

# Cyberinfrastructure: challenges & path forward

By Patrick Brady

Cyberinfrastructure for Major Facilities: Challenges and a Path Forward

March 2, 2022

CI Compass Cyberinfrastructure for NSF Major Facilities Workshop

# Cyberinfrastructure: challenges & path forward

Patrick Brady  
University of Wisconsin-Milwaukee  
LIGO Scientific Collaboration, LIGO Data Grid & SCiMMA

# One CI problem I would like to solve

- Sustainability
  - Infrastructure is defined as “the basic physical and organizational structures and facilities needed for the operation of a society or enterprise.”
  - To me, research cyberinfrastructure is the combination of physical hardware, software, services, and organizational structures needed to undertake and complete research.
- Cyberinfrastructure
  - Needs to be continuously operated, maintained, and renewed to remain competitive.
- At present
  - Large facilities often only have sufficient funding to support minimal CI for the research they support
  - Research CI to support research in the gaps between big facilities limited to small projects that are siloed, have short life cycles, and have no **real** sustainability plan beyond submitting winning another grant typically 2 years after the project has started.
    - We were recently told that planning on another NSF grant is not a path to sustainability
  - Current economic conditions make retention of staff difficult

# Using shared services from broader CI ecosystem

- The biggest challenge is person-power:
  - Needs people who understand both the research requirements and the CI well enough to identify useful services, match requirements to them, and facilitate conversations about adoption.
- Experience in LIGO(-Virgo-KAGRA)
  - HTCondor was our first and is our longest successful engagement. The mature parts were easily adopted by users making it easier to identify high-value feature requests.
  - Open Science Grid is a long-term (multi-decade) engagement that was extremely difficult in the early years. Immature technologies, and the usual problem of multiple solutions to the same problem, in the early years made it extremely difficult for LIGO to muster sufficient effort to make the engagement successful.
    - Nevertheless, maturation of the tech, changes in approach (on both parts), and modest increase in effort from LIGO has made this collaboration successful for LIGO.
  - Authentication and authorization ....

# Collaborations and/or sharing best practices for CI?

- **Biggest technical challenge**
  - Avoiding the reinvention of solutions multiple times. Typically not done by the CI staff in projects, but by users who think they can do it better.
    - Sometimes they do! But long-term support and operations then must transition somehow to another project or facility
  - Training in best practices for software development, uniform adoption of gitlab, training on CI/CD methodologies
- **A costly challenge**
  - Every NSF CI project that involves more than a few people requires tools like Slack/Mattermost, GitHub/GitLab, Google Workspace, etc.
  - For smaller projects, the solution is to piece together the free personal versions, although some providers have good licensing for research organizations who don't need SSO or other features.
  - Can we solve this problem together?

# Balancing adoption of new tech with current ops?

- Delicate balance
  - Security considerations for end-of-life software, services, etc tend to drive retirement because there is insufficient effort to strategically manage the full portfolio of CI.
  - Migration away from old services or approaches is extremely difficult
    - Researchers are getting work done and don't want the pain of porting to a new system
    - Too often, the same individuals are trying to deploy a new technology, support the old one, and bring the users along.
- The biggest challenge is also an opportunity:
  - Smart people have lots of ideas for new CI that could be useful or better than existing CI
  - How to harness that creativity and their effort to enhance the CI ecosystem

# Socio-technical challenges and how to solve them?

- An important part of the socio-technical solution has been
  - Agile, open development practices which help to draw in talent (often students or postdocs) in the research domain help with development efforts
    - It also helps reduce the risk of reinvention while new tech is being explored and migration plan is developed.
  - But our communities are untrained in specific agile frameworks and have often arrived at agile practices through trial and error
  - SCiMMA formally adopted scrum for development teams. Small amount of training in agile roles and responsibilities was critical to the success
- Hesitance about hype
  - In LIGO, GitLab has helped to informally drive better agile practices with widespread adoption of the self-hosted premium service. But we don't have sufficient effort to properly maintain it at this scale.
    - Compare: gitlab ultimate (at \$99 per user per month) would cost \$2.4M per year for 2000 users
    - LIGO currently has 25% of an FTE to keep it going.

# CI solution that you would like to share with others?

- <https://hop.scimma.org/>
  - (Beta) Event Streaming Hub for Multi-Messenger Astronomy
    - Hop.SCIMMA allows observatories to publish streams of candidates and discoveries, and scientists, brokers, and TOMs to subscribe to these streams. Hop.SCIMMA is powered by Hopskotch.
- SCiMMA Analysis Platform
  - SCIMMA's analysis platform provides a full-featured remote Jupyter environment, with all the required Hopskotch packages pre-installed.

