ESnet
ENERGY SCIENCES NETWORK

The Superfacility Model

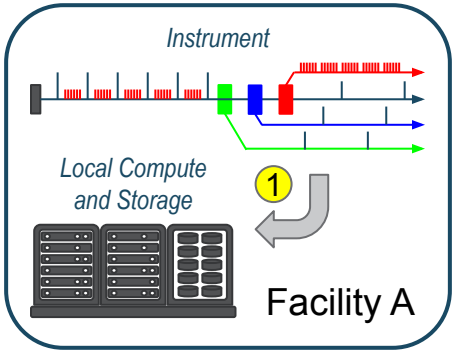
Chin Guok
Planning and Architecture Group Lead
Energy Sciences Network
Lawrence Berkeley National Laboratory

TNC22
Trieste, Italy
June 16, 2022

U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

Example of a Traditional Workflow for Large Scale Distributed Science Experiments

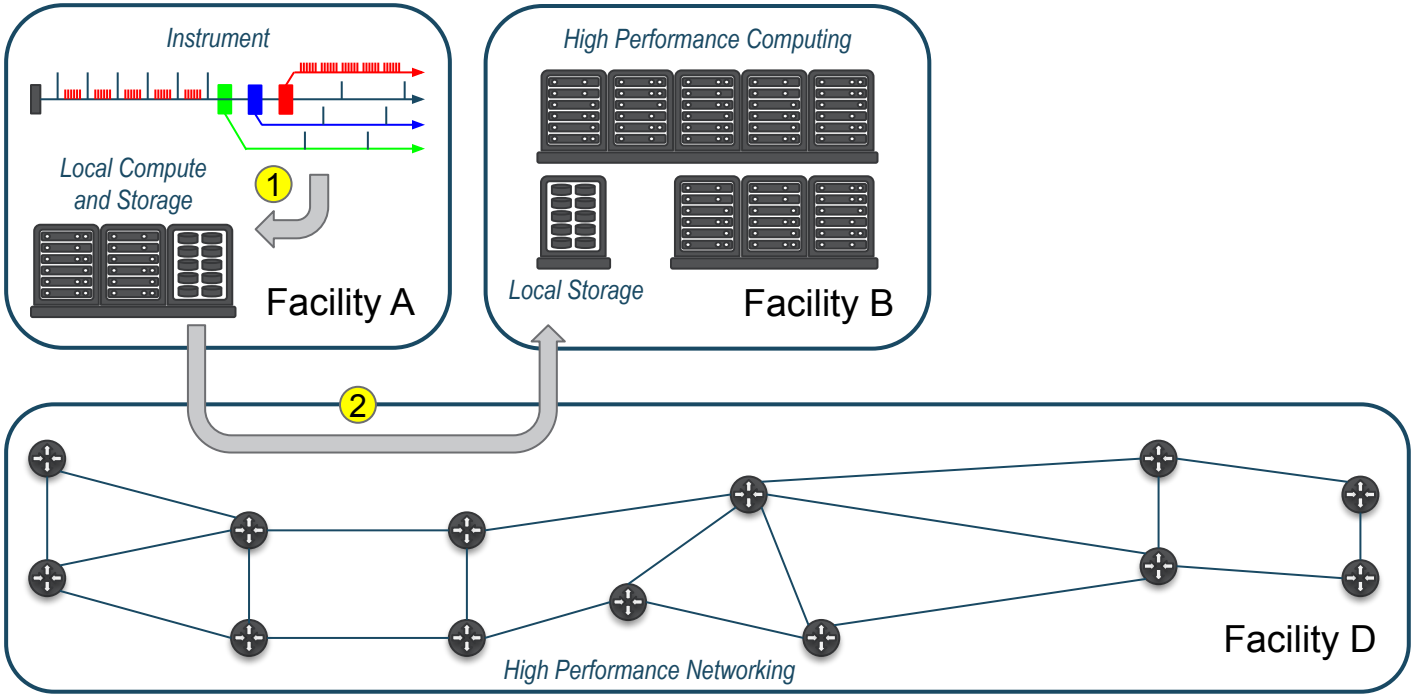


Discrete Data Flow

1. Initial data processing on local compute and staged in local storage.

ESnet

Example of a Traditional Workflow for Large Scale Distributed Science Experiments

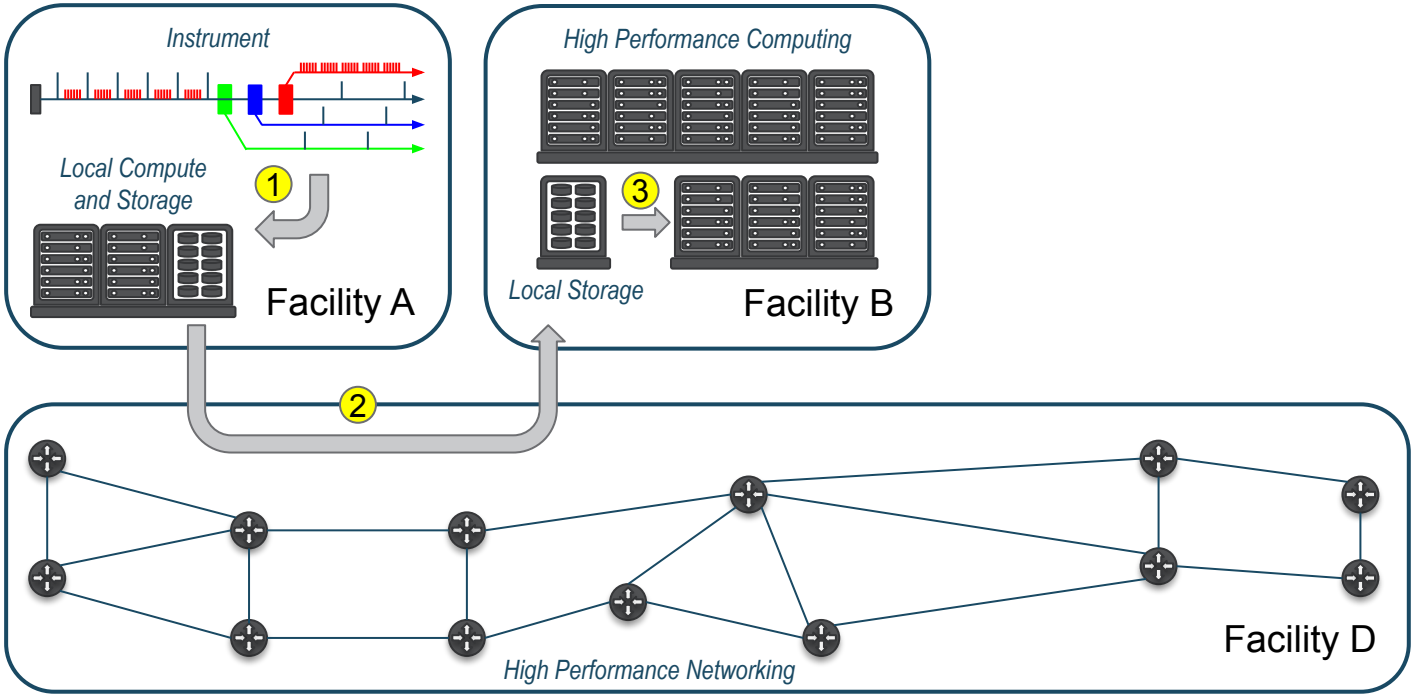


Discrete Data Flow

1. Initial data processing on local compute and staged in local storage.
2. Data transfer over WAN and staged in HPC local storage.

ESnet

Example of a Traditional Workflow for Large Scale Distributed Science Experiments

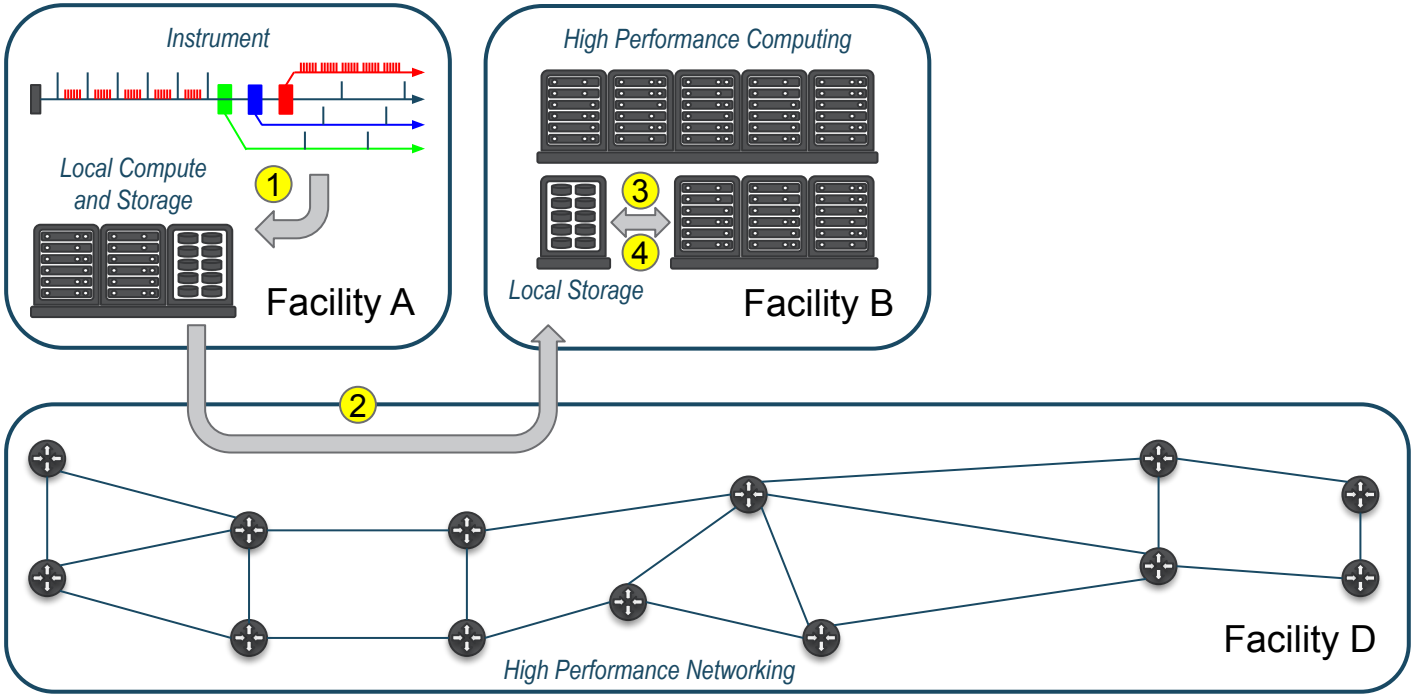


Discrete Data Flow

1. Initial data processing on local compute and staged in local storage.
2. Data transfer over WAN and staged in HPC local storage.
3. HPC fetches data from local storage for processing.

ESnet

Example of a Traditional Workflow for Large Scale Distributed Science Experiments

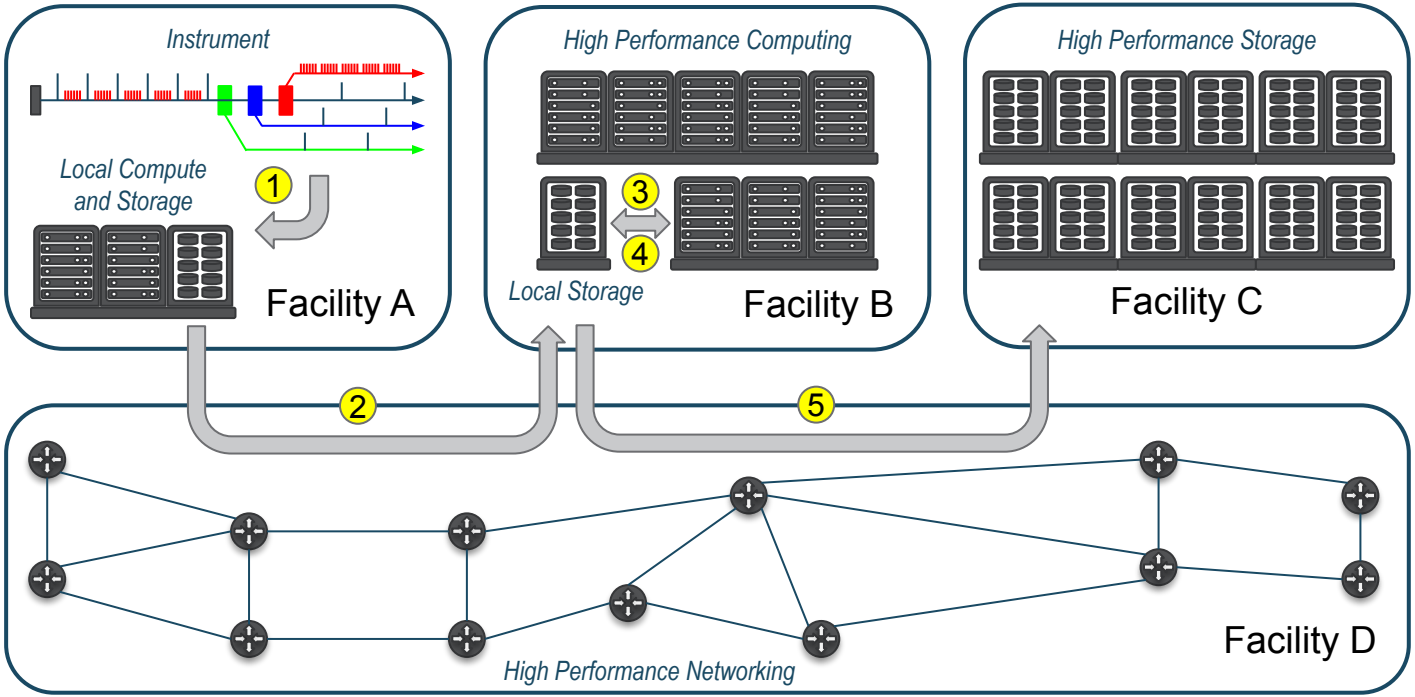


Discrete Data Flow

1. Initial data processing on local compute and staged in local storage.
2. Data transfer over WAN and staged in HPC local storage.
3. HPC fetches data from local storage for processing.
4. HPC stages processed data in local storage for transfer.

ESnet

Example of a Traditional Workflow for Large Scale Distributed Science Experiments



Discrete Data Flow

1. Initial data processing on local compute and staged in local storage.
2. Data transfer over WAN and staged in HPC local storage.
3. HPC fetches data from local storage for processing.
4. HPC stages processed data in local storage for transfer.
5. Data transfer over WAN into HPS for long term storage.

ESnet

Where are the inefficiencies?

- Experiments use local compute to compensate for the lack of schedulable (shared) compute resources.
  - Not every job is a good fit for an HPC.
  - Discrepancy between HPC job schedule and need for real-time computing to support the experiment.
- Networks are treated as “black-box” utilities, e.g., unpredictable performance, unknown status.
  - (Temporary) storage is used to stage data and compensate for lack of network performance predictability, resulting in multiple data transfers.
- Access to resources across domains is inconsistent
  - Different security mechanisms and allocation policies.
  - Different APIs and architectures, affecting job portability.
- ○○○

ESnet

History of DOE Office of Science

Manhattan Project
(1942 - 1946)

Energy Research and Development Administration
(1974 - 1977)

U.S. DEPARTMENT OF ENERGY
Office of Science

Department of Energy
Office of Science
(1998 - Present)

1940s | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s | Present

Atomic Energy Commission
(1946 - 1974)

Department of Energy
Office of Energy Research
(1977 - 1998)

ESnet

History of DOE Office of Science

Manhattan Project
(1942 - 1946)

Energy Research and Development Administration
(1974 - 1977)

U.S. DEPARTMENT OF ENERGY
Office of Science

Department of Energy
Office of Science
(1998 - Present)

1940s | 1950s | 1960 | 1970s | 1980s | 1990s | 2000s | 2010s | Present

Atomic Energy Commission
(1946 - 1974)

Department of Energy
Office of Energy Research
(1977 - 1998)

TEAM SCIENCE

ESnet

DOE Office of Science - Largest supporter of basic research in the physical sciences in the US

The mission of the **Advanced Scientific Computing Research (ASCR)** program is to discover, develop, and deploy computational and networking capabilities to analyze, model, simulate, and predict complex phenomena important to the Department of Energy (DOE).

Basic Energy Sciences (BES) supports fundamental research to understand, predict, and ultimately control matter and energy at the electronic, atomic, and molecular levels in order to provide the foundations for new energy technologies and to support DOE missions in energy, environment, and national security.

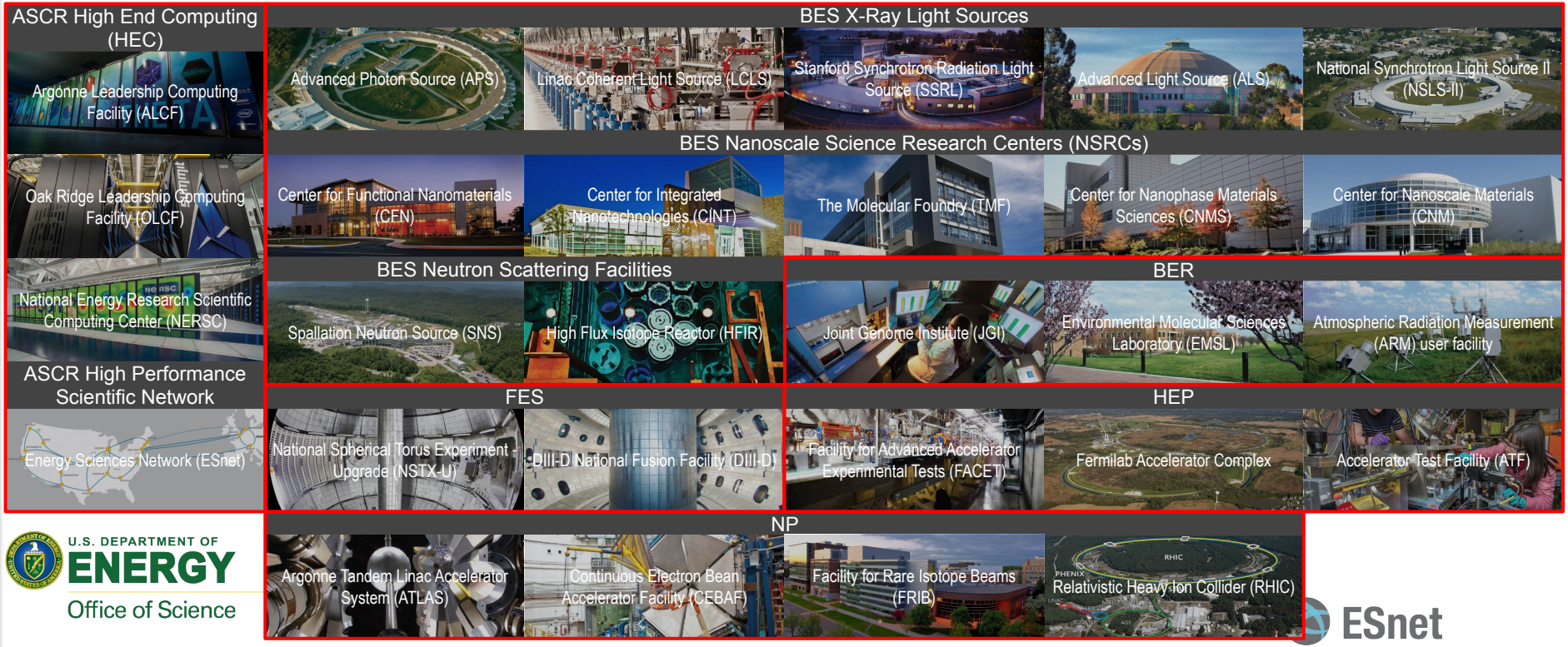
The mission of the **Biological and Environmental Research (BER)** program is to support transformative science and scientific user facilities to achieve a predictive understanding of complex biological, earth, and environmental systems for energy and infrastructure security, independence, and prosperity.

The **Fusion Energy Sciences (FES)** program mission is to expand the fundamental understanding of matter at very high temperatures and densities and to build the scientific foundation needed to develop a fusion energy source.

The mission of the **High Energy Physics (HEP)** program is to understand how our universe works at its most fundamental level.

The mission of the **Nuclear Physics (NP)** program is to discover, explore, and understand all forms of nuclear matter.

DOE Office of Science - Uniquely positioned for large scale collaborative science*

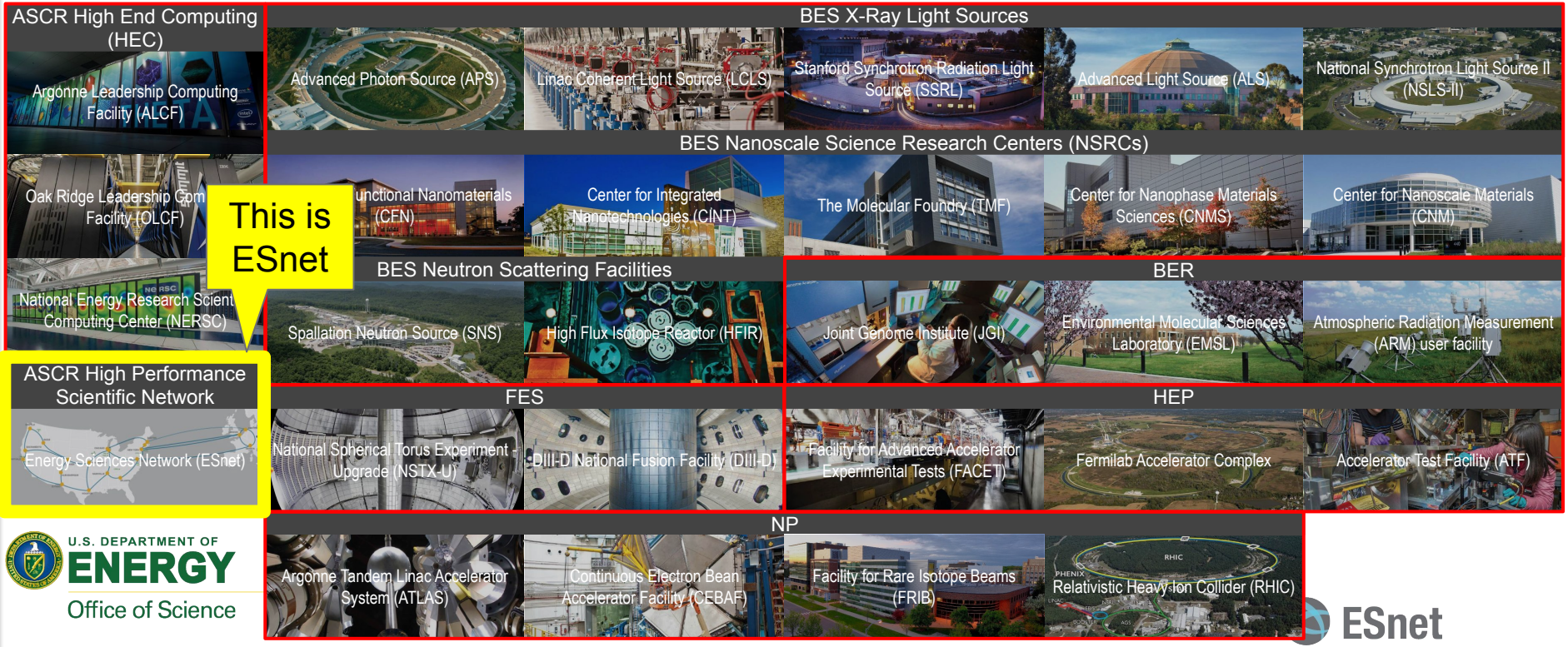


U.S. DEPARTMENT OF ENERGY
Office of Science

ESnet

*DOE Office of Science facilities also support other collaborations, e.g., LHC, LSST, etc

DOE Office of Science - Uniquely positioned for large scale collaborative science*



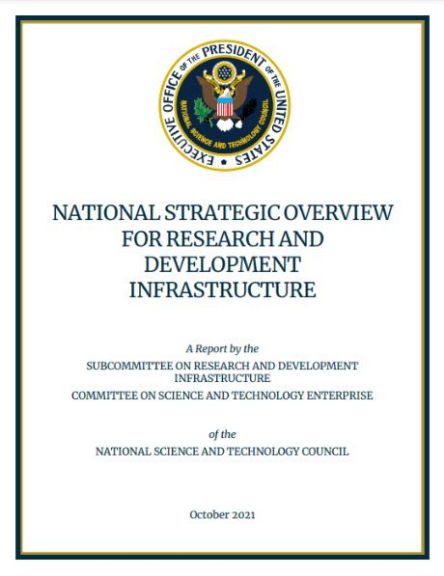
U.S. DEPARTMENT OF ENERGY
Office of Science

ESnet

*DOE Office of Science facilities also support other collaborations, e.g., LHC, LSST, etc

Interconnectivity and integration of instrumentation, data and computing have been explicitly recognized as strategic requirements for national R&D

The 2021 National Strategic Overview from the Subcommittee on Research and Development Infrastructure formally redefined "federal R&D Infrastructure" to now include computing, data, and networking facilities, resources and services.

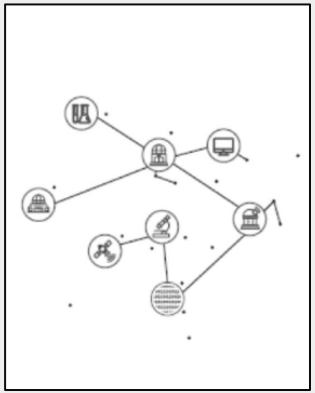

NATIONAL STRATEGIC OVERVIEW FOR RESEARCH AND DEVELOPMENT INFRASTRUCTURE

A Report by the
SUBCOMMITTEE ON RESEARCH AND DEVELOPMENT INFRASTRUCTURE
COMMITTEE ON SCIENCE AND TECHNOLOGY ENTERPRISE

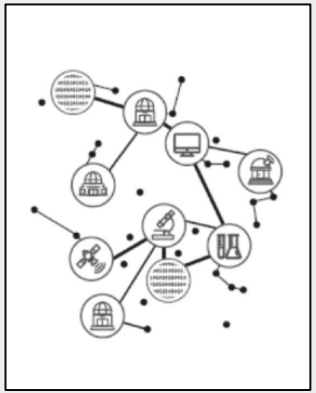
of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

October 2021

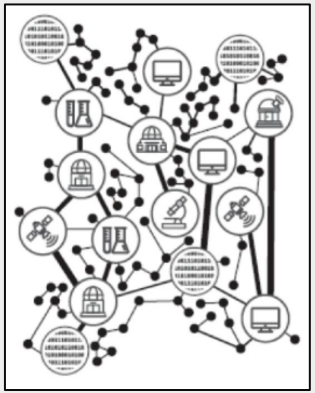
"R&D continues to shift from smaller to bigger science, driven in large part by advances in computing and other research cyberinfrastructure, which interlink[s] research data, analytics, … and experimental instrumentation."



Past



Present



Future

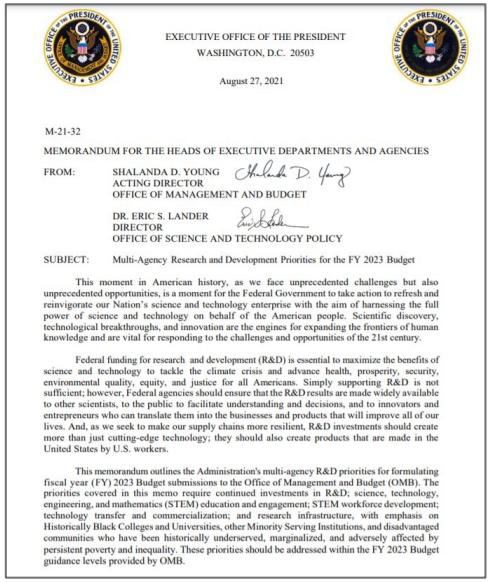
U.S. DEPARTMENT OF ENERGY | Office of Science

ESnet

https://www.whitehouse.gov/wp-content/uploads/2021/10/NSTC-NSO-RDI-_REV_FINAL-10-2021.pdf

Brown et al

National imperatives for US leadership in strategic areas all require an interlinked ecosystem of instruments, compute, data



EXECUTIVE OFFICE OF THE PRESIDENT
WASHINGTON, D.C. 20503

August 27, 2021

M-21-32

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: SHALANDA D. YOUNG
ACTING DIRECTOR
OFFICE OF MANAGEMENT AND BUDGET

DR. ERIC S. LANDER
DIRECTOR
OFFICE OF SCIENCE AND TECHNOLOGY POLICY

SUBJECT: Multi-Agency Research and Development Priorities for the FY 2023 Budget

This moment in American history, as we face unprecedented challenges but also unprecedented opportunities, is a moment for the Federal Government to take action to refresh and reinvigorate our Nation's science and technology enterprise with the aim of harnessing the full power of science and technology on behalf of the American people. Scientific discovery, technological breakthroughs, and innovation are the engines for expanding the frontiers of human knowledge and are vital for responding to the challenges and opportunities of the 21st century.

Federal funding for research and development (R&D) is essential to maximize the benefits of science and technology to tackle the climate crisis and advance health, prosperity, security, environmental quality, equity, and justice for all Americans. Simply supporting R&D is not sufficient; however, Federal agencies should ensure that the R&D results are made widely available to other scientists, to the public to facilitate understanding and decisions, and to innovators and entrepreneurs who can translate them into the businesses and products that will improve all of our lives. And, as we seek to make our supply chains more resilient, R&D investments should create more than just cutting-edge technology; they should also create products that are made in the United States by U.S. workers.

This memorandum outlines the Administration's multi-agency R&D priorities for formulating fiscal year (FY) 2023 Budget submissions to the Office of Management and Budget (OMB). The priorities covered in this memo require continued investments in R&D; science, technology, engineering, and mathematics (STEM) education and engagement; STEM workforce development; technology transfer and commercialization; and research infrastructure, with emphasis on Historically Black Colleges and Universities, other Minority Serving Institutions, and disadvantaged communities who have been historically underserved, marginalized, and adversely affected by persistent poverty and inequality. These priorities should be addressed within the FY 2023 Budget guidance levels provided by OMB.

Multi-Agency R&D Priorities for the FY 2023 Budget (August 27, 2021)

Pandemic readiness and prevention.

Tackling climate change.

Catalyzing research and innovation in critical and emerging technologies.

- *Artificial intelligence (AI)*
- *Quantum Information Science (QIS)*
- *High-performance computing*
- *Advanced communications technologies, Microelectronics, Biotechnology, Robotics, Space technologies.*

Innovation for equity.

National security and economic resilience.

Emerging interagency concepts and initiatives towards a ***National Research Ecosystem***

DOE is uniquely placed to lead and shape the national conversation and initiatives.

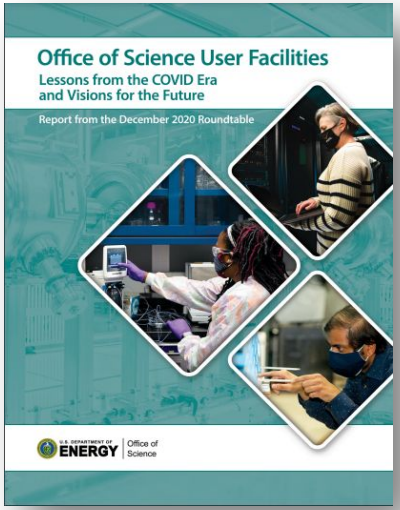
U.S. DEPARTMENT OF ENERGY | Office of Science

ESnet

Brown et al

The Roundtable report contained key insights regarding the need for distributed and interoperable computing resources

"Facilities may be able to deploy a distributed network of connected and interoperable computing resources that enable all scales of computing, data exploration, and analysis."



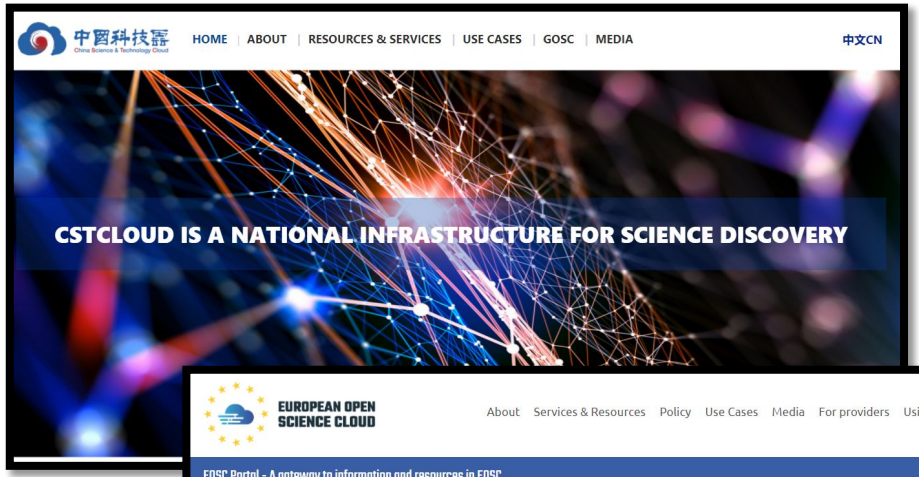
- The pandemic highlighted the importance of **secure remote collaboration and facilitated access to data and computing resources**.
- **Data management and data stewardship** present another critical opportunity. User facilities generate exabytes of unique, irreplaceable data which must be managed, curated and made available for analysis and computation.
- This requires a host of **user-connecting operational approaches and technologies** such as high-quality interfaces, collaboration tools, federated identity management, automation of experiments and workflows, and more.
- "With collaboration among all its user facilities, DOE SC is in a position to facilitate all aspects of the data lifecycle across its facility complex, including simulations, experiment design, data generated at scientific instruments, data analysis, and data archiving for future use.
- "Seamlessly connecting a user with data and computing enables more uniform and egalitarian data exploration and analysis capabilities."

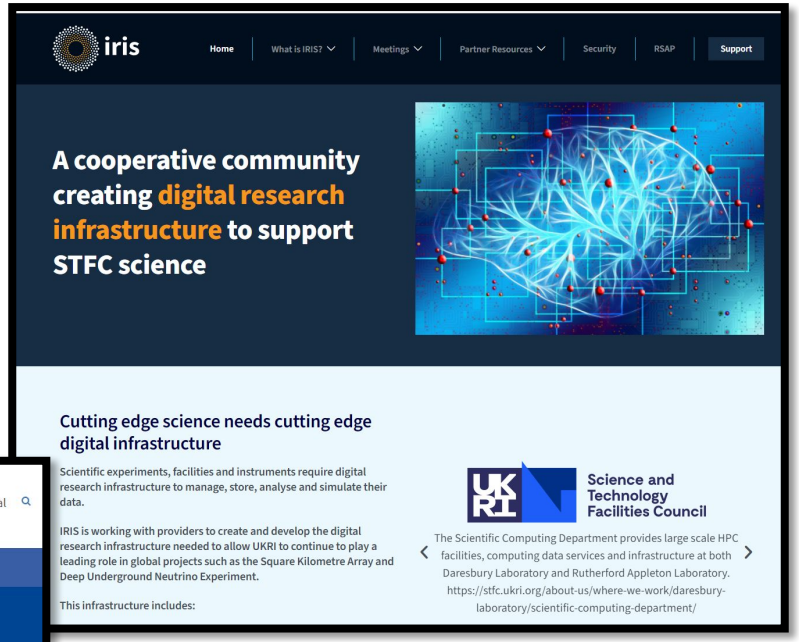
U.S. DEPARTMENT OF ENERGY | Office of Science

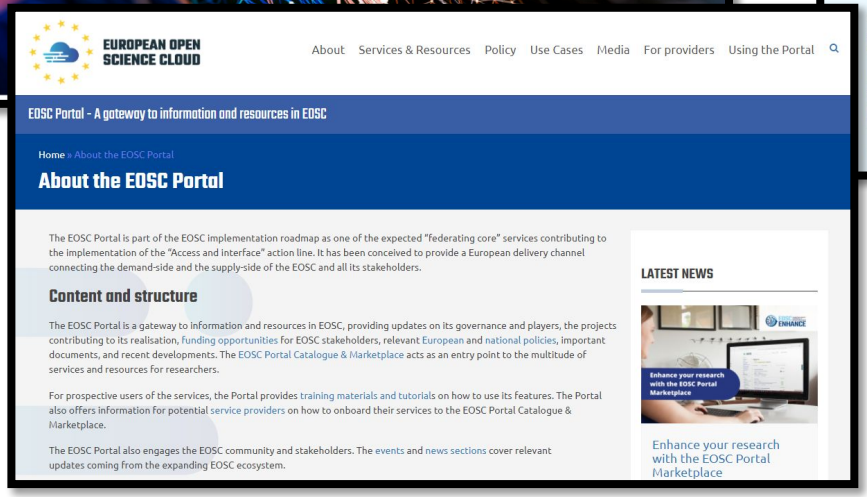
ESnet

Brown et al

International efforts in Integrated Research Infrastructure are expanding too







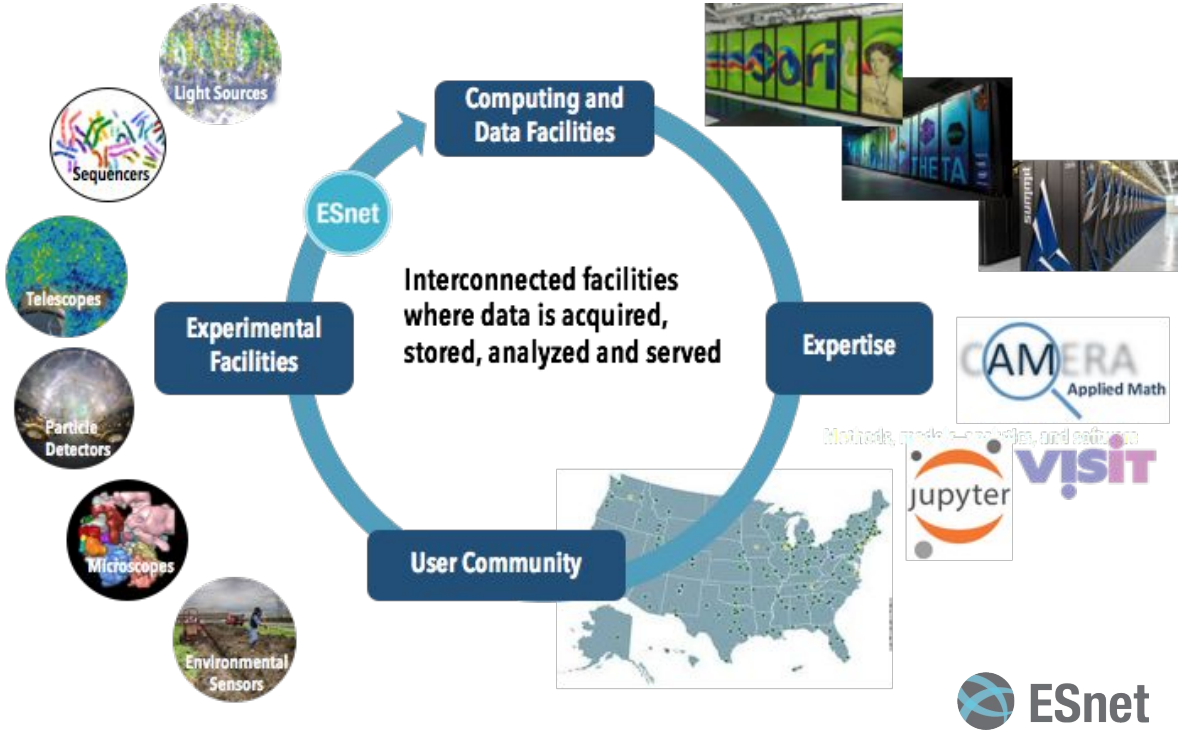
- China Science and Technology Cloud
- European Open Science Cloud
- IRIS UKRI SFTC initiative

ESnet

Brown et al

Superfacility: A model to integrate experimental, computational, networking, and storage facilities for reproducible science

Enabling new discoveries by coupling experimental science with extreme scale data analysis and simulations



Bard et al

What does this mean for networks*?

Promoting networks as "first class" resources, similar to instruments, compute and storage, e.g.,

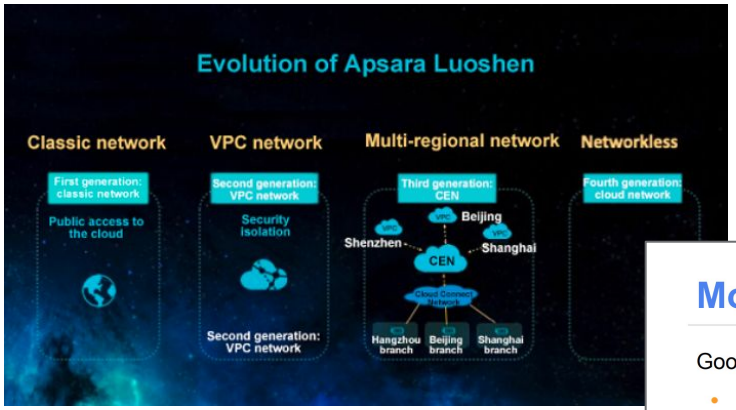
- Accessibility
  - Security frameworks for accessing (selected) services
  - APIs to interact with services
- Controllability
  - Resource/service selection/negotiation
  - Service scheduling
- Transparency
  - Resource (general) availability
  - Service (specific) status

Networking is an end-to-end service, inter-domain interoperability and service consistency is critical!

ESnet

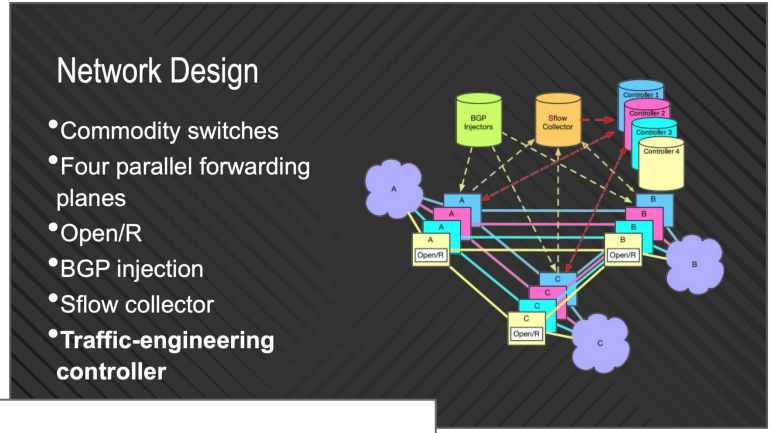
Hyperscalers get it*!

Alibaba Cloud's Apsara Luoshen



Z. Zong, "Apsara Luoshen, a High Performance Network Engine that Drives Alibaba Cloud", GTNC 2018, Nov 15, 2018

Facebook Express Backbone (EBB)



H. Kwok, "Express Backbone: Moving Fast with Facebook's Long-Haul Network", Networking @Scale 2017, July 9, 2015

Google B4 WAN

More Than the Sum of Parts

Google

Google Networking works together as an integrated whole

- **B4: WAN interconnect**
- GGC: edge presence
- Jupiter: building scale datacenter network
- Freedome: campus-level interconnect
- Andromeda: isolated, high-performance slices of the physical network

Publications in **INFOCOM 2012, SIGCOMM 2013, SIGCOMM 2014, CoNEXT 2014, EuroSys 2014, SIGCOMM 2015**

planet

S. Mandal, "Lessons Learned from B4, Google's SDN WAN", 2015 USENIX ATC'15, July 9, 2015

Deploying a vanilla best-effort delivery network is not optimal!

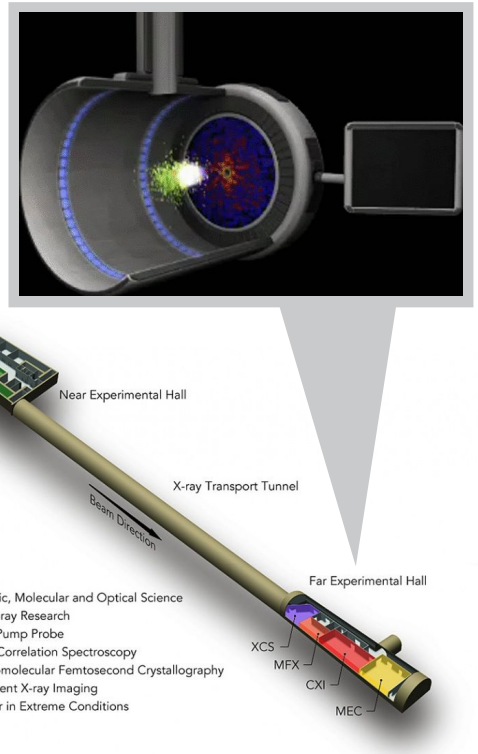
**NB: Solutions deployed are only within a single administrative domain*

ESnet

Superfacility Use Case 1 - Fast feedback to adjust experiment parameters

Linac Coherent Light Source (LCLS)

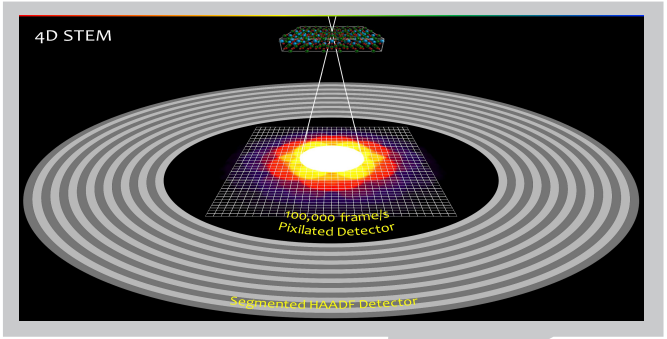
- Ultrafast X-ray pulses from LCLS are used like flashes from a high-speed strobe light, producing stop-action movies of atoms and molecules.
- Both data processing and scientific interpretation demand intensive computational analysis.
- Leverage HPC resources to process initial results to verify proper alignment. Misalignment results in wasted experiment.



ESnet

Amedeo et al

Superfacility Use Case 2 - Reduction or elimination of site local compute and storage



National Center for Electron Microscopy (NCEM) - 4D STEM Development

- NCEM is developing a high frame rate (100KHz) 4D detector system to enable fast real-time data analysis of scanning diffraction experiments in scanning transmission electron microscopy (STEM)
- High frame rate development aims to improve scanning diffraction experiments and will be installed on the Transmission Electron Aberration-corrected Microscope (TEAM)
- Direct high speed data transfer of raw image sets from microscope to HPC for online analysis and storage of data.

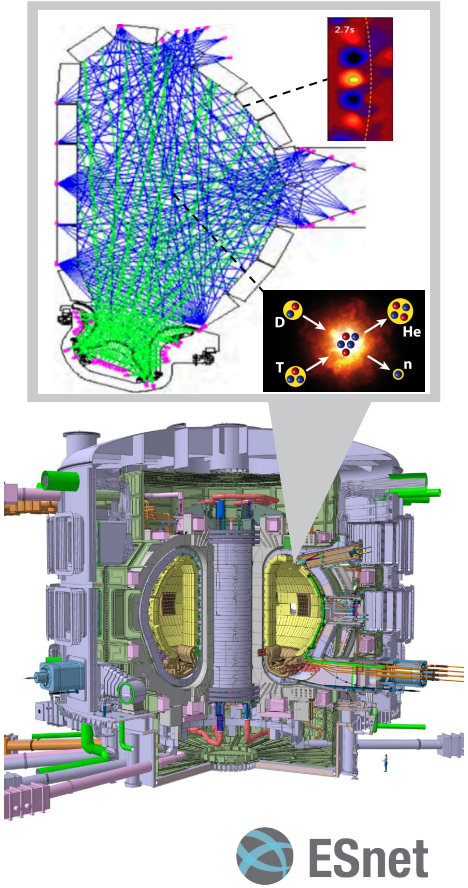
ESnet

Bard et al

Superfacility Use Case 3 - Real-time analysis for monitoring and control

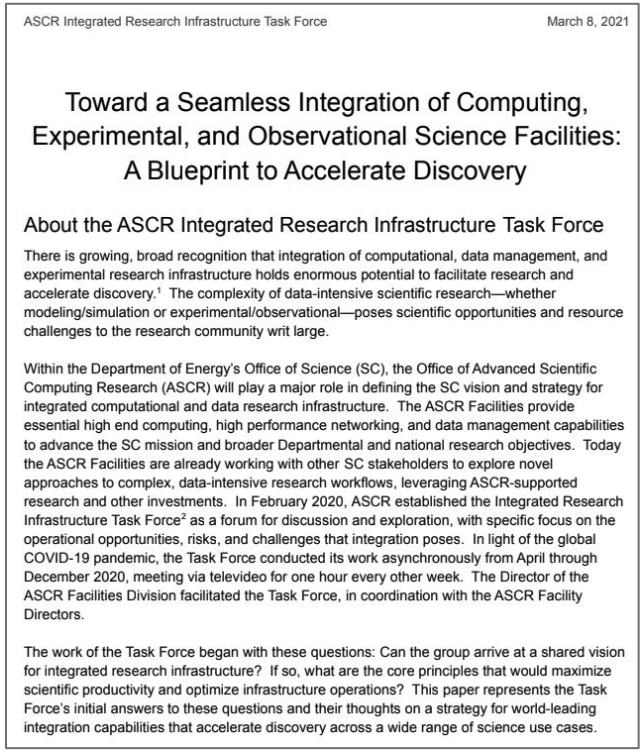
ITER (*originally* International Thermonuclear Experimental Reactor)

- First fusion device to produce net energy and maintain fusion for long periods of time with ten times the plasma volume of largest machine operating today.
- ITER is designed to produce **500 MW** of fusion power from 50 MW of input heating power.
- Real-time analysis and control is needed to flag potentially dangerous issues within the reactor and mitigate accordingly.



ESnet

DOE ASCR IRI Task Force contemplated operational models and guiding principles.



ASCR Integrated Research Infrastructure Task Force March 8, 2021

Toward a Seamless Integration of Computing, Experimental, and Observational Science Facilities: A Blueprint to Accelerate Discovery

About the ASCR Integrated Research Infrastructure Task Force

There is growing, broad recognition that integration of computational, data management, and experimental research infrastructure holds enormous potential to facilitate research and accelerate discovery.[1] The complexity of data-intensive scientific research—whether modeling/simulation or experimental/observational—poses scientific opportunities and resource challenges to the research community writ large.

Within the Department of Energy's Office of Science (SC), the Office of Advanced Scientific Computing Research (ASCR) will play a major role in defining the SC vision and strategy for integrated computational and data research infrastructure. The ASCR Facilities provide essential high end computing, high performance networking, and data management capabilities to advance the SC mission and broader Departmental and national research objectives. Today the ASCR Facilities are already working with other SC stakeholders to explore novel approaches to complex, data-intensive research workflows, leveraging ASCR-supported research and other investments. In February 2020, ASCR established the Integrated Research Infrastructure Task Force[2] as a forum for discussion and exploration, with specific focus on the operational opportunities, risks, and challenges that integration poses. In light of the global COVID-19 pandemic, the Task Force conducted its work asynchronously from April through December 2020, meeting via televideo for one hour every other week. The Director of the ASCR Facilities Division facilitated the Task Force, in coordination with the ASCR Facility Directors.

The work of the Task Force began with these questions: Can the group arrive at a shared vision for integrated research infrastructure? If so, what are the core principles that would maximize scientific productivity and optimize infrastructure operations? This paper represents the Task Force's initial answers to these questions and their thoughts on a strategy for world-leading integration capabilities that accelerate discovery across a wide range of science use cases.

B. Brown, C. Adams, K. Antypas, D. Bard, S. Canon, E. Dart, C. Guok, E. Kissel, E. Lancon, B. Messer, S. Oral, J. Ramprakash, A. Shankar, T. Uram

"Our vision is to integrate across scientific facilities to accelerate scientific discovery through productive data management and analysis, via the delivery of pervasive, composable, and easily usable computational and data services."

Areas

- AL Allocations
- AC Accounts
- DA Data
- AP Applications
- SC Scheduling
- WF Workflows
- PB Publication
- AR Archiving

Principles

Flexibility. Assembly of resource workflows is facile; complexity is concealed

Performance. Default behavior is performant, without arcane requirements

Scalability. Data capabilities without excessive customizations

Transparency. Security, authentication, authorization should support automation

Interoperability. Services should extend outside the DOE environment

Resiliency. Workloads are sustained across planned and unplanned events

Extensibility. Designed to adapt and grow to meet unknown future needs

Engagement. Promotes co-design, cooperation, partnership

Cybersecurity. Security for facilities and users is essential.

Questions…

Chin Guok <chin@es.net>

ESnet