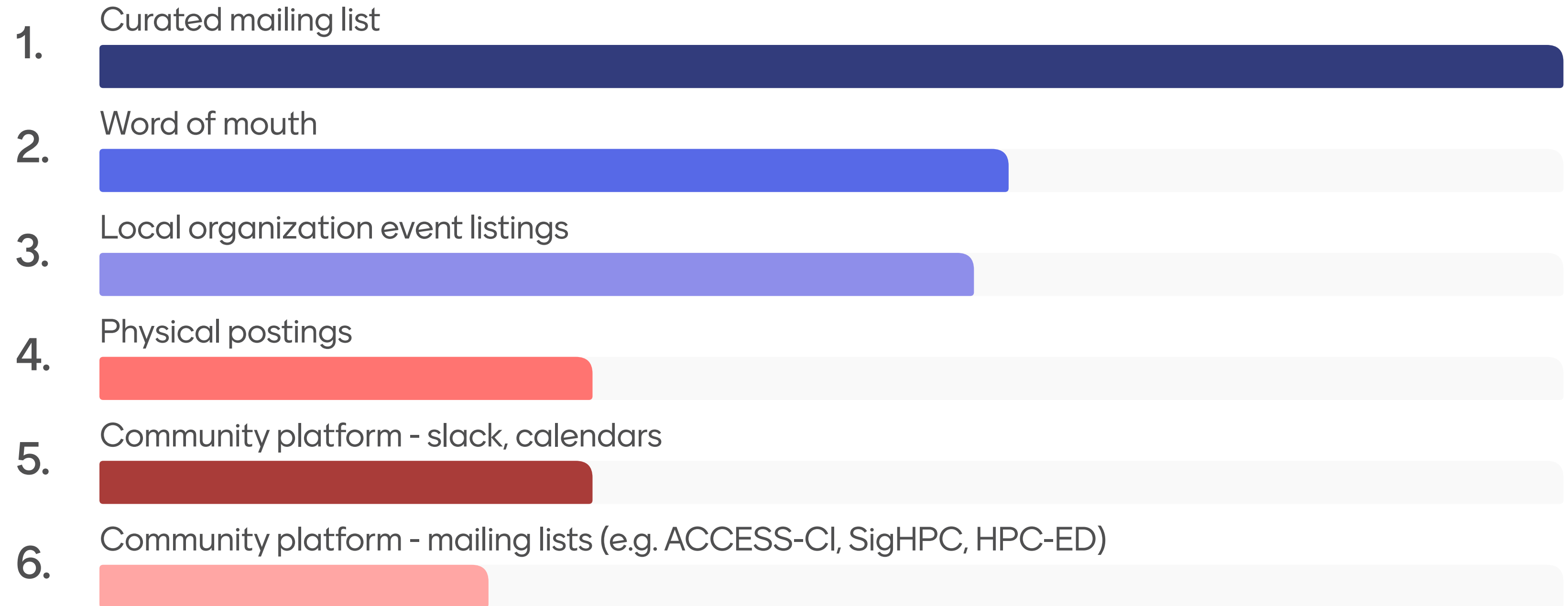


How can we reach out?

- Word of mouth
- Physical postings
- Curated mail list
- Local organization event listings
- Community platform
 - General posting
- Community platform
 - Curated events/materials listing



How do you currently communicate events/offerings?



What is currently your biggest outreach problem?

Reaching prospective users

attendance

Unsure if the audience is right

Translating how I can help them into terms they'll understand

The people who need to the info don't know they need it.

Getting people to read enough to know what's in it for them

lack of attention

Reaching prospective users who are not aware of the research computing group

What is currently your biggest outreach problem?

Finding potential users

Connecting with under
represent groups -
academic majors /
disciplines

What would make us stronger together?





Join the Discussion!

Let's align our metadata, cross-reference resources, and build a unified catalog.

Questions? Comments? Rants? Raves?

 **Stronger Together: PEARC 2026 BoF**

