U.S. DEPARTMENT OF ENERGY | Office of Science

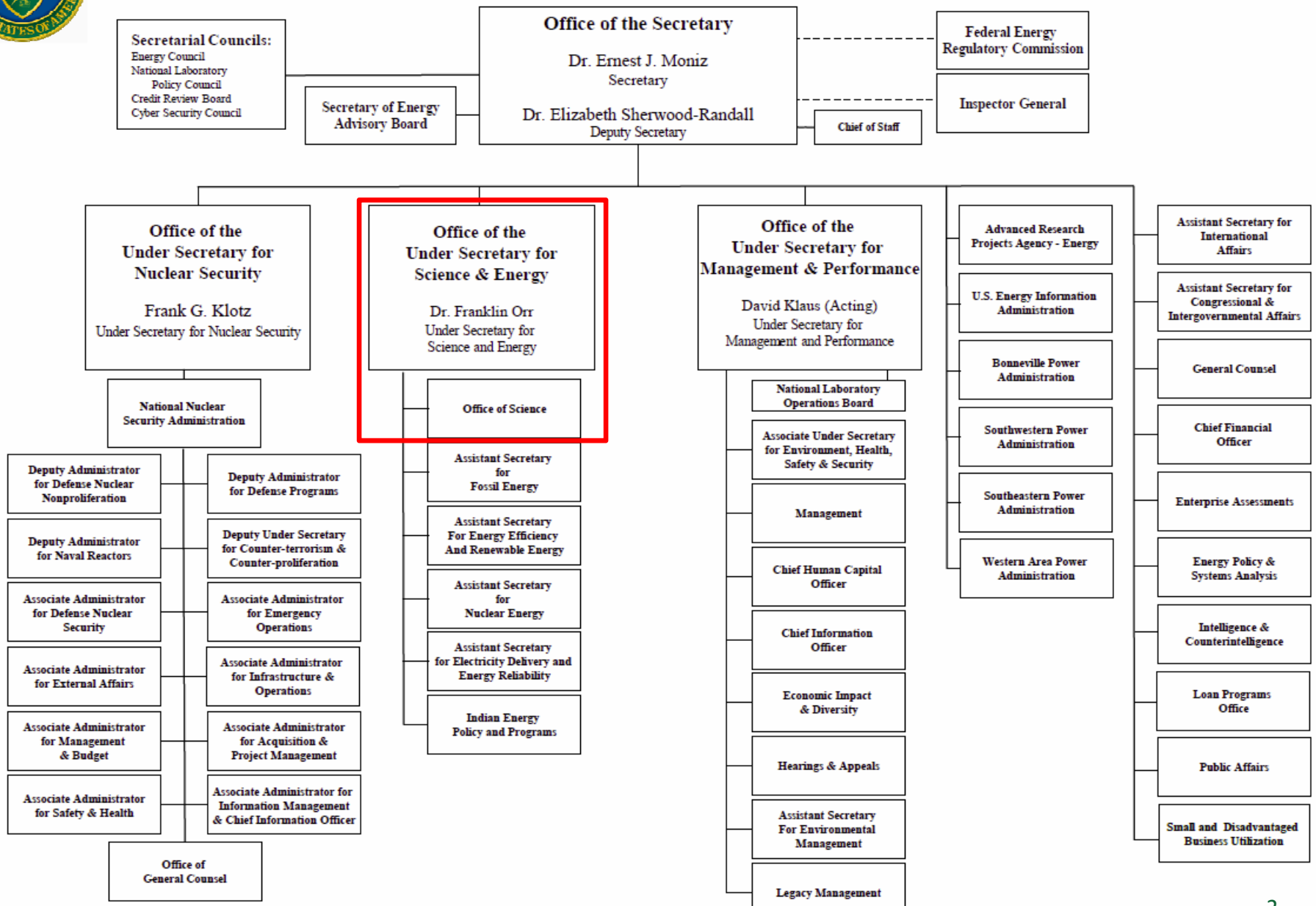
Cybersecurity for Scientific Computing Integrity

Dr. Robinson Pino

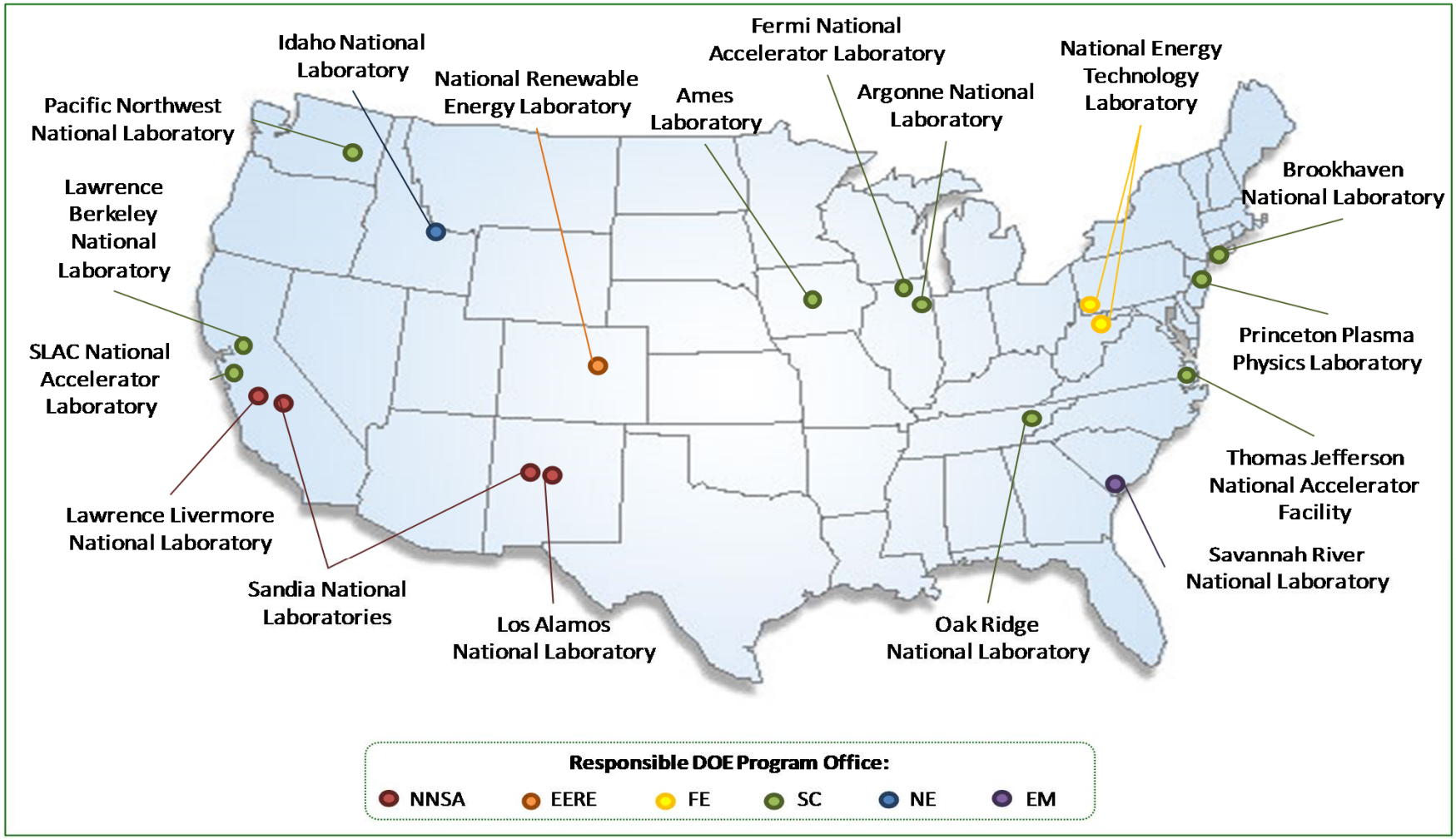
http://science.energy.gov/ascr/

January 10, 2017

DEPARTMENT OF ENERGY



DOE's 17 National Laboratories



Quick-Facts about the DOE Office of Science

Advanced Scientific Computing Research (ASCR)

Basic Energy Sciences (BES)

Biological and Environmental Research (BER)

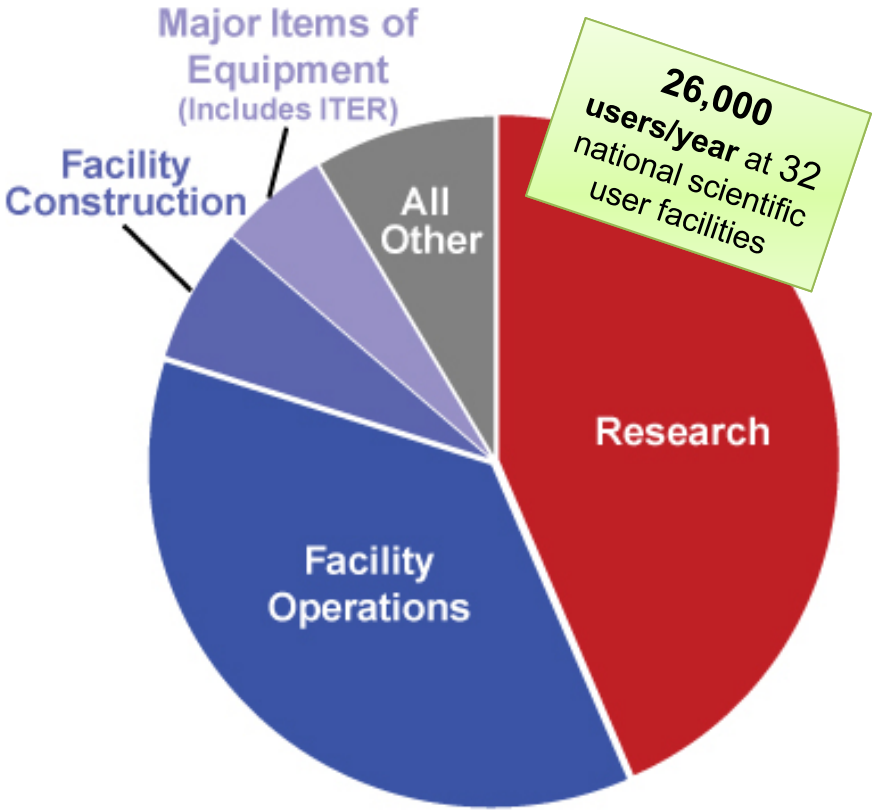
Fusion Energy Sciences (FES)

High Energy Physics (HEP)

Nuclear Physics (NP)

DOE Office of Science – At a Glance

FY 2015 Funding
Total = $5.140 billion

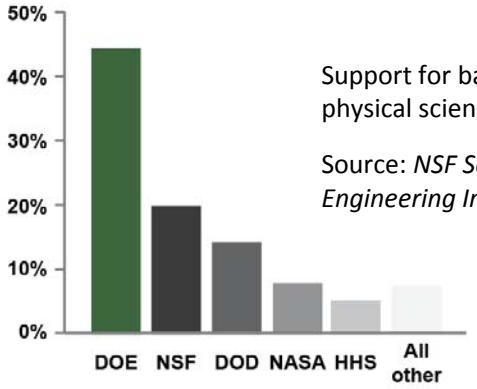


Source: http://science.energy.gov/about/

Mission: SC delivers scientific discoveries and tools to transform our understanding of nature and advance the energy, economic, and national security of the U.S.

Research

- Support for 47% of the U.S. Federal support of basic research in the physical sciences;
- ~22,000 Ph.D. scientists, grad students, engineers, and support staff at >300 institutions, including all 17 DOE labs;
- U.S. and world leadership in high-performance computing and computational sciences;
- Major U.S. supporter of physics, chemistry, materials sciences, and biology for discovery and for energy sciences.



Support for basic research in the physical sciences by agency.

Source: *NSF Science and Engineering Indicators 2012*

Scientific User Facilities

- The world's largest collection of scientific user facilities (aka research infrastructure) operated by a single organization in the world, used by 31,000 researchers each year.

Advanced Scientific Computing Research

Computational and networking capabilities to extend the frontiers of science and technology

- **Mathematics research** to address challenges of increasing complexity within DOE's mission areas from a mathematical perspective. This requires integrated, iterative processes across multiple mathematical disciplines.
- **Computer science research** to increase the productivity and integrity of HPC systems and simulations, and support data management, analysis, and visualization techniques.
- **SciDAC partnerships** to dramatically accelerate progress in scientific computing that delivers breakthrough scientific results.
- **Exascale computing** research and development of capable exascale hardware architectures and system software, including the deployment of programming environments for energy-efficient, data-intensive applications, and engagement with HPC vendors to deliver systems that address the exascale challenges.
- **Facilities** operate with at least 90% availability while continuing planned upgrades – begin deployment of 10-40 petaflop upgrade at NERSC and continue preparations for 75-200 petaflop upgrades at each LCF.
- **CSGF** - Continue a postdoctoral program at the ASCR facilities and provide funding for the Computational Science Graduate Fellowship to address DOE workforce needs.

Leadership Computing for Scientific Discovery

OLCF Titan System Specifications:

- Peak performance of 27.1 Petaflops
  - 24.5 GPU + 2.6 CPU
- 18,688 Hybrid Compute Nodes with:
  - 16-Core AMD Opteron CPU
  - NVIDIA Tesla "K20x" GPU
  - 32 + 6 GB memory
- 200 Cabinets; 710 TB total system memory; 8.9 MW peak power

- Peer reviewed projects are chosen to advance science, promote innovation, and strengthen industrial competitiveness.
- Demand for these machines has grown each year, requiring recent upgrades of both.

ALCF Mira System Specifications:

- Peak performance of 10 Petaflops
- 49,152 Compute Nodes each with:
  - 16-Core Power PC A2 CPU with 64 Hardware Threads and 16 Quad FPUs
  - 16 GB memory
- 56 Cabinets; 786 TB total system memory; 4.8 MW peak power

FY 2013 research projects include; advancing materials for lithium air batteries, solar cells, and superconductors ; improving combustion in fuel-efficient, near-zero-emissions systems; understanding how turbulence affects the efficiency of aircraft and other transportation systems; designing next-generation nuclear reactors and fuels ; developing fusion energy systems

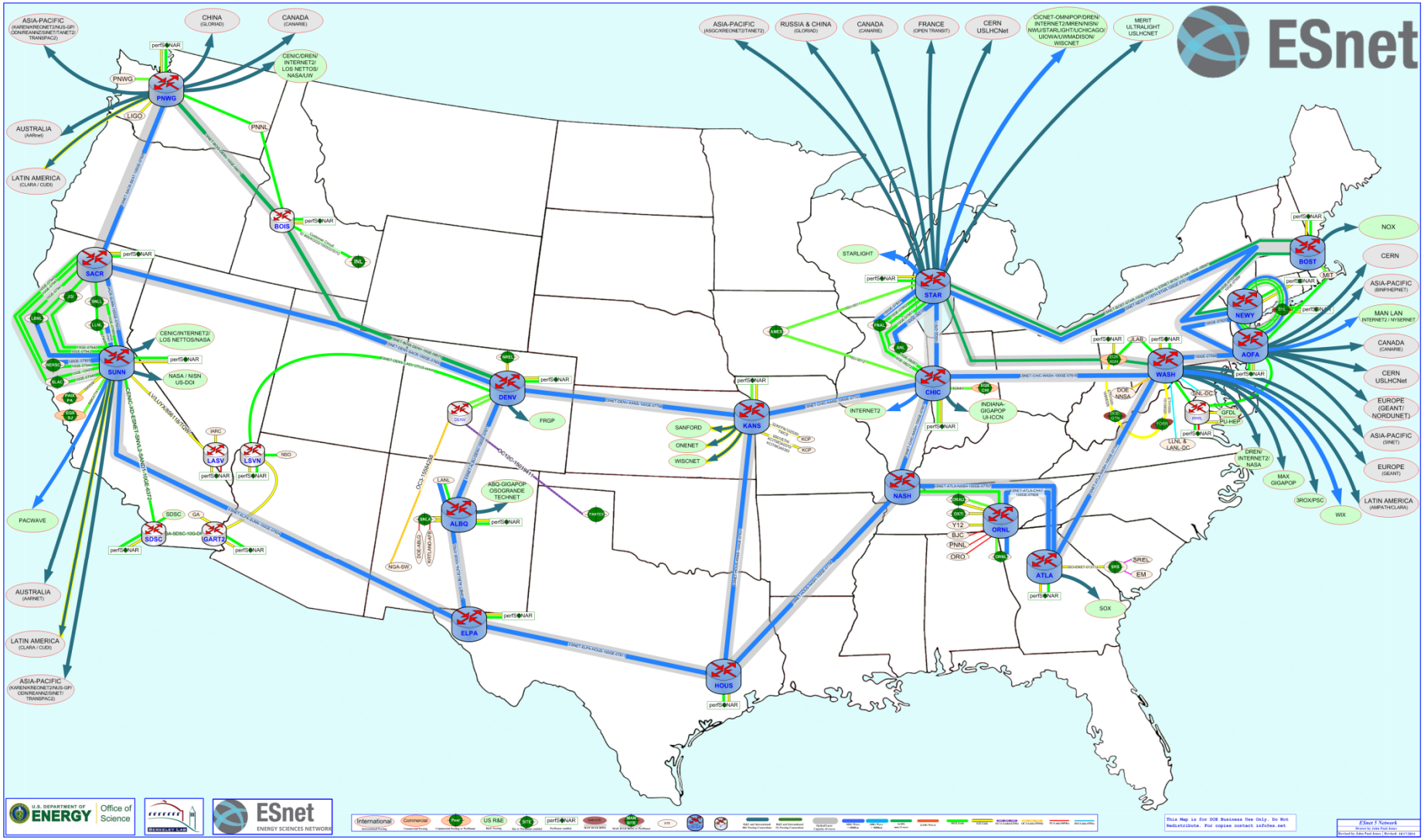
NERSC: 40 years of High Performance Computing for DOE

System Specifications:

- *Hopper XT5 (2010)*
  - 1.3PF, 212TB, 2.9 MW peak power
- *Edison XC30 (in acc*
  - Based on DARPA/DOE HPCS system
  - 2.4PF, 333TB, 2.1 MW peak power
- *400TF mixed use clusters*
  - NERSC, JGI, HEP/NP, Materials, Kbase

Computational Research and Theory Building will provide 12 MW power and cooling for future NERSC computing resources

Energy Sciences Network (Esnet)



Cybersecurity in HPC

2015 DOE Report: Cybersecurity for Scientific Computing Integrity

- Research and develop means to collect extreme-scale data and knowledge, and develop and apply analytics in order to understand and improve scientific computing integrity and computer security

- Develop means to learning and maintaining interdependent causal models of the scientific computation, exascale system, and computer security in real-time to enable better, faster recovery to reduce disruptions to scientists' efforts

- Metrics for quantifying the trustworthiness of scientific data, capturing the likelihood and potential magnitude of errors due to uncertain inputs, incomplete models, incorrect implementations, silent hardware errors, and malicious tampering

https://www.orau.gov/cybersecurity2015/
https://www.orau.gov/integrity2015/

What is scientific computing integrity?

Scientific Computing Integrity:

The ability to have high confidence that the scientific data that is generated, computed, processed, stored, or transmitted by computers and computer-connected devices has a process, provenance, and correctness that is understood

Why DOE Scientific Integrity?

- **DOE has the responsibility to address the energy, environmental, and nuclear security challenges that face our nation**
- **Much of DOE's enterprise involves distributed, collaborative teams**
  - A significant fraction involve "open science"
  - Depends on multi-institutional, often international collaborations that must access or share information between sites around the world
- **Office of Science mission is the delivery of scientific discoveries and major scientific tools to transform our understanding of nature and to advance the energy, economic, and national security of the United States.**
  - To execute its responsibilities, DOE must be able to assure the integrity and availability of scientific facilities and computer systems, and of the scientific, engineering, and operational software and data that support its mission

Workshop Report Vision and Goal

- **Identify fundamental research challenges to enable scientific computing integrity and computer security by:**
  - Achieving repeatable, reproducible workflows
  - Produce computing results whose process, origin, and data provenance is understood,
  - Correctness is understood, and for which uncertainty estimates are provided.
- **Capabilities must be enhanced by systems with autonomous decision-making capabilities**
  - Give scientists the ability to make informed decisions about the integrity of their data.

Requirements for Open Science Cybersecurity

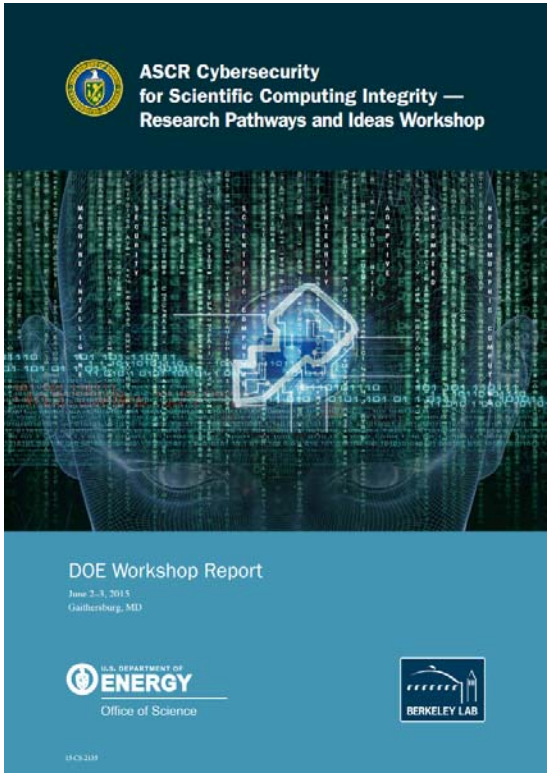
- **Office of Science labs function in completely open environments**
  - Availability and data sharing are of great importance
  - Are often accessed by authorized users around the world
  - Multiple projects from multiple countries run on the same machine
- **DOE HPC and large-scale science workflows differ from general-purpose computing**
  - DOE machines might run one program for weeks on tens of thousands of processors
  - Even other distributed, high-bandwidth services like Netflix or YouTube are smaller

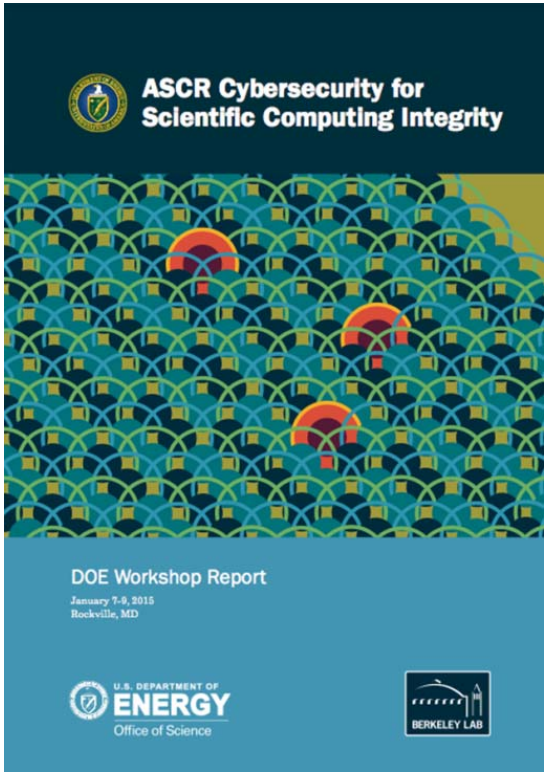
Workshop Results: Basic Research Areas

- **Trustworthy Supercomputing:**
  - Research within extreme high performance computing that can be influenced to incorporate a cybersecurity mindset during R&D
- **Trust within Open, High-End Networking and Data Centers:**
  - Research ideas that enable trust within an open shared environment among communications and data at rest or in transit minimizing security overhead
- **Extreme Scale Data, Knowledge, and Analytics for Understanding and Improving Cyber Security:**
  - Research to correlate, find, or detect patterns from heterogeneous sources of information such as the network, computing nodes, operating system, runtime, applications, etc. for use for cybersecurity, esp. scientific data integrity and provenance

Thank you!

Dr. Robinson Pino, *Program Manager, DOE Office of Science*
Robinson.Pino@science.doe.gov





http://science.energy.gov/ascr/
https://www.orau.gov/cybersecurity2015/
https://www.orau.gov/integrity2015/