U.S. DEPARTMENT OF ENERGY | Office of Science

ESnet

THE DOE OFFICE OF SCIENCE

ESnet Requirements Review Program Through the IRI Lens

A Meta-Analysis of Workflow Patterns Across DOE Office of Science Programs

FINAL REPORT
2023

ESnet Requirements Review Program Through the IRI Lens

A Meta-Analysis of Workflow Patterns Across DOE Office of Science Programs

ESnet is funded by the US DOE, Office of Science, Office of Advanced Scientific Computing Research (ASCR). Carol Hawk is the ESnet Program Manager. ESnet is operated by Lawrence Berkeley National Laboratory (Berkeley Lab), which is operated by the University of California for the US Department of Energy under contract DE-AC02-05CH11231. This work was supported by ASCR.

Authors: Dart, Eli; Zurawski, Jason; Hawk, Carol; Brown, Benjamin L.; Monga, Inder.

Cover Graphic: Boyd Productions, LLC
Creative: Samantha Trieu, Creative Services, Lawrence Berkeley National Laboratory
Editorial: Bonnie Powell, Computing Sciences Area, Lawrence Berkeley National Laboratory

DOI: 10.2172/2008205
LBNL-2001552
24-SN-35053

Table of Contents

Executive Summary

The Department of Energy (DOE) ensures America's security and prosperity by addressing its energy, environmental, and nuclear challenges through transformative science and technology solutions[1]. The DOE's Office of Science (SC) delivers groundbreaking scientific discoveries and major scientific tools that transform our understanding of nature and advance the energy, economic, and national security of the United States[2]. The SC's programs[3] advance DOE mission science across a wide range of disciplines and have developed the research infrastructure needed to remain at the forefront of scientific discovery.

The DOE SC's world-class research infrastructure — exemplified by the 28 SC scientific user facilities[4] — provides the research community with premier observational, experimental, computational, and network capabilities. Each user facility is designed to provide unique capabilities to advance core DOE mission science for its sponsor SC program and to stimulate a rich discovery and innovation ecosystem.

Research communities gather and flourish around each user facility, bringing together diverse perspectives. A hallmark of many facilities is the large population of students, postdoctoral researchers, and early-career scientists who contribute as full-fledged users. These facility staff and users collaborate over years to devise new approaches to utilizing the user facility's core capabilities. The history of the SC user facilities has many examples of wildly inventive researchers challenging operational orthodoxy to pioneer new vistas of discovery; for example, the use of the synchrotron X-ray light sources for study of proteins and other large biological molecules. This continual reinvention of the practice of science — as users and staff forge novel approaches expressed in research workflows — unlocks new discoveries and propels scientific progress.

Within this research ecosystem, the high performance computing (HPC) and networking user facilities stewarded by SC's Advanced Scientific Computing Research (ASCR) program play a dynamic cross-cutting role, enabling complex workflows demanding high performance data, networking, and computing solutions. The DOE SC's three HPC user facilities and the Energy Sciences Network (ESnet) high-performance research network serve all of the SC's programs as well as the global research community. Argonne Leadership Computing Facility (ALCF), the National Energy Research Scientific Computing Center (NERSC), and Oak Ridge Leadership Computing Facility (OLCF) conceive, build, and provide access to a range of supercomputing, advanced computing, and large-scale data-infrastructure platforms, while ESnet interconnects DOE SC research infrastructure and enables seamless exchange of scientific data. All four facilities operate testbeds to expand the frontiers of computing and networking research. Together, the ASCR facilities enterprise seeks to understand and meet the needs and requirements across SC and DOE domain science programs and priority efforts, highlighted by the formal requirements reviews (RRs) methodology.

In recent years, the research communities around the SC user facilities have begun experimenting with and demanding solutions integrated with HPC and data infrastructure. This rise of integrated-science approaches is documented in many community and high-level government reports. At the dawn of the era of exascale science and the acceleration of artificial intelligence (AI) innovation, there is a broad need for integrated computational, data, and networking solutions.

In response to these drivers, DOE has developed a vision for an Integrated Research Infrastructure (IRI): **To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation.**

1 https://www.energy.gov/about-us
2 https://www.energy.gov/science/mission
3 https://science.osti.gov/Programs
4 https://science.osti.gov/User-Facilities

The IRI vision is fundamentally about establishing new data-management and computational paradigms within which DOE SC user facilities and their research communities work together to improve existing capabilities and create new possibilities by building bridges across traditional silos. Implementation of IRI solutions will give researchers simple and powerful tools with which to implement multi-facility research data workflows.

In 2022, SC leadership directed the Advanced Scientific Computing Research (ASCR) program to conduct the Integrated Research Infrastructure Architecture Blueprint Activity (IRI ABA) to produce a reference framework to inform a coordinated, SC-wide strategy for IRI. This activity convened the SC science programs and more than 150 DOE national laboratory experts from all 28 SC user facilities across 13 national laboratories to consider the technological, policy, and sociological challenges to implementing IRI.

Through a series of cross-cutting sprint exercises facilitated by the IRI ABA leadership group and peer facilitators, participants produced an IRI Framework[5] based on the IRI Vision and comprising:

- IRI Science Patterns spanning DOE science domains;
- IRI Practice Areas needed for implementation;
- IRI blueprints that connect Patterns and Practice Areas;
- Overarching principles for realizing the DOE-wide IRI ecosystem.

The resulting IRI framework and blueprints provide the conceptual foundations to move forward with organized, coordinated DOE implementation efforts. The next step is to identify urgencies and ripe areas for focused efforts that uplift multiple communities.

Upon completion of the IRI ABA framework, ESnet applied the IRI Science Patterns lens and undertook a meta-analysis of ESnet's Requirements Reviews (RRs), the core strategic planning documents that animate the multi-year partnerships between ESnet and five of the DOE SC programs. Between 2019 and 2023, ESnet completed a new round of RRs with the following SC programs: Nuclear Physics (2019-20), High Energy Physics (2020-21), Fusion Energy Sciences (2021-22), Basic Energy Sciences (2021-22), and Biological and Environmental Research (2022-23). Together these ESnet RRs provide a rich trove of insights into opportunities for immediate IRI progress and investment.

Our meta-analysis of 74 high-priority case studies reveals that:

- There are a significant number of research workflows spanning materials science, fusion energy, nuclear physics, and biological science that have a similar structure. Creation of common software components to improve these workflows' performance and scalability will benefit researchers in all of these areas.
- There is broad opportunity to accelerate scientific productivity and scientific output across DOE facilities by integrating them with each other and with high performance computing and networking.
- The ESnet RRs' blending of retrospective and prospective insight affirms that the IRI patterns are persistent across time and likely to persist into the future, offering value as a basis for analysis and strategic planning going forward.

5 https://doi.org/10.2172/1984466

The Integrated Research Infrastructure Architecture Blueprint Activity (IRI ABA)

The IRI Architecture Blueprint Activity[6] (IRI ABA) brought together domain experts from all DOE SC programs to look for common patterns within diverse workflows across a range of scientific disciplines. Participants identified three common patterns:

- **Time-Sensitive Patterns,** which have urgency, requiring real-time or end-to-end performance with high reliability, e.g., for timely decision-making, experiment steering, and virtual proximity;
- **Data Integration–Intensive Patterns,** which require combining and analyzing data from multiple sources, e.g., sites, experiments, and/or computational runs;
- **Long-Term Campaign Patterns,** which require sustained access to resources over a long period to accomplish a well-defined objective.

IRI case studies that exhibit the Time-Sensitive Pattern tend to represent scientific domains that rely on low-latency, near-real-time workflows spanning multiple facilities. They often require high-throughput connection of experimental facilities with high performance computing (HPC) platforms: for example, materials science research that uses real-time simulation and analysis of experimental data to steer running experiments toward the most fruitful avenues of exploration, potentially using artificial intelligence/machine learning (AI/ML).

IRI case studies that exhibit the Data Integration-Intensive Pattern tend to represent scientific domains that use multi-modal data from a variety of different data sources, such as simulations, experiments, and observations. They require facilities designed to capture, compute, store, categorize, and share data from a wide range of experiments, as well as from long-running observational efforts that acquire data from hundreds of diverse instruments and sensors, preserving metadata and provenance. Examples include earth systems observational data acquired globally using diverse sensors, and light source experimental facilities that use simulated and empirical data together to advance materials science.

IRI case studies that exhibit the Long-Term Campaign Pattern tend to represent scientific domains driven by large collaborations centered around rare or unique scientific instruments designed, built, and operated over years or decades. They require sophisticated data management systems, effective large-scale collaboration structures, and the ability to use computing for data analysis over a period of many years. Examples include high energy physics collaborations, astronomical sky surveys, or multi-decadal studies of climate models.

The IRI ABA revealed a subset of experiments in many different fields (e.g., materials science, fusion energy, nuclear physics, and biology, among others) that all have a similar structure. These experiments observe specific processes in nature using high-precision detectors and move data from those detectors to computing resources for analysis. Many of these workflows are similar enough that it makes sense to explore the creation of common software components to improve their performance and scalability, regardless of the specific scientific field of a particular experiment.

More than 15 years ago, ESnet began documenting data movement between SC facilities and resources at routine intervals, in the form of ESnet Requirements Reviews conducted in partnership with each SC program. These documents and their strategic importance for SC are discussed in the next section.

6 https://www.osti.gov/biblio/1984466

ESnet Requirements Reviews

ESnet is SC's high-performance networking user facility, engineered and optimized for large-scale science. However, ESnet is much more than a high-capacity data network tailored to handle science's enormous data flows: the ESnet team's applied research results in a range of innovative services that seek to optimize scientific data acquisition, transport, placement, and sharing. ESnet interconnects all 17 DOE national laboratories, 28 user facilities, and peers with over 270 other research and education (R&E) and commercial networks within the U.S. and around the world. ESnet enables tens of thousands of scientists to transmit vast quantities of data quickly and without loss; to access instruments, high performance computing, and other scientific resources remotely; and to collaborate across the globe.

Beginning in 2007, ESnet has conducted an ESnet Requirements Review[7] (RR) regularly with each SC program. Each RR comprises an in-depth survey and discussion of existing and planned data-intensive science experiments, instruments, and user facilities and utilizes a case study methodology to ensure documentation of each SC program's highest priority use cases over the following five years. Although networking is the focus, the collected requirements are in many ways data-centric and provide a window into the data needs of each SC program, especially for cross-facility, multi-facility, or multi-institution workflows. ESnet's view "from the middle" across the breadth of SC programs makes the RR historical archive a strategic source of insight.

The ESnet RRs focus on data output, transfer, and storage needs as well as science workflows and collaborations. Each principal investigator from the SC program facility, collaboration, or project is asked to complete one or more case studies, with assistance from ESnet and laboratory networking representatives. The information is used to plan for the high-performance networking needs of the SC program in the upcoming five-year timeframe. Specifically, the resulting RR report positions ESnet to invest appropriately in network bandwidth and services by gathering information about:

- Major science experiments and facilities, both those currently in operation and in the pipeline;
- The volume of data produced now and anticipated in the future, with an emphasis on geographical location — where the data must be shared, computed, and/or stored;
- The network, software, and cloud resources currently in use, as well as any planned upgrades, additions, or improvements.

This strategic partnership between ESnet and each SC program has delivered far-reaching outcomes, such as highlighting:

- Workflow changes in CERN's Large Hadron Collider (LHC) experimental model that favored increased use of networking between the LHC data and computing facilities. This supported ESnet's trans-Atlantic extension to CERN.
- The fundamental challenges for data movement between facilities. This inspired creation of the Science DMZ model[8], use of Data Transfer Nodes[9], and the Globus[10] application.
- The importance of centralized technology resources at DOE HPC facilities as part of the experimental workflow. This innovation, made possible via the high-capacity resources of ESnet, resulted in new multi-facility use cases across the DOE SC ecosystem.

7 https://www.es.net/requirements/
8 E. Dart, L. Rotman, B. Tierney, M. Hester and J. Zurawski, "The Science DMZ: A network design pattern for data-intensive science," SC '13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, Denver, CO, USA, 2013, pp. 1-10, doi: 10.1145/2503210.2503245.
9 https://fasterdata.es.net/DTN/
10 https://www.globus.org/

IRI Meta-Analysis of Recent ESnet Requirements Reviews

Methodology

The meta-analysis applied the IRI framework devised in the IRI Architecture Blueprint Activity to the body of insight collected in the recent ESnet Requirements Reviews. ESnet evaluated each case study from the most recent RRs to determine which of the three IRI patterns, if any, the case study reflected. The meta-analysis process reviewed 74 case studies that spanned five DOE SC programmatic areas — Nuclear Physics (NP), High Energy Physics (HEP), Fusion Energy Sciences (FES), Basic Energy Sciences (BES), and Biological and Environmental Research (BER) — and were collected via the ESnet RR process between 2019 and 2023. Representing a cross-section of facilities and experiments, these case studies offer important insight into the process of science, collaboration space, and technology requirements. Many of these use cases highlighted multi-facility workflows; most were able to identify areas of friction related to the use of current technology, or adoption of new technology, as they scale operations into the coming years.

IRI Meta-Analysis Results

The meta-analysis showed that IRI patterns can reveal common features in programs or collaborations that might not otherwise be easily compared. Opportunities for common development become apparent when different workflows align within the common criteria of an IRI pattern. Importantly, while the three IRI patterns are very different from each other, each pattern is clearly associated with different scientific domains stewarded by different SC programs, revealing the persistent presence of the IRI-ABA patterns across the breadth of SC programs.

Key findings from the ESnet RR process that dovetail with the vision of IRI include the need to offer:

- long-term custodial storage of research data;
- flexibility in computing resources;
- intelligent middleware to manage data mobility;
- increasing performance in keeping with data scale; and
- scalability of human resources to design, implement, and operate cyberinfrastructure in the coming years.

The ESnet RR meta-analysis also identified gaps in the data infrastructure that supports the SC mission. Several of these gaps align with the IRI vision, including a growing need to:

- Establish data infrastructure that can manage the storage, cataloging, and sharing of research data with a wide community over time;
- Address workflows that are not portable across the ecosystem of resources, e.g. that constrained by limitations of hardware, software, and the policies surrounding access and use;
- Reduce operational friction that impacts productivity (e.g., lack of common tools, policies, or performance baselines); and
- Develop more multi-facility capabilities to better utilize SC resources and improve science outcomes.

The meta-analysis validates that all three IRI patterns exist across the SC programs (see Figure 1), that a significant number of use cases are dominated by one pattern, and that a significant number of use cases span more than one of the three patterns.

Insight 1: The meta-analysis indicates that nearly all use cases align with at least one IRI pattern.

The broad applicability of the IRI patterns to DOE SC use cases and workflows confirms the validity of the IRI patterns and illustrates their utility in reasoning about common solutions to science workflow challenges. Many use cases may have created a hardware and software platform to operate locally, because it was infeasible to coordinate it externally due to funding or technology constraints; future versions scaled this up until data volumes prevented further growth. The prevalence of inter-facility overlapping Time-Sensitive, Data Integration-Intensive, and Long-Term Campaign Patterns reflects this, and serves as a fertile ground for future cross-facility workflows delivered via IRI approaches.

Specific use cases that are dominated by one pattern and thus could be coordinated to accelerate progress are highlighted by:

- Growing collaborations between SC experimental facilities, ASCR HPC facilities, and ESnet to provide experimentalists with rapid or near-real-time data analysis to inform the next experimental run or shot, thereby enhancing productivity and revealing new insight. These workflows reflect the IRI Time-Sensitive Pattern.
- Increased use of simulations and experiments together in the scientific process; for instance, the use of simulations to guide experiments or to help interpret the data from detectors. These workflows reflect the IRI Data Integration-Intensive Pattern.
- Sky surveys, the LHC experiments, and similar large-scale collaborations that use very large facilities designed, built, and operated by big teams over years and decades to answer specific questions at the cutting edge of science. These workflows reflect the Long-Term Campaign Pattern.

Figure 1 shows the percentage of ESnet RR case studies that reflect a particular IRI pattern within each SC program. It is important to note that, although an IRI pattern may appear in only one or two use cases in an SC program's case study portfolio, that IRI pattern could be essential for achieving the program's mission objectives.

Note: The analysis represented in Figure 1 omits the two use cases that do not currently align with an IRI pattern due to existing workflow limitations, but which need integrated workflows and would benefit from alignment with an IRI pattern to accelerate progress.

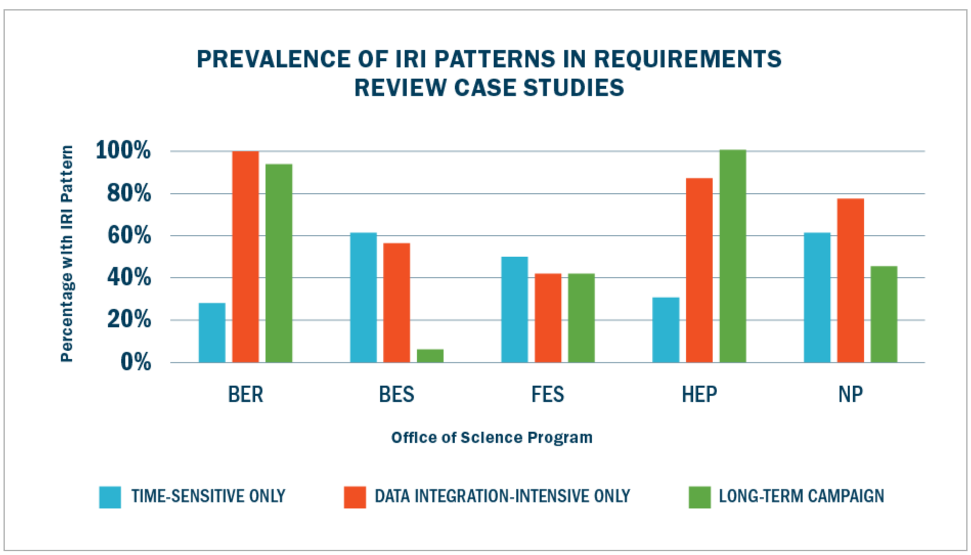

Figure 1: How each SC program case-study portfolio breaks down into percentage of IRI patterns.[11]

11 Excludes two case studies that did not currently align with an IRI pattern due to existing workflow limitations.

Insight 2: The meta-analysis shows that a majority of use cases align with more than one IRI pattern.

Figure 2 shows the percentage of the analyzed ESnet RR case studies that reflect a single IRI pattern or combinations of IRI patterns. More than a third of use cases reflect both the Long-Term Campaign and the Data Integration-Intensive patterns. However, only one use case reflects the Long-Term Campaign pattern alone, and 4 percent of use cases align with both the Long-Term Campaign pattern and the Time-Sensitive pattern. This suggests that if a workflow aligns with the Long-Term Campaign pattern, it is highly likely that that workflow will also be Data Integration-Intensive, indicating fertile opportunities for common solutions and leveraging of lessons learned.

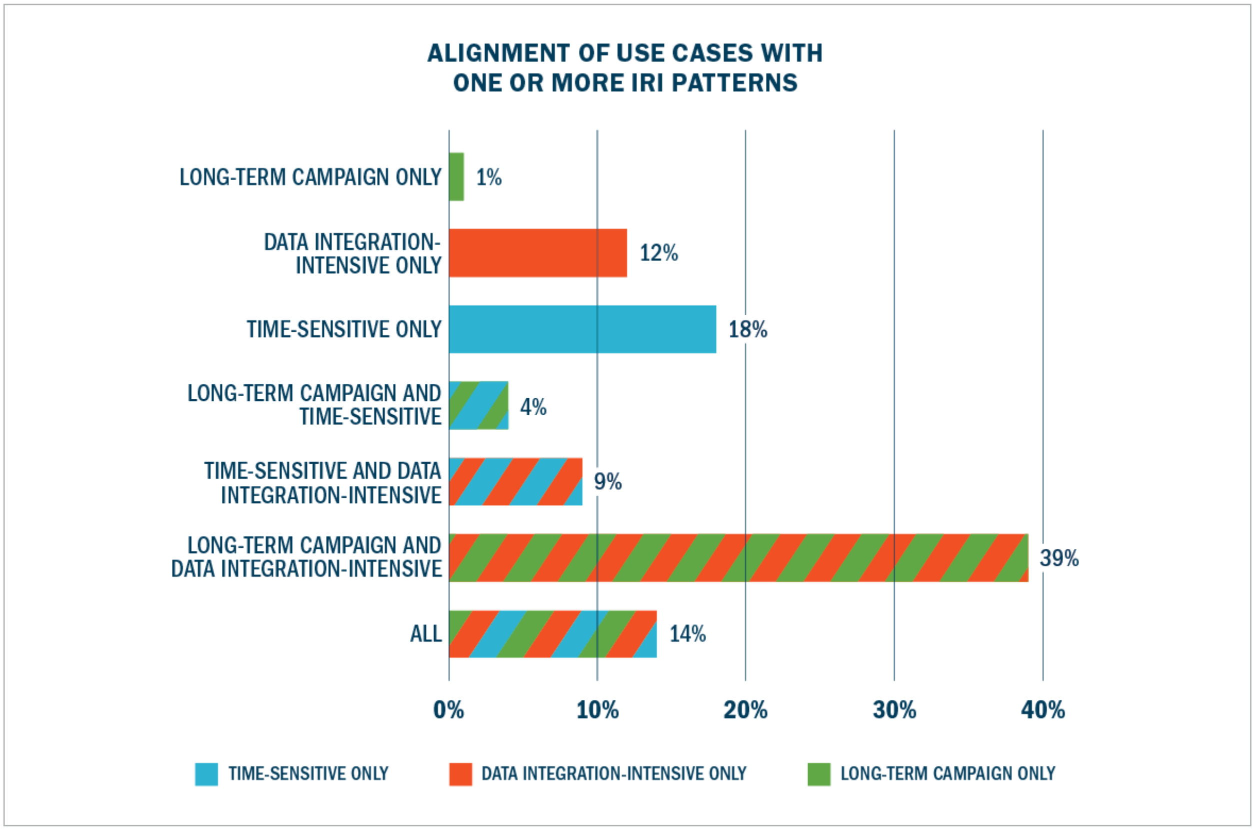

Figure 2: The percentage of ESnet RR case studies that reflect a particular IRI pattern or combination of patterns.[12]

12 Excludes two case studies that did not currently align with an IRI pattern due to existing workflow limitations.

Insight 3: Looking across the portfolio of use cases for each SC program reveals previously unseen commonalities.

Figure 3 shows the number of use cases that each SC program contributes to the percentages shown in Figure 2. It is of interest to note that most of the use cases that align with both the Long-Term Campaign and Data Integration-Intensive Patterns originate from the SC programs Biological and Environmental Research and High Energy Physics, signaling the potential for these programs — not typically considered as close cousins — to explore future IRI solutions based on the aggregation of insight from past experiences and investments.

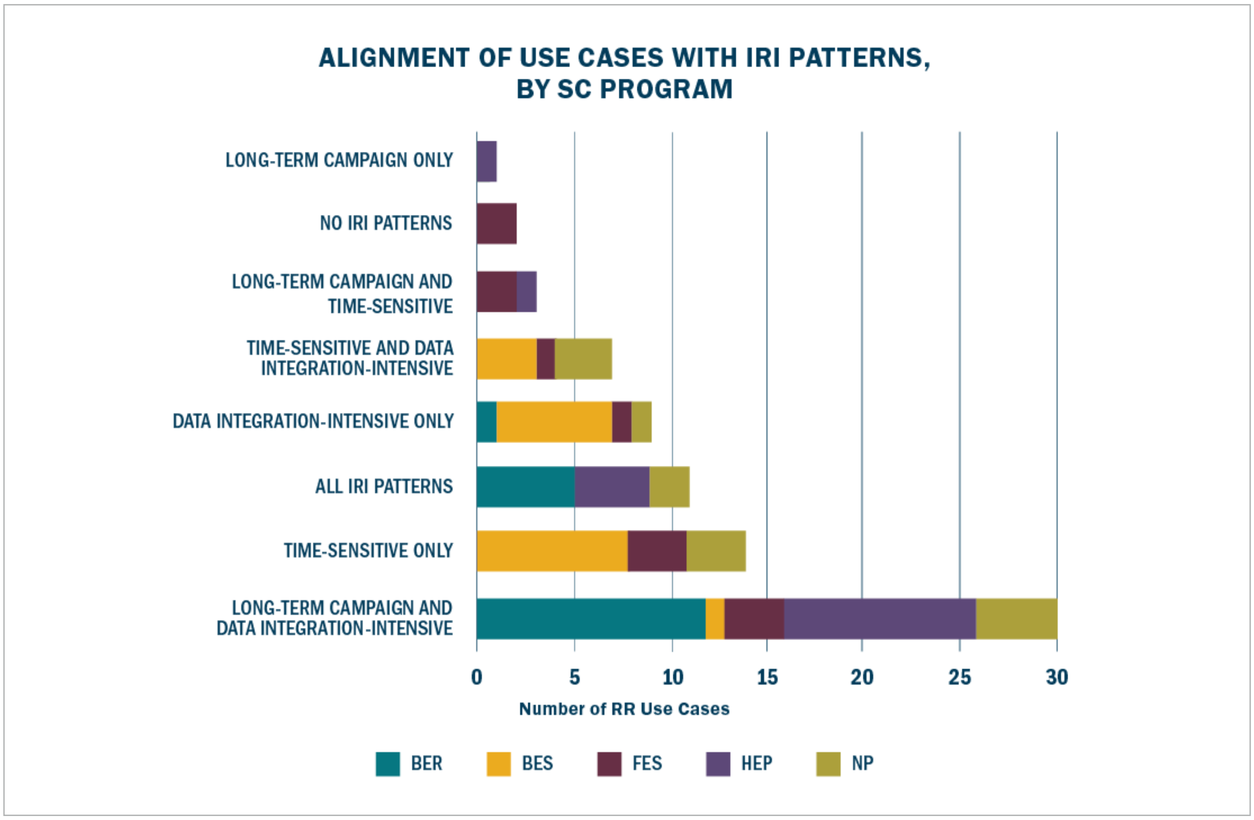

Figure 3: The number of ESnet Requirements Reviews (RR) case studies that reflect a particular IRI pattern or combination of patterns, for each SC program.

Summary Conclusions

The IRI Architecture Blueprint Activity, the ESnet Requirements Reviews, and this meta-analysis applying the IRI framework to ESnet RR case studies reveal an intensifying need across all SC programs and DOE user facilities for higher-performance integrated workflows that result from better integration of research infrastructure. This meta-analysis demonstrates that this increasing need is pervasive and growing across the SC programs in recent years. The IRI patterns offer a strategic framework that invites common solutions to shared challenges and opens the possibility of new capabilities that reside at the interfaces and within the intersections of an integrated research infrastructure.

The IRI framework offers a new paradigm for organizing coordinated efforts, through which stakeholders will achieve faster progress towards their individual goals. This meta-analysis of the ESnet RR case studies through the IRI framework lens underscores several strategic opportunities for computing, data, and network infrastructure design:

- Near real-time data analysis by HPC platforms will steer experiments;
- Shared data-management solutions will allow meaningful integration of diverse datasets, expanding and deepening the context within which these data are interpreted.
- Large collaborations will gain insights from diverse data integrated in meaningful ways and acquired over decades.

The meta-analysis points the way to new opportunities to accelerate progress for the SC programs' highest-priority activities. Through implementation and continual refinement of the IRI framework, the DOE research community's ability to recognize, act on, and achieve future integrated science possibilities not yet envisioned will become increasingly clear. The IRI framework will be invaluable as the SC research community works together to bridge traditional silos and create opportunities for scientific discovery across all SC programs.

Appendix A: Nuclear Physics

NP Case Study	Time-Sensitive	Data Integration-Intensive	Long-Term Campaign
Thomas Jefferson National Accelerator Facility (JLab): Facilities	X	X	
JLab: Measurement of a Lepton-Lepton Electroweak Reaction (MOLLER)	X	X	
JLab: The Solenoidal Large Intensity Device (SoLID)	X		
JLab: Theory Group & Lattice Quantum Chromodynamics (LQCD)		X	X
Facility for Rare Isotope Beams (FRIB)	X	X	
Gamma-Ray Energy Tracking Array (GRETA)[13]	(X)		
Argonne National Laboratory (ANL): Gammasphere / Argonne Tandem Linear Accelerator System (ATLAS)	X		
ANL: CEBAF Large Acceptance Spectrometer for 12 GeV (CLAS12) / Electron-Ion Collider (EIC)		X	
Brookhaven National Laboratory (BNL): The RHIC and ATLAS Computing Facility (RACF)		X	X
BNL: The Solenoidal Tracker At RHIC (STAR)	X	X	X
BNL: Pioneering High-Energy Nuclear Interaction eXperiment (PHENIX) / sPHENIX	X	X	X
Compact Muon Solenoid (CMS) Heavy Ion Experimentation		X	X
ALICE (A Large Ion Collider Experiment) Project and ALICE-USA Computing		X	X

Table 1: Categorization of case studies from the Nuclear Physics (NP) program office.

13 Workflows are designed with the capability to operate with a Time-Sensitive Pattern

Reviewing the NP case studies through the IRI criteria leads to several general observations:

- **Time-Sensitive Pattern:** NP experiments frequently have a series of time-dependent steps. These typically include calibration of instruments, acquisition and filtering of data, and allocation of resources for storage and computation. Due to increases in data volume, the later steps are growing beyond the capabilities of in-house resources; as a whole, the community has the use of scalable processing through grid computation methods. When local resources are insufficient, the use of grid computing can be leveraged to acquire capabilities from other sources.
- **Data Integration-Intensive Pattern:** NP data sources come in two main forms — observations from sensors within experimental machinery and simulations of the experimental outcome. During active experimental runs, these multi-modal data requirements place pressure on the availability of computing, storage, and networking resources.
- **Long-Term Campaign Pattern:** A number of experiments have a rich multi-decadal run time and may re-process historic results, using this data as input into the creation of simulations or to train AI/ML models.

Appendix B: High Energy Physics

HEP Case Study	Time-Sensitive	Data Integration-Intensive	Long-Term Campaign
Cosmological Simulation Research		■	■
Dark Energy Science Collaboration (DESC)			■
Dark Energy Spectroscopic Instrument (DESI)	■		■
The Vera C. Rubin Observatory (Rubin Observatory) and the Legacy Survey of Space and Time (LSST)	■	■	■
Cosmic Microwave Background — Stage 4 (CMB-S4)	■	■	■
LZ (LUX-ZEPLIN) Dark Matter Experiment	■	■	■
Muon experimentation at Fermilab: Muon G minus two (g-2)		■	■
Muon experimentation at Fermilab: Muon-to-electron-conversion experiment (Mu2e)		■	■
Belle II experiment		■	■
Neutrino experiments at Fermilab: Short-Baseline Neutrino Program (SBN)		■	■
Neutrino experiments at Fermilab: Deep Underground Neutrino Experiment (DUNE)	■	■	■
Large Hadron Collider (LHC) experimentation and operation		■	■
LHC experimentation and operation: ATLAS (A Toroidal LHC ApparatuS) experiment		■	■
LHC experimentation and operation: Compact Muon Solenoid (CMS) experiment		■	■
LHC experimentation and operation: LHC operations		■	■
LHC experimentation and operation: High-luminosity (HL) era of the LHC		■	■

Table 2: Categorization of case studies from the High Energy Physics (HEP) program office.

Reviewing the HEP case studies through the IRI lens leads to several general observations:

- **Time-Sensitive Pattern:** A limited number of HEP uses cases require time-sensitive processes, typically related to observation of rare phenomena (e.g., supernovae) or possibly due to limited computational and storage resources at an experimental site and a requirement to operate in a multi-facility mode at all times.
- **Data Integration-Intensive Pattern:** Most HEP use cases combine observations from multiple sensors, along with coupling observed results with those of simulation. These multi-modal data requirements place pressure on the availability of computing, storage, and networking resources during active experimental runs.
- **Long-Term Campaign Pattern:** A number of experiments have a rich multi-decadal run time and may re-process historic observed results, or build to model and simulation models by increasing size, scale, and intensity.

Appendix C: Fusion Energy Sciences

FES Case Study	Time-Sensitive	Data Integration-Intensive	Long-Term Campaign
International fusion collaborations	■		■
Remote observation and participation of fusion facilities	■	■	
General Atomics: DIII-D National Fusion Facility	■		
MIT Plasma Science and Fusion Center (PSFC)		■	■
Princeton Plasma Physics Laboratory (PPPL)	■		■
Planning for ITER operation		■	■
Public-private partnerships in fusion research.			
Material Plasma Exposure eXperiment (MPEX) at Oak Ridge National Laboratory (ORNL)		■	
Matter in Extreme Conditions (MEC) Experiment at the SLAC National Accelerator Laboratory (SLAC)	■		
LaserNetUS Program			
Multi-facility FES workflows	■		
Whole-device modeling (WDM) and FES high-performance computing (HPC) activities.		■	■

Table 3: Categorization of case studies from the Fusion Energy Sciences (FES) program office.

Reviewing the FES case studies through the IRI criteria leads to several general observations:

- **Time-Sensitive Pattern:** With limited facilities that are capable of performing FES experiments, many use cases rely on remote observation and control of instrumentation. This implies a strong tie to network resources to facilitate domestic and international collaborations. FES experimentation is by nature time dependent, since a single "shot" of a reactor produces a burst of data that must be quickly analyzed, with the output of the analysis supporting decision making for the configuration of the next shot.
- **Data Integration-Intensive Pattern:** FES relies on data acquisition from multiple sources and is beginning to explore how simulation may be used during the experimental process to improve outcomes.
- **Long-Term Campaign Pattern:** While not as common across the FES ecosystem, a number of simulation and modeling efforts, along with historical data from major FES facilities, retain and routinely use previous datasets as a part of the scientific process.

Appendix D: Basic Energy Sciences

BES Case Study	Time-Sensitive	Data Integration-Intensive	Long-Term Campaign
Advanced Light Source (ALS)	■		
Advanced Photon Source (APS)	■		
National Synchrotron Light Source II (NSLS-II)	■		
Linac Coherent Light Source (LCLS)	■	■	
Stanford Synchrotron Radiation Light Source (SSRL)	■		
High Flux Isotope Reactor (HFIR) and Spallation Neutron Source (SNS)	■		
Center for Functional Nanomaterials (CFN)		■	
Center for Integrated Nanotechnologies (CINT)		■	
Center for Nanoscale Materials (CNM)		■	
Center for Nanophase Materials Sciences (CNMS)		■	
The Molecular Foundry (TMF)	■	■	
Autonomous Experiment Steering for BES Facilities	■	■	
BES Design and Development of Digital Twin Strategies		■	
Multifacility Experimentation and Analysis Workflows: X-ray Light Source Perspective	■		
Multifacility Experimentation and Analysis Workflows: Neutron Scattering Perspective	■		
Multifacility Experimentation and Analysis Workflows: Nanoscale Science Research Center (NSRC) Perspective	■		
Use of the ESnet for Quantum Simulations of Materials and Molecules		■	
The Materials Project (MP): Status and Future Directions		■	■

Table 4: Categorization of case studies from the Basic Energy Sciences (BES) program office.

Reviewing the BES case studies through the IRI criteria leads to several general observations:

- **Time-Sensitive Pattern:** Most BES facilities are, by design, operating high-throughput experimental resources. As the data volumes for these facilities has grown over time — and continues to grow with upgrades — in many cases it is infeasible to perform computation and storage close to the instrument. A number of beamlines have coupled their facilities with major HPC resources and will continue to do so as software and workflow management are updated to facilitate this shift. Additionally, autonomous experimental steering (e.g., real-time feedback on results and simulations through the use of AI/ML approaches) is emerging as a new time-dependent operational model.
- **Data Integration-Intensive Pattern:** BES data sources come in two main forms — observations from multiple sensors within experimental machinery and coupling observed results with those of simulation. These multi-modal data requirements place pressure on the availability of computation, storage, and networking resources during active experimental runs.
- **Long-Term Campaign Pattern:** Many BES facilities do not currently have long-term campaigns, as the data generated from experiments is returned to individual users. This has the potential to change in future years, as historical data is sought to train AI/ML models for analysis.

Appendix E: Biological and Environmental Research

BER Case Study	Time-Sensitive	Data Integration-Intensive	Long-Term Campaign
Joint Genome Institute (JGI)	■	■	■
Department of Energy Systems Biology Knowledgebase (KBase)		■	■
National Microbiome Data Collaborative (NMDC)		■	■
Environmental Molecular Sciences Laboratory (EMSL)	■	■	■
The Atmospheric Radiation Measurement Program (ARM)		■	■
The Earth System Grid Federation (ESGF)		■	■
The Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE)[14]	□	■	■
The Center for Advanced Bioenergy and Bioproducts Innovation (CABBI)		■	■
The Center for Bioenergy Innovation (CBI)		■	
The Great Lakes Bioenergy Research Center (GLBRC)		■	■
The Joint BioEnergy Institute (JBEI)	■	■	■
Energy Exascale Earth System Model (E3SM)		■	■
AmeriFlux Network		■	■
Next Generation Ecosystem Experiments (NGEE)		■	■
Coastal Efforts		■	■
Calibrated and Systematic Characterization, Attribution, and Detection of Extremes (CASCADE)		■	■
The Cooperative Agreement To Analyze variabiLity, change and predictabilitY in the earth SysTem (CATALYST)		■	■
Multi-Facility Workflows in BER	■	■	■

Table 5: Categorization of case studies from the Biological and Environmental Research (BER) program office.

14 Future plans indicate alignment with the Time-Sensitive Pattern, but today's workflows do not yet align.

Reviewing the BER case studies through the IRI criteria leads to several general observations:

- **Time-Sensitive Pattern:** There are limited time-sensitive use cases within BER, and usually these are limited to facilities that are trying to perform high-throughput experiments (e.g., numerous "runs" for samples) and requiring a way to share the raw data with processing and storage facilities. Most experiments and facilities are not capable of doing real-time analysis currently, due to sensor locations being remote or there being no pressing need to see instantaneous results in order to influence a subsequent run.
- **Data Integration-Intensive Pattern:** BER operates on multi-modal data constantly, and as a result most of the facilities and experiments are intrinsically data-intensive. This can be seen for some facilities that are designed to capture, store, categorize, and share results from a wide range of experiments, as well as long-running observational efforts that capture results from hundreds of instruments and sensors.
- **Long-Term Campaign Pattern:** BER data, whether from observations, sequences, or historic models, remains useful over time. A number of facilities are being developed to help better categorize and serve this information, and many long-running efforts report use of historic models and datasets over time.