

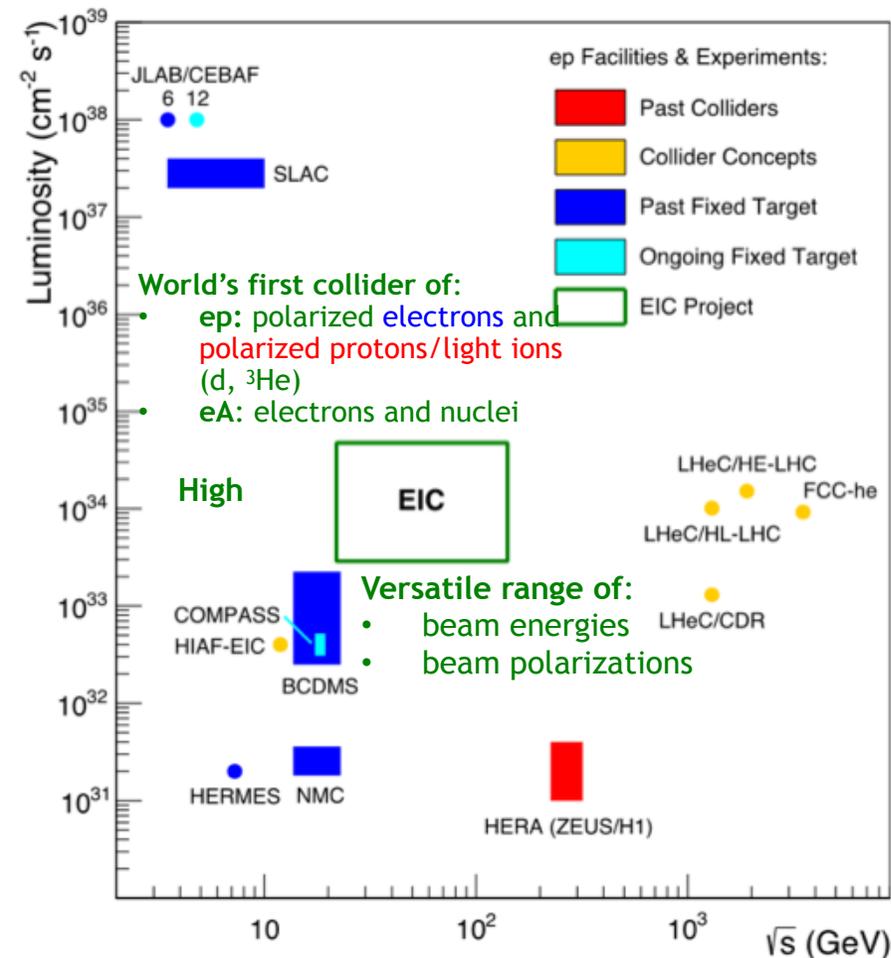
# FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

## Workshop Summary

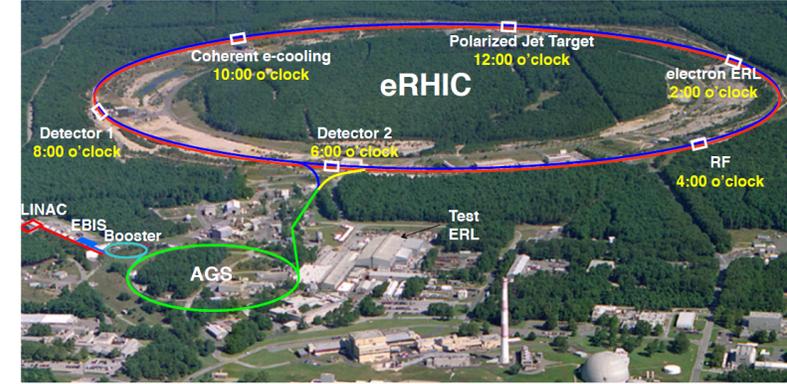
Ofer Rind, BNL (for the organizers)  
GDB Meeting  
Dec. 9th, 2020

# The Nuclear Physics Frontier

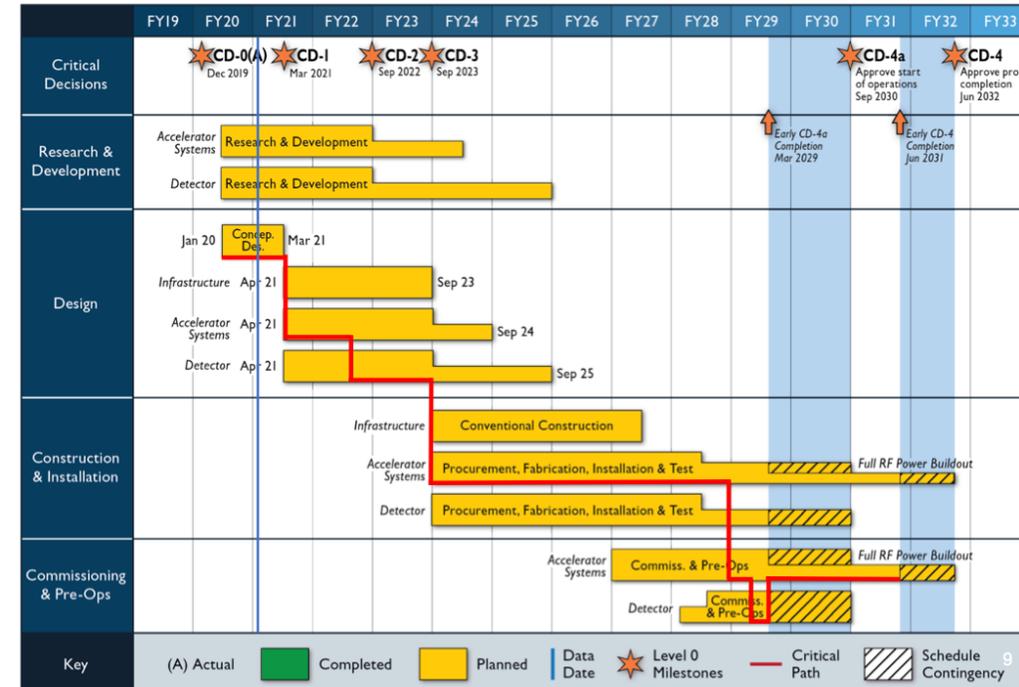
The mission of the Nuclear Physics program in the U.S. is to discover, explore, and understand all forms of nuclear matter.



The Electron-Ion Collider will operate at BNL with JLAB as a major partner in realizing the project



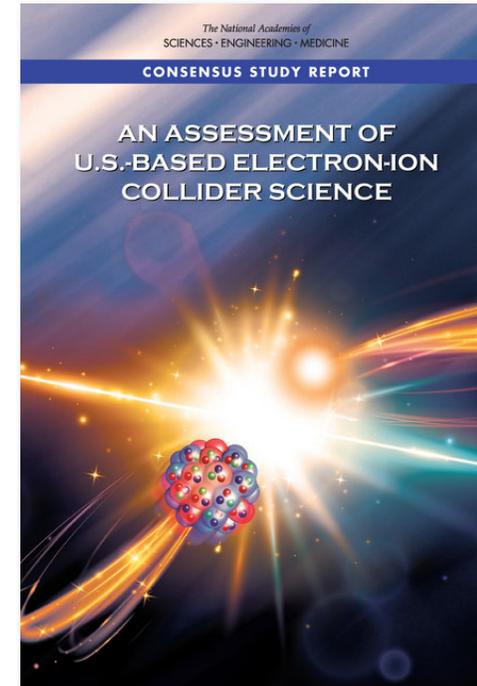
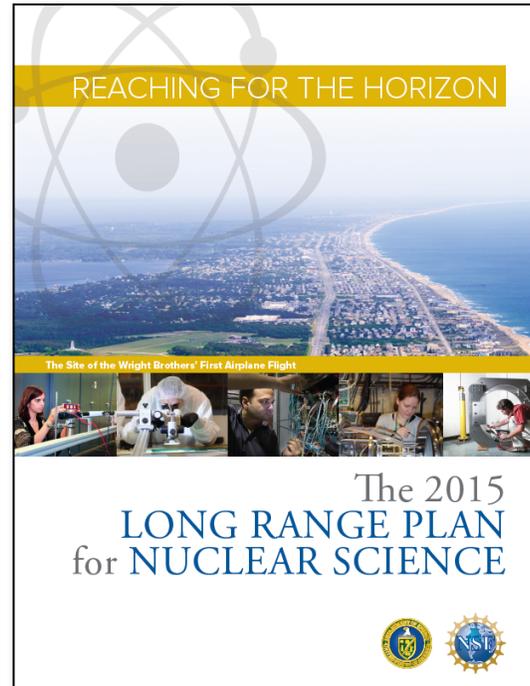
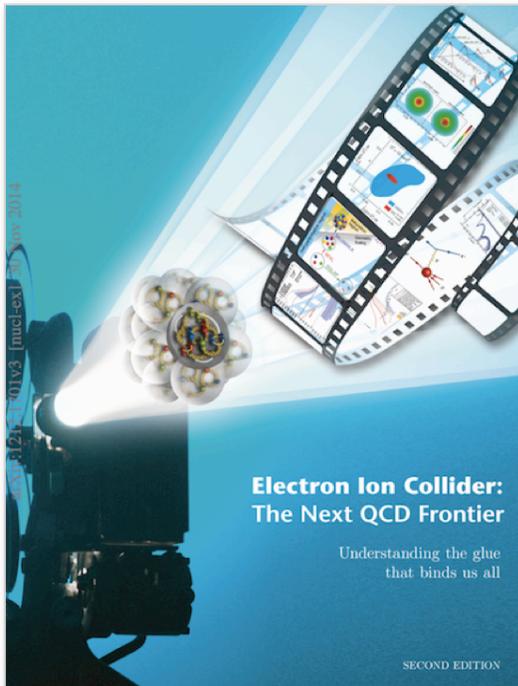
In the meantime, a robust NP program continues at BNL (sPHENIX, STAR) and at JLAB (12 GeV) with its own evolving computing challenges



# The EIC User Group (EICUG)



How will the computing effort in the nuclear physics community need to evolve in order to meet the needs of this evolving experimental landscape?



## Future Trends in Nuclear Physics Computing in 2016



76 participants

### Goals

- Examined computing strategy at a time horizon of ten years
- Defined common vision for NP computing
- Recommended future directions for development

### Website

- <https://www.jlab.org/conferences/trends2016/>

## Future Trends in Nuclear Physics Computing in 2017



74 participants

### Goals

- Discussed trends in scientific computing
- Collected ideas on how to improve analysis
- Worked towards next-generation analysis techniques and tools

### Website

- <https://www.jlab.org/conferences/trends2017/>

**BROOKHAVEN** & Jefferson Lab  
NATIONAL LABORATORY

## FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

SEPT. 29 - OCT. 1, 2020

The workshop focuses on the Nuclear Physics Software & Computing community. We will identify what is unique about our community and we will discuss how we can strengthen common efforts and chart a path for Software & Computing in Nuclear Physics for the next ten years.

### TOPICS:

- Common Scientific Software
- The Role of Data Centers in Scientific Discovery
- Unique Software Challenges for Nuclear Physics

### Focus on the Nuclear Physics Software & Computing community

- Identify what is unique about our community
- Discuss how we could strengthen common efforts
- Chart a path for Nuclear Physics Software & Computing for the next ten years

### Website

- <https://indico.bnl.gov/event/9023/timetable/>

# Joint Organization by BNL and JLab



Alexander Kiselev



Amber Boehnlein



Graham Heyes



Mark Ito



Markus Diefenthaler



Ofer Rind



Paul Laycock



Torre Wenaus



**207 participants**

**THANK YOU FOR JOINING US!**

# Organization

- We met for four hours each day in a time window chosen to be as inclusive as possible for participants around the world:  
**9:00 a.m. - 1:00 p.m. (EDT)**
- Substantial discussion time was included in the agenda.

The screenshot shows a Zoom meeting agenda for a session titled "Common Scientific Software: Password for watching the recording: NPTRENDS20". The session is moderated by Paul Laycock (BNL) and Torre Wenaus (BNL). The agenda includes the following items:

- 9:00 AM** - Workshop Goals (15m): Speakers: Alexander Kiselev (BNL), Dr Amber Boehnlein (Jefferson Lab), Dr Graham Heyes (Jefferson Lab), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab), Ofer Rind (BNL), Paul Laycock (BNL), Torre Wenaus (BNL). Includes a "TRENDS2020-Welc..." document.
- 9:15 AM** - Developing Common Software: ACTS (30m): Speaker: Andreas Salzburger (CERN). Includes a "2020-ACTS-Future-T..." document.
- 9:45 AM** - Questions on Talk (10m)
- 10:00 AM** - Software Sustainability (30m): Speaker: Daniel S. Katz (University of Illinois Urbana-Champaign). Includes a "nuclear-physics-co..." document.
- 10:30 AM** - Questions on Talk (10m)
- 10:45 AM** - Developing Common Software: Rucio (30m): Speaker: Mario Lassnig. Includes a "Rucio @ Future Tren..." document.
- 11:15 AM** - Questions on Talk (10m)
- 11:30 AM** - Discussion on Common Software (1h 30m)

## Common Scientific Software

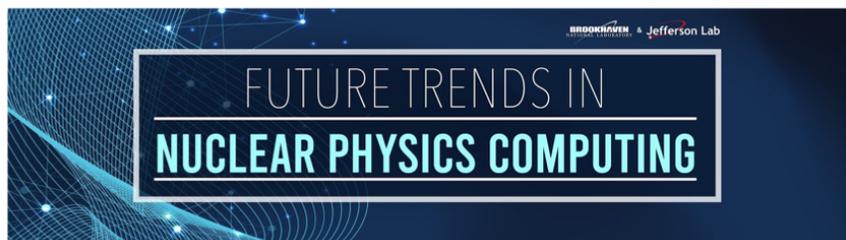
- Tuesday, September 29
- Moderated by P. Laycock and T. Wenaus

## The Role of Data Centers in Scientific Discovery

- Wednesday, September 30
- Moderated by G. Heyes and O. Rind

## Unique Software Challenges for Nuclear Physics

- Thursday, October 1
- Moderated by M. Diefenthaler, M. Ito, and A. Kiselev



## Future Trends in Nuclear Physics Computing Meeting Notes

[Timetable](#)

This is the live meeting notes document for the [Future Trends in Nuclear Physics Computing Workshop](#) held on September 29 - October 1, 2020. This workshop, the third of the series (previous editions were in [2017](#) and [2016](#)), focuses on the Nuclear Physics Software & Computing community itself. Goals for the workshop are to identify what is unique about our community, find ways to strengthen common efforts, and chart a path for Software & Computing in Nuclear Physics for the next ten years.

We meet for four hours each day in a time window chosen to be as inclusive as possible for participants around the world. Substantial discussion time is included in the agenda, and session conveners will keep speakers to time in order to preserve the discussion time. This google doc will be used in advance to give the discussions structure and focus, as well as during the workshop itself to moderate and record the discussion and gather input from all participants, and after the workshop as the basis for summarizing and report writing. Editing is on, and all participants are encouraged to contribute in all phases.

Each day has a theme. In advance of the workshop, questions and discussion points for each day will be gathered here to guide a moderated common discussion following the talks. A short discussion period will follow each talk to address questions specific to the talk. The content prepared in advance will be augmented during the presentations and discussions.

A brief synopsis of the previous day will be part of an intro talk on days two and three.

The workshop will conclude with a short summary, but summarizing and report writing proper will proceed after the workshop. All participants are welcome and encouraged to join the meeting organizers in this work. The outcome will be a written report, with presentation and discussion of the report in the subsequent meeting of the "Software & Computing Round Table" that is jointly organized

## Live notes

### Scope

- Moderate and record the discussion
- Gather questions and discussion points from all participants
- **After the workshop:** summarizing and report writing
- [https://docs.google.com/document/d/1mug\\_UB31WngFvvlLv8CFRbd4dSdKZR0iROL3m3lFY5Y/edit](https://docs.google.com/document/d/1mug_UB31WngFvvlLv8CFRbd4dSdKZR0iROL3m3lFY5Y/edit)
- 26 (!) pages

# Workshop Overview

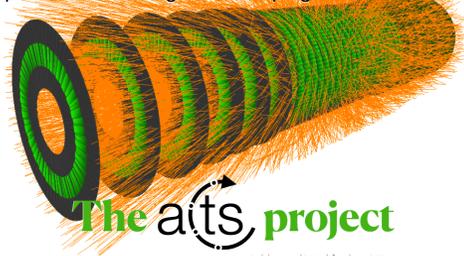
Common Scientific Software	The Role of Data Centers in Scientific Discovery	Unique Software Challenges for Nuclear Physics
Developing Common Software: ACTS Andreas Salzburger (CERN)	Data Centers in a Decade Eric Lancon (BNL)	Experience From the 12 GeV Science Program David Lawrence (JLAB)
Software Sustainability Daniel Katz (UIUC)	Data and Analysis Preservation Maxim Potekhin (BNL)	Software Challenges in Streaming Readout Jan Bernauer (SBU/Riken BNL)
Developing Common Software: Rucio Mario Lassnig (CERN)	User Perspectives and Requirements Graham Heyes (JLAB)	Survey From Nuclear Physics Students and Young Postdocs Markus Diefenthaler (JLAB)

- We thank all our speakers for doing an excellent job of making the sessions as interesting and useful as we could have hoped
- Thank you also to the workshop participants for stimulating discussions, challenging questions, and frank open answers

## The keys to success

- Looked at two software development case studies: ACTS (tracking), Rucio (data mgmt)
- **The team is most important**
  - Both cases benefited from crucial experience developing and operating a worse software package
  - Do not separate development and operations
  - Developers keen to use modern software paradigms, open-source and open-minded, proactively searching out and adopting best practices
- **Focused Project** Clear short-term goals, grounded in real-world deliverables
  - Aligned with the long-term plan of building something sustainable
  - Designed to be used by outside collaborators.
- **Disciplined Management**
  - Accept that the long view takes longer to deliver the short-term product
  - Manage expectations of the collaboration and funders to ensure the team have sufficient time and space to succeed.

Experiences & findings in developing common software



Experiences building a common system

# Themes: Community Support for Common Software

## Career Support for Scientific Software Development

- Active, passionate, recurring theme!
- Recognition, encouragement and reward: need to make software citations a priority
- Provide career paths and funding that allow for and value software development
  - Research Software Engineers (RSE) need to be supported and not only at the labs

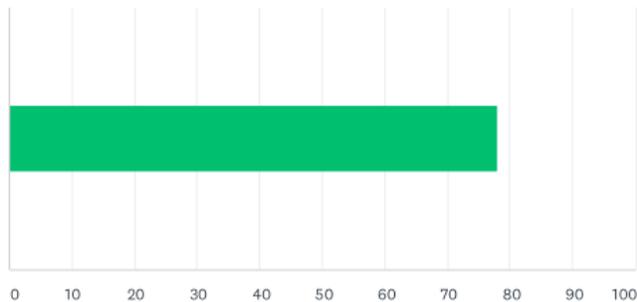


## Support for education and training in software development

- Quality tutorials are extremely important for teaching young scientists

Q1 What fraction of your time do you spend on the software and computing aspects of your research, such as programming, analysis jobs, etc.?

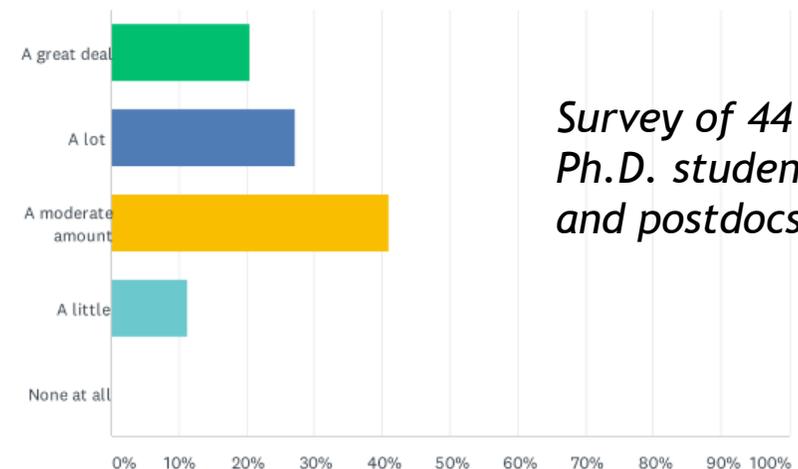
Answered: 44 Skipped: 0



ANSWER CHOICES	AVERAGE NUMBER	TOTAL NUMBER	RESPONSES
	78	3,435	44
Total Respondents: 44			

Q7 Are you confident performing software and computing tasks needed for your research?

Answered: 44 Skipped: 0



*Survey of 44  
Ph.D. students  
and postdocs*

## Organizing Community Software Efforts

- The NP field is going through a transition toward larger data size and more complex analyses
  - Traditional model of smaller groups developing in-house solutions is not sustainable
  - Can benefit from a more unified approach to effectively address problems at the scale of future experiments
  - Finding the right balance between in-house and community development is a challenge, both technical and cultural
- Analysis approaches/models can differ from those in HEP and may require unique software and computing strategies
  - For example, looking at multi-dimensional, strongly correlated relationships among data opposed to search of rare events with novel topologies

## Should the NP community participate in HSF or build its own organization?

- Pros and cons, but the balance of opinion favored NP participation in HSF
  - HSF is a do-ocracy, active participation will yield the biggest rewards.

## Evolution of Data Center Infrastructure

- Adapting to support increasing number of data-intensive scientific programs
  - Edge Computing, Superfacility Model
- Use of disruptive technologies - AI/ML and coprocessors (GPU, FPGA)
  - Tension between HTC and HPC in resource planning
- Role of tape vs. disk and cloud storage
- Integration of distributed computing resources
  - Where do the LCFs fit in?
  - Federated authentication/authorization
- Evolution of high bandwidth networking - how will this affect compute models?

## Long term Data and Analysis Preservation (DAP)

- Importance of preserving metadata and code alongside raw data
- Importance of documentation and proper choice of tools
- Importance of building DAP into the infrastructure and policies at an early stage

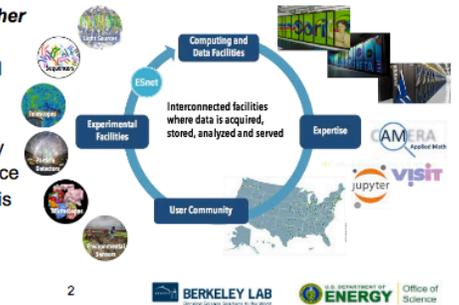
Containerization as mechanism for DAP as well as for operating on distributed resources

Challenges in developing common, relevant computing benchmarks for NP

Superfacility: an ecosystem of connected facilities, software and expertise to enable new modes of discovery

Superfacility@LBNL: NERSC, ESnet and CRD working together

- A model to integrate experimental, computational and networking facilities for reproducible science
- Enabling new discoveries by coupling experimental science with large scale data analysis and simulations



NERSC

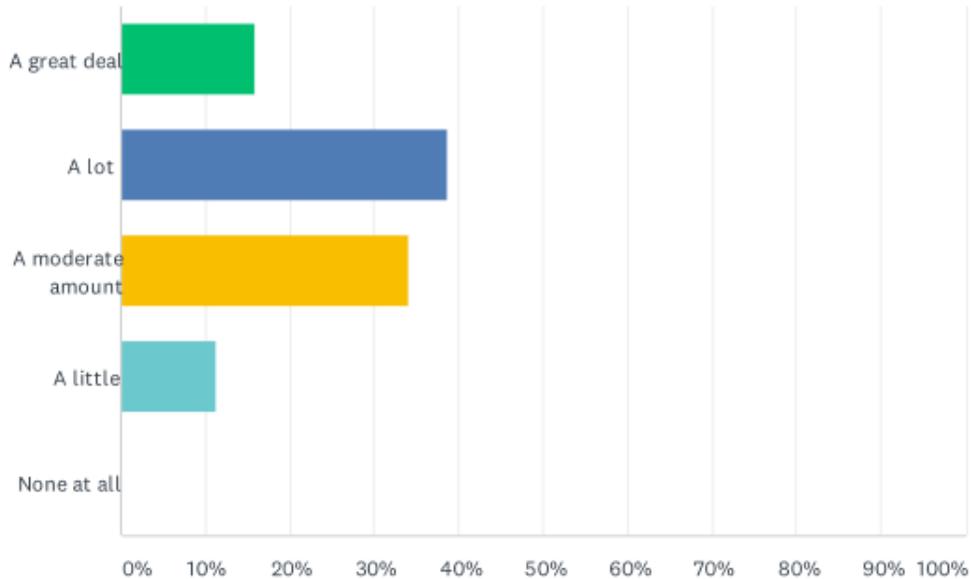
2

BERKELEY LAB  
BERKELEY NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY  
Office of Science

## Q9 Are the computing resources you need readily available?

Answered: 44 Skipped: 0

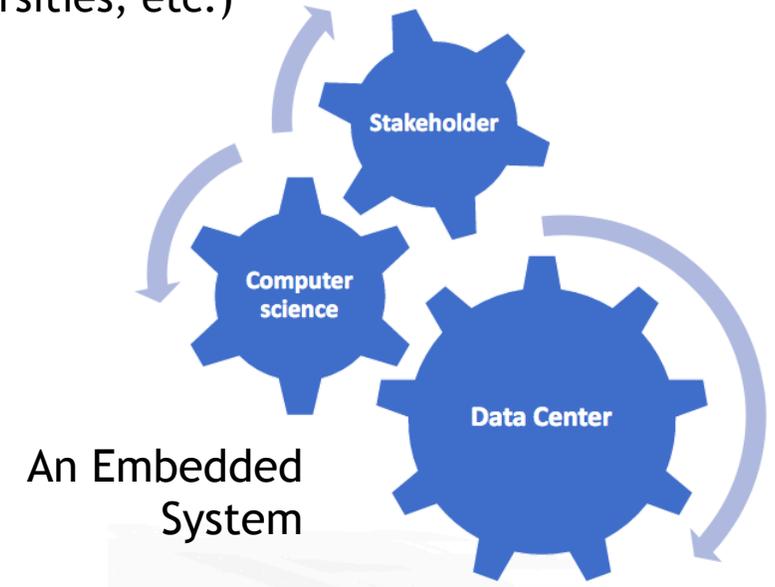


### A survey comment:

It seems that funds are spent mostly on computing resources, but very little on computing expertise (if we had half the computers and twice the number of software experts we would be in a much better place)

## Interactions between facilities and stakeholders

- Need for improved two-way communication
  - Educating users on how to use the data center
  - Communicating stakeholder needs to the facility
- Embedding facility personnel within experiments and vice versa
- Connecting ops personnel with CS researchers (at labs, at universities, etc.)



BROOKHAVEN  
NATIONAL LABORATORY & Jefferson Lab

## WORKSHOP REPORT

# FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

SEPT. 29 - OCT. 1, 2020

### EDITORS

Alexander Kiselev (BNL)  
Amber Boehnlein (JLAB)  
Graham Heyes (JLAB)  
Mark Ito (JLAB)

Markus Diefenthaler (JLAB)  
Ofar Rind (BNL)  
Paul Laycock (BNL)  
Torre Wenaus (BNL)

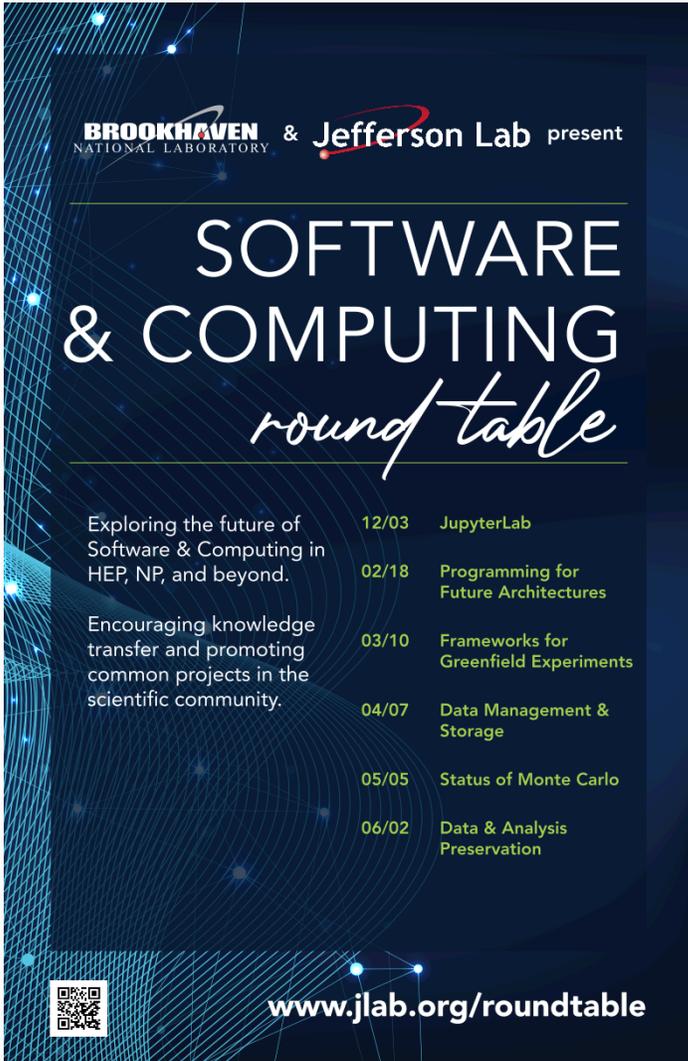
<https://indico.bnl.gov/event/9023/>

## Workshop Report

- **Draft** (28 pages and growing)
  - <https://docs.google.com/document/d/1HT0dQJwZkUGaxGS0bUhZgkz0l2G5QR-4Dm08-hr8DJc/edit?usp=sharing>
- Everyone was welcome and encouraged to provide comments between **Nov. 10 - Nov. 17.**
- **Status**
  - Finalizing the report.
  - Will be shared on arXiv.

## Action Item: Community Forum

- There was consensus support for starting a new **community forum for NP** to discuss common projects, the role of data centers, unique challenges etc.
- Ongoing discussion of details and possible connection to HSF.
- Possible **goals of the community forum**
  - Inform on building successful scientific software projects, taking the unique challenges of NP into consideration
  - Foster collaborative common software projects in NP
  - Promote scientific software career support



**BROOKHAVEN** & **Jefferson Lab** present  
NATIONAL LABORATORY

## SOFTWARE & COMPUTING *round table*

Exploring the future of Software & Computing in HEP, NP, and beyond. 12/03 **JupyterLab**

Encouraging knowledge transfer and promoting common projects in the scientific community. 02/18 **Programming for Future Architectures**

03/10 **Frameworks for Greenfield Experiments**

04/07 **Data Management & Storage**

05/05 **Status of Monte Carlo**

06/02 **Data & Analysis Preservation**

 [www.jlab.org/roundtable](http://www.jlab.org/roundtable)

Jointly organized by BNL and JLab

- Role of Software & Computing Round Table in 2016 to now
  - Encourage knowledge transfer
  - Promote common projects
- Exploring the future of software & computing in HEP, NP, and beyond
- Emphasis on the interplay of software & computing and science
- Website: <https://www.jlab.org/roundtable>
- Have now added Graeme Stewart to organizing committee in order to encourage cross-promotion with and participation in HSF

### NUCLEAR PHYSICS DATA DEMAND MORE POWERFUL PROCESSING



*Jefferson Lab and Brookhaven National Lab partner on a Software & Computing Round Table to track the leading edge of computing and foster collaboration*

Fans of the popular TV show “The Big Bang Theory” can picture the sitcom’s physicists standing at a whiteboard, staring hard at equations.

It’s an iconic image. But is that the future — or even the present — of how nuclear physicists do their jobs? Not really. Not when new experiments demand ever-more powerful data processing and thus ever-more-powerful software and computing.

“Scientists being at a blackboard and writing up some equations — that is not always the reality,” said Markus Diefenthaler, an experimental nuclear physicist at the U.S. Department of Energy’s **Thomas Jefferson National Accelerator Facility** in Newport News, Virginia.

## [News story on Software & Computing Round Table](#)

# FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

Questions?