

Instructions: The following document is a template for notes/questions during the breakout session. Please note that the top 3-5 findings and 3-5 recommendations (at the end) should be summarized and presented back at the main session after the breakouts.

What are the opportunities for collaboration between LFs and other Large CI projects?

List of participants:

Ilya Baldin, Brian Bockelman, Peter Couvares, Brian Glendenning, Ben Hauger, John Haverlack, Frank Wuerthwein, Angela Murillo, John Towns, Devin Bougie, Chris Romsos, Charles Vardeman, Sanjay Padhi, Jane Wyngaard, Sharon Broude Geva, Rachel Adams, Cheryl Ann Blain, Surajit Chaudhuri

Problem statement:

Large facilities - and large CI projects - need large vertical stacks to provide end-to-end solutions to their communities. There is always a danger in “vertical stacks” becoming “vertical silos”, causing the facilities and projects to miss opportunities to collaborate.

How can we create opportunities, make them “discoverable”, and mutually beneficial to both parties?

Notes:

- Not bipartite - both “LF to LF” and “LF to CI” and “CI to CI”
- It’s clear the “top” of the stack is the domain science. How deep is it?
- Comment on the silo metaphor: Farmers will tell you silos are a good thing - don’t necessarily need to break them!
 - Rather, can we connect silos? Get these to talk to each other. Let silos evolve. Can be broader / narrower – but can they talk to each other?
- Needs to have a dedication of time to look at other facilities. Communicate with them.
- What are the risks in “breaking the silo”? Delegate the control to others... but the facilities are fairly risk-averse by design.
 - Can we avoid blaming the LF when the risk is larger than the benefit?
 - Actual risks vs perceived risks (same with opportunities).
 - What makes TrustedCI / cybersecurity a place where the LF are more accepting of outside help?
 - Maybe this is only perceived?
 - Maybe this is more unique to the Open Science setup?
 - Seen as not being here to judge but here to help.
-
-

Current approach:

- CI-CoE pilot
- What are examples of successful collaborations between LF and CI?
 - LIGO & XSEDE: Work together to help optimize / review codes.
 - Catalyst really was at the NSF-level.

- Fear of being judged. Took some time to understand / work-through cultural differences. Build trust.
 - Core application sped up by 4-5; afterward, LIGO was a total of 10x faster
 - LIGO & HTCondor team: partnership spanning many years, based on past history. More willing to take risk.
 - HTCondor has similar stories with NRAO and CMS.
 - Difference between using a capability and bringing software into a software stack.
 - Can reduce the risk if you build the software in the right way.
- What are the significant abstractions that we can share? If the commonality is only “we use turing complete machines”, then we don’t have much to talk to.
 - The collaboration is broader than just software - think about staff expertise, for example.
- FKW: What if there was a “pot of gold” that was only accessible for projects between >2 LFs + CIs?
 - What’s the difference between this and CSSI?
 - FKW: think less about full-scale proposals and more like an XRAC allocation. Something that is time-limited and gets access to expertise.
 - Ewa: This will allow us to see commonalities and foster collaboration.
 - There’s a lot more than writing proposals - where individuals think it is important (training, diversity), people move forward regardless of funding.
 - Knowledge can go both directions (LF->CI and CI->LF).
 - Community is important and doesn’t cost money.
 - What’s the difference versus ECSS? FKW: Requirements for multiple LFs and to bring their own effort to the table. Maybe more of a focus on CI and avoid the domain science.
- Collaboration is where we want to get to, but just sharing is a great place to start!
 - Example: There’s an enormous amount of training in XSEDE that could be leveraged. Over half of the trainees in XSEDE have no allocations.
 - Additional pieces can be built on top of that. For the training example, perhaps XSEDE is just missing a particular module for a community?
 - Some amount of this is awareness.
 - Other examples: OSG School, which focuses on dHTC techniques. Carpentry, which focuses on other base skills.
- We can take a look at base layers (best example, WWW) to help collaborate.
 - Other examples: schema.org.
 - What does industry do? What can we leverage - versus where we are indeed different.
 - Example: IPFS
- Is there an idea of a “CI Champion”?
 - Dedicated contact, lower level (not the PI).
 - XSEDE Campus Champions are not necessarily limited to XSEDE.
- How many cases are there where LFs use infrastructure in a dependable way?
 - There are some common technologies where the CIs are sharing - HTCondor. Think of other areas such as resource provisioning, data management, AAI.
 - But what are the mechanisms to do this? *Incentives...*
 - Collaborations can do things that they couldn’t do independently.
 - What are the obstacles?
 - LFs that have sufficient CI internally.
 - Risk-management (creating undesirable dependencies). Mismatch in time horizons between LFs and CI projects

- There are disincentives - socially, some of us like to have bigger projects.
 - Inefficiency of costs.
 - Historically, there wasn't a lot of pressure to reuse NSF CI investments. Is this turning a corner?
- Are LFs and CIs seen as peers in this workshop?
 - There are opportunities for bidirectional learning.
- Given the problems with workforce, we have issues building large enough teams (both in terms of quantity and quality).
 - There is a risk in having someone with the wrong background developing the software - as opposed to outsourcing this.
 - Here, community makes a large difference. Many examples here, both at the management and at the technical level.
- Are we taking a short-term extrapolation from today to the near-term future?
 - Likely we will be able to buy all our computing from an external provider -- but we still get a finite budget.
 - We should be prepared to take advantage of this.
 - But how does this affect collaboration between CI & LFs?
 - How do we make sure that CI & LFs build on top of common industry investments? What are the appropriate common industry investments?
Example: should we be writing query optimizers for databases? Maybe not...
 - This goes back into risk management again on the LF side.
 - Focus should be on going up the layers. What is hard today is maybe easy tomorrow?
 - Partnership is longer term than the specific technology.
- Requires leadership within the facilities to be outward-looking.

Current challenges and opportunities:

- How do we describe services that CIs offer?
 - Is there a set of principles for CI services equivalent to [FAIR](#) (findable, accessible, interoperable and reusable)?
- How do we perform "matchmaking" between CI and LFs?
- How do we fund the effort to transition existing Large Facilities CI to more common stacks?

Short term vision:

- There's a lot of simple sharing where we can derive benefits.
 - Some steps to improve the discoverability of services.
- Guidance on open source licenses and good software "best practices".
 - Licenses can be outside of a software engineer's area of expertise.
- Architectures on what has been done in the past - has this been shared with the community?

Long term vision:

- Co-development in areas of common interest in LFs and CIs.
 - Supporting latest technologies (e.g. FPGAs).
 - Biomedical Hub / NIH STRIDES / All of Us
- How do we differentiate between Python and CORBA? One worked out well but the other different.
 - Always be prepared to move - for example, for data, the minute you put data in, have a plan on how to get it out.
 - Have a mechanism to evolve services, especially when you do the initial adoption.

Top 3-5 findings (to be presented back at main session):

1. Why do LFs and CIs collaborate?
 - a. Economies of scale.
 - b. Continuity.
 - c. Ability to deliver on something you can't do on your own. Helps with the workforce quality problem.
2. Barriers to collaboration:
 - a. Risk management: Loss of control. Unaligned time scales.
 - b. Inefficiencies of cost.
 - c. Physical / fiscal barriers (e.g., on a ship, at the south pole).
3. Existing models of collaboration:
 - a. "Shotgun marriage": Thou shalt collaborate now, per the funding agency.
 - b. Long-term partnerships: Teams working together across many projects; works best when both sides have vested interests.
 - c. Education, workforce development, training, ECSS.

Top 3-5 recommendations (to be presented back at main session):

1. Create opportunities for discovery:
 - a. E.g., catalog of existing solutions, share what we do at the architectural level (using a common template)
 - b. Sharing of available services - esp. training resources.
 - c. Following a common template.
2. Help provide a mechanism to "matchmake" between LFs and CIs and understand when the partnership works for both.
3. Provide a new mechanism to incentivize collaboration: Have a team dedicated to working on short-term projects involving multiple LFs (and requiring LFs to provide their own effort).

Some questions to ask

Problem statement:

All LFs are snowflakes - each one different from the next. Different user communities, different goals, different time horizons.

Can LFs collaborate with each other and Large CI projects to leverage funding and efforts? Share resources or services?

Are LFs so vertically integrated that it makes it difficult for them to collaborate? What must be changed?

What are other obstacles for these collaborations?

Current approach:

What are some of the current examples of fruitful collaborations between LFs and large CIs?

What attributes or circumstances made them a success?

What were the benefits of those collaborations? To LFs? To large CIs? To users?

Current challenges and opportunities:

What are some of the challenges in building collaborations?

Can they be categorized (electro-political, technological, budgetary)?

Short-term vision:

What can be changed to start the process of fostering collaborations?

- Workshops
- People-to-people contacts
- NSF using money as a stick/carrot
- Other external facilitation

Long-term vision:

What are the necessary attributes of a successful collaboration between LFs? Between LFs and Large CIs? From their own perspectives? From NSF perspective? From user perspective?

What types of mechanisms can be engaged in order to make those collaborations more common? Ad-hoc vs. structured approach?

Incentives vs. disincentives of collaborations:

For: continuity, being able to deliver something you can't do on your own, budget

Against: additional risks, loss of control, inefficiency of cost, mismatch in time horizons between CI projects and LFs