# Analysis Facilities for HL-LHC (DOE)

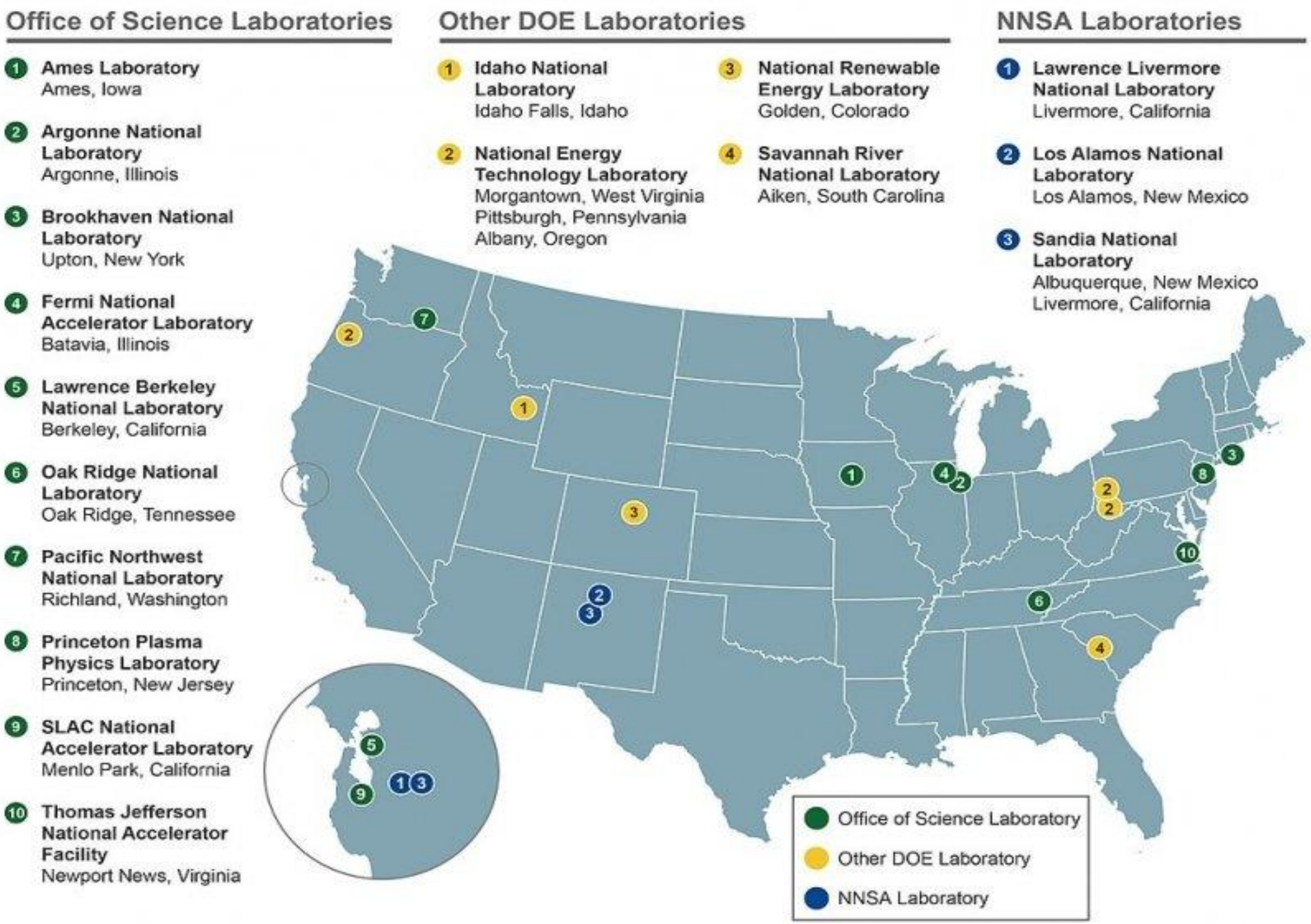
**Doug Benjamin** (BNL), Burt Holzman (FNAL), Ofer Rind (BNL), Wei Yang (SLAC)

4/7/2022

# Acknowledgements

- Ken Bloom, Brian Bockelman, Lincoln Bryant, Kyle Cranmer, Rob Gardner, Chris Hollowell, Eric Lancon, Ofer Rind, Oksana Shadura and Wei Yang
- And especially Burt Holzman (for several slides)

# DOE National Laboratories




- 17 national labs
- 4 with large HEP funding: **Fermilab**, **Brookhaven**, **SLAC**, Lawrence Berkeley

- Resources for analysis exist at all DOE Computing Centers

Brookhaven National Laboratory Fermilab SLAC NATIONAL ACCELERATOR LABORATORY

# Analysis Facilities at National Labs

- Pre-existing computing facilities
  - **Long history** of providing user analysis facilities

RHIC Computing Facility (RCF)

➢ Organizationally established in 1997

The first scientific non-data computer acquisition by the Laboratory occurred in 1970. About $500K had been allocated for the acquisition of a medium-sized computer to service the bubble-chamber film-measuring and analysis needs generated by FAF. The

  - In the future we will focus on the **AFs in development** (support fast columnar analyses) that complement our existing AFs
- Security
  - As .gov sites, labs are generally subjected to increased scrutiny and oversight
  - Certification of software / path to FedRAMP certification is helpful
- Multi-tenancy
  - Serve broad communities, not single experiments (and not necessarily just HEP)

# Fundamental principles:

- Create a user-oriented analysis facility based on our own experiences supporting scientists
.
- Explore, deploy and collaborate on industry-level technologies and strategies for optimizing data analysis partly in preparation for HL-LHC and upcoming experiments with large data demands such as DUNE.
- Foster collaboration with BES, NP and HEP experiments in order to better understand science analysis needs and provide computing solutions accordingly.

| Secure | Integrated & functional | Multi-VO | DevOps (operational sustainability) | Active collaboration |
| --- | --- | --- | --- | --- |

# Common Needs

- Both ATLAS and CMS need a flexible cyberinfrastructure suitable for quickly deploying additional services (potentially including off-premises resources) and serving the US analysis community and beyond.
- The LHC community needs to share common software substrates and approaches amongst the sites in order to be sustainable.
- Facilities must integrate with the existing distributed infrastructure; a successful analysis facility program will likely be a small percentage (<10%) of the overall hardware investment for HL-LHC computing and an even smaller portion of the global investment in scientific computing. Hence, future analysis facilities, like the current ones, **will be successful only by leveraging the larger computing resources, including those national scale resources.**

# Existing Analysis Facility Gaps

- Leveraging HPC centers:
  - High Performance Computing centers, such as DOE's Leadership Class Facilities at Lawrence Berkeley National Laboratory or the NSF-funded "Frontera" Leadership-Class Computing resource at TACC, are world-class computing facilities that provide unparalleled capabilities.
- Federated Authentication and Authorization:
  - The "Authentication and Authorization Infrastructure" (AAI) is a key design criteria for a facility.
  - Traditionally, each facility offering interactive access created a local Unix user account for each individual in the experiment desiring access, whereas Grid access can use global identities authenticating with an X.509 credential issued by a certificate authority.
  - There is activity amongst the DOE National Labs to allow Federated ID.
    - For Example at BNL – we have a jupyterhub instance that allows ATLAS users to use either their BNL, SLAC or CERN credentials to create a lightweight account. (is this good enough for most users?)

# Existing Analysis Facility Gaps (2)

- Authoring and sharing environments/data:
  - We need to enable end users to easily share their software environments within their and other groups.
  - ATLAS/CMS users often share data through EOS. This is one of reasons that users gravitate to CERN.
  - Several facilities have begun a transition from exclusively traditional batch environments with shared file systems, to ones that include container-based environments.
    - The shared file system for code and libraries, which is extremely limited in terms of reproducibility, portability, and scalability, is a simple and familiar model for many users and provides a mechanism to share software environments across groups.
    - Using CVMFS, experiments provide a shared file system-based environment for collaboration-wide software;
    - Groups often install analysis-specific software and modules on NFS servers at a given facility.
    - Containers can be helpful but require more technical competence from the end-users

# Crystal Ball Gazing and Conclusions:

- Cyber Security landscape is ever changing – Analysis facilities need to be adaptive without sacrificing usability for the users
  - New services will need to be deployed.
- Over the next decade and beyond –
  - Data volumes for analysis will increase.
  - How users do their analysis is evolving with new techniques and tools from outside of HEP. Our analysis centers must be responsive to this.
    - Likely need to leverage additional resources (for example: ASCR/NSF HPC machines for ML)
  - Labor to support scientific computing is not expanding, so we have to be more efficient with the labor that we have through increased automation.
  - Scientist time is our most precious resource we must figure out how to make the scientists more effective.