

**ESNET REQUIREMENTS REVIEW**  
**DOE Office of Science - Nuclear Physics Program Office**  
**MAY 2019**

### **The ESnet User Facility**

The Energy Sciences Network (ESnet) is the Office of Science's high-performance network user facility, delivering highly-reliable data transport capabilities optimized for the requirements of large-scale science. In essence, ESnet is the circulatory system that enables the DOE science mission. ESnet is stewarded by the Advanced Scientific Computing Research (ASCR) Program, and managed and operated by the Scientific Networking Division at Lawrence Berkeley National Laboratory. ESnet is widely regarded as a global leader in the research and education networking community.

ESnet connects DOE national laboratories, user facilities, and major experiments so scientists can use remote instruments and computing resources as well as share data with collaborators, transfer large datasets and access distributed data repositories. ESnet is much more than an internet service provider; it provides a range of services that are tailored to meet the needs of data-intensive science.

### **Review Purpose and Process**

The primary purpose of an ESnet Requirements Review is to discuss and analyze current and planned science use cases and anticipated data output of a particular program, user facility, or project to inform ESnet's strategic planning, including network operations, capacity upgrades, and other service investments.

The Requirements Review is an in-person event. By design a highly conversational process through which all participants gain shared insight into the salient data management challenges of the subject program/facility/project. The ESnet Science Engagement Team organizes, convenes, and executes the Review in close coordination with the ESnet Program Manager and one or more leads from the subject program/project/facility.

### **Review Objectives, Outcomes, and Artifacts**

Through a case study methodology, the Review provides ESnet with information about:

- Existing and planned data intensive science experiments and/or user facilities, including the geographical locations of experimental site(s), computing resource(s), data storage, and research collaborator(s).
- For each experiment/facility project, a description of the "process of science," including the goals of the project, how experiments are performed and/or how the facility is used. This description includes information on the systems and tools used to analyze, transfer, and store the data that is produced.
- Current and anticipated data output on near and long term time scales.
- Timeline(s) for building, operating, and decommissioning of experiments, to the degree these are known.

- Existing and planned network resources, usage, “pain points” or bottlenecks in transferring or productively using the data produced by the science.

ESnet will focus the Review to deliver the following outcomes:

- Identification and analysis of any data management gaps and/or network bottlenecks that are barriers to achieving the scientific goals.
- A forecast of capacity/bandwidth needs by area of science, particularly in geographic regions where data production/consumption is anticipated to increase or decrease.
- Development of a shared understanding of the data management needs, challenges, and capability gaps that could inform strategic investments in solutions.

The Review will result in the following artifacts:

- A published report with an analysis of the findings from the review.
- A set of followup actions, if appropriate.

#### **Review Participants**

- The ESnet Program Manager and ASCR leadership
- Office of Nuclear Physics (NP) Program Managers
- NP Principal Investigators
- DOE Lab Networking and IT Staff, as appropriate
- ESnet staff, specifically: ESnet Leadership, Network Engineering Team members, and Science Engagement Team members

#### **Review Timeline**

1. Week of April 1st, 2019: Requirements review case study template shared with attendees.
  2. Week of April 10th and April 18th, 2019: ESnet Requirements Review Introductory Webinars: *A brief video-conference for participants to discuss the basics of the review, including expectations/best practices/Q&A for filling out the case study and preparing for the in-person review.*
  3. Friday April 26th, 2019: Draft written case studies due
  4. May 8th and 9th, 2019: In person review
  5. August 2019: Final report published
-

## CASE STUDY QUESTIONNAIRE

### BEFORE YOU BEGIN, PLEASE NOTE:

- ***\*\*ESnet recommends collaborating with the IT or networking staff that support your collaboration, facility or lab, particularly with respect to questions 6, 7, 8, 9 and 10. If you are unsure who may be able to assist, please email the ESnet Science Engagement Team ([engage@es.net](mailto:engage@es.net)).\*\****
- *Consider all contributions as a draft form that can be changed/discussed at any point before the case study is finalized in the report after the in-person review.*
- *Please direct any questions about this process directly to the ESnet Science Engagement Team ([engage@es.net](mailto:engage@es.net)).*

### **PROJECT/FACILITY NAME**

**PI name(s), institution(s), email address(es)**

## 1. Science Background (3-5 paragraphs total)

**Tell us about your science.**

*Please briefly describe your scientific research or the research you support; provide the high level context for the research: What are the goals of the science? Which department(s) and/or laboratory(s) are involved? Who are the stakeholders?*

*Provide an overview of the data life cycle; the narrative should summarize:*

- *The “story of the data” i.e. why and how the project/research creates, collects, analyzes, transfers, shares, and stores data.*
- *The transformation, longevity, and general usage patterns of the scientific data by those that use it.*
- *Make special note of the origin/chain of custody for data sets, if known.*

## 2. Collaborators

**Who are your collaborators? Where and how does the collaboration share data?**

*This section aims to capture the breadth of the science collaborations involved in your experiment or facility. In particular, the geographic location of your collaborators and how data sets are created, shared, computed and stored.*

*Using the table below:*

- *List facilities, significant users/collaborators, and/or virtual organizations (VOs) that are critical to the workflow in terms of transferring or sharing datasets.*

- *List geographical endpoints for collaborators, being as specific as possible, if known (estimates are also ok, e.g. city, the state, territory, or country).*
- *Please also capture a rough estimate of the breadth and depth of the collaboration space (e.g. number of users, number of participating facilities).*

*Please populate the table below adding additional rows as needed. If additional descriptions or explanation is needed, please provide in the space below the table.*

User/Collaborator and Location	Do they store a primary or secondary copy of the data?	Data access method, such as data portal, data transfer, portable hard drive, or other? (please describe "other")	Avg. size of dataset? (report in bytes, e.g. 125GB)	Frequency of data transfer or download? (e.g. ad-hoc, daily, weekly, monthly)	Is data sent back to the source? (y/n) If so, how?	Any known issues with data sharing (e.g. difficult tools, slow network)?

Please provide any supporting narrative text.

## **The Science Case Study – Instruments and Facilities, Process of Science, and Local vs. Remote components**

The below Sections 3, 4, and 5 are intended to collectively provide ESnet with a full view of the “who, what, how, and where” of each science collaboration or experiment represented in the requirements review. Section 3 focuses on the “what”, particularly the hard assets that generate or process the data. Section 4 focuses on the “how” these assets are used, i.e. the process of the science, and section 5 captures the geographic location of all the instruments and facilities that are part of the science effort.

This information allows ESnet to analyze the needs of the science collaboration in the context of the ESnet network footprint and its distributed capabilities.

### **3. Instruments and Facilities**

***What instruments and/or facilities are used to perform the science? how do these instruments and/or facilities work or operate?***

- *Please briefly describe the Instruments and Facilities used in your work, including any plans for major upgrades, new facilities, or similar changes.*
- *Please describe the Instrument or Facility’s compute, storage, and network capabilities.*
- *If you are a facility or manage an instrument, please describe the resources you make available to your users, or that users deploy at your facility.*
- *Please describe the composition of the data sets produced by the instrument or facility (e.g. file size, number of files, number of directories, total data set size)*

*Please provide these descriptions in the following time scales:*

- *Present-2 years (current budget horizon)*
- *Next 2-5 years (current technology horizon)*
- *Beyond 5 years (strategic planning)*

*Please note: If this section includes a shared resource (e.g. supercomputer center, or multi-user experimental facility), please describe the specific components that apply to your scientific use case.*

### **4. Process of Science**

***How do you use the instrument or facility and the data produced?***

*Please describe the way in which the instruments and facilities (as discussed above) are and will be used for knowledge discovery, emphasizing the role of networking in enabling the science - where applicable.*

*Please include (as appropriate): descriptions of the science workflows; methods for data analysis and data reduction; the integration of experimental data with simulation data, or other use cases.*

*This enables ESnet to forecast the impact of the science on network capabilities – including where and when to deploy additional bandwidth and services. Please provide these descriptions the following time frames:*

- *Present-2 years (current budget horizon)*
- *Next 2-5 years (current technology horizon)*
- *Beyond 5 years (strategic planning)*

*Please note: If this section includes a shared resource (e.g. supercomputer center, or multi-user experimental facility), please describe the specific components that apply to your scientific use case.*

## **5. Remote Science Activities**

### ***What remote resources are used in your science?***

*Please describe any remote instruments or resources used in the process of science, and how this work impacts or may impact the network. This could include any connections to or between instruments, facilities, people, or data at different sites (i.e., supercomputers, particle accelerators, genome sequencers, satellite data...)?*

*Please describe how these remote instruments are used and or how they are planned to be used in the following time frames:*

- *Present-2 years (current budget horizon)*
- *Next 2-5 years (current technology horizon)*
- *Beyond 5 years (strategic planning)*

## **6. Software Infrastructure**

### ***What data management software tools do you use to perform your research and analyze your data?***

*Please include descriptions of tools that perform the following tasks: (note: If commercial or open source software is being utilized, please indicate this)*

- *Locally or remotely manages data resources.*

- Facilitates the transfer of data sets from or to remote collaborators (e.g. Aspera, Globus, ftp, scp, etc.)
- Processes raw data into final and intermediate formats or data products.

Please describe the use and future use of software tools in the following timeframes:

- Present-2 years (current budget horizon)
- Next 2-5 years (current technology horizon)
- Beyond 5 years (strategic planning)

**\*\*NOTE\*\*** ESnet recommends collaborating with the IT or networking staff that support your collaboration, facility or lab to address this question. If you are unsure who may be able to assist, email ESnet ([engage@es.net](mailto:engage@es.net)).\*\*

## 7. Network and Data Architecture

### **What is the network architecture and bandwidth for your facility and/or laboratory and/or campus?**

It is critical for ESnet to understand the network resources used to move data from the data source location to the wider facility/campus network, and to external collaborators or other data resources.

Please describe in detail:

- The Local Area Network (LAN), Metro Area Network (MAN) and Wide Area Network (WAN) capabilities that connect your science experiment/facility/data source to external resources and collaborators.
- The local network configuration
- Network bandwidth connection speed(s)
- Any other relevant network capabilities.

**NOTE:** Please provide network diagrams where possible. These can be obtained from IT or networking staff that supports your facility or collaboration.

Please include information on whether or how the following tools are integrated into your network architecture, or if such tools are desired in the future:

- High-performance data transfer technology (e.g. Globus);
- New Network architectures (e.g., Science DMZ <http://fasterdata.es.net/science-dmz/>); or
- Network performance management tools (e.g. perfSONAR, <https://fasterdata.es.net/performance-testing/perfsonar/>).

Please describe the above existing and planned capabilities in the following timeframes:



- Present-2 years (current budget horizon)
- Next 2-5 years (current technology horizon)
- Beyond 5 years (strategic planning)

**\*\*NOTE\*\*** ESnet recommends collaborating with the IT or networking staff that support your collaboration, facility or lab to address this question. If you are unsure who may be able to assist, please email ESnet ([engage@es.net](mailto:engage@es.net)). **\*\***

## 8. Cloud Services

### ***If applicable, what cloud services do you use or plan to use?***

*Please describe current or planned use of cloud services for data analysis, storage, computing, or other purposes.*

*Please note: "Cloud" in this case could include research & education computing clouds such as Chameleon, commercial clouds such as Amazon, Google, IBM, or Microsoft, or private clouds hosted by some other organization not mentioned. Please specify what type of cloud service(s) you are using or intend to use.*

*Please share your cloud service plans in these time frames:*

- Present-2 years (current budget horizon)
- Next 2-5 years (current technology horizon)
- Beyond 5 years (strategic planning)

**\*\*NOTE\*\*** ESnet recommends collaborating with the IT or networking staff that support your collaboration, facility or lab to address this question. If you are unsure who may be able to assist, please email ESnet ([engage@es.net](mailto:engage@es.net)). **\*\***

## 9. Data-Related Resource Constraints

### ***Please name any current or future network or data-related constraints to your work?***

*Please describe any current or anticipated future constraints that affect your productivity. Examples might include insufficient data transfer performance, insufficient storage system space or performance, difficulty finding or accessing data in community data repositories, or unmet computing needs.*

*Please share constraints as they are known or expected in the following time frames:*

- Present-2 years (current budget horizon)

- Next 2-5 years (current technology horizon)
- Beyond 5 years (strategic planning)

**\*\*NOTE\*\*** ESnet recommends collaborating with the IT or networking staff that support your collaboration, facility or lab to address this question. If you are unsure who may be able to assist, please email ESnet ([engage@es.net](mailto:engage@es.net)). \*\*

## 10. Outstanding Issues

### *What have we missed?*

*Please use this space to address or discuss any challenges, barriers, or concerns that aren't discussed elsewhere in the case study. In particular, if there are current network or data transfer performance problems that impact scientific productivity, please describe them.*

**11. Please provide a list of names/affiliations/emails of any contributors to this report.**

#END#