**CASE STUDY QUESTIONNAIRE**

**BEFORE YOU BEGIN, PLEASE NOTE:**

* ***\*\*ESnet recommends coordination with the IT or networking staff that support your collaboration, facilities, or lab, particularly with respect to questions 6 through 11. If you are unsure who may be able to assist, please email the ESnet Science Engagement Team (******engage@es.net******) and we can help you locate the appropriate resources.\*\****
* *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**).*

***PLEASE ENTER YOUR FACILITY’S NAME /TITLE***

***PLEASE DENOTE ALL: Author name(s), institution(s), email address(es) that contributed***

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

***Tell us about your science.***

*Please briefly describe the scientific research of the facility, or the research you directly 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 a brief 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 major collaborators? Where and how do collaborations with the facility share data? What future collaborations are possible?***

*This section aims to capture the breadth of the science collaborations involved in your facility or experiment. In particular, the geographic location of your collaborators and how data sets are created, shared, computed and stored. To the extent that it is known, what future possibilities exist in this space?*

*If you are representing a complete facility, this information may be available from your annual reports, and doesn’t need to reflect any specific experimental team you support.*

*Using the table below:*

* *List other facilities, significant users/collaborators, and/or virtual organizations (VOs) that are critical to the general workflow in terms of transferring or sharing datasets.*
* *List known geographical endpoints for collaborators, being specific if possible (estimates are also ok, e.g., city, the state, territory, or country).*
* *Please capture a rough estimate of the breadth and depth of the collaboration space (e.g., number of users, number of participating external facilities).*
* *Emerging or future collaborations can be denoted along with a possible timeline.*

*Please populate the table below adding additional rows as needed. If additional descriptions or explanations are 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 for the entries in the table.

## 3. Integrated Research Infrastructure (IRI) Readiness

***Does your workflow rely on, or could it benefit from, the use of multiple DOE SC resources?***

*Research communities that utilize DOE SC User Facilities are experimenting with and demanding solutions integrated with high performance computing and data infrastructure. In response to these drivers, DOE has embraced the IRI[[1]](#footnote-0) vision 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.*

*The IRI Architecture Blueprint Activity (IRI-ABA[[2]](#footnote-1)) brought together domain experts from all DOE SC Programs to look for common patterns within diverse workflows across a range of scientific disciplines. Participants discovered three common patterns:*

* *The* ***Time-Sensitive*** *pattern is characteristic of low-latency workflows requiring real-time, or near-real-time, response across more than one Facility or resource.*
* *The* ***Data Integration-Intensive*** *pattern is characteristic of workflows combining diverse datasets to deepen and expand context.*
* ***The Long-Term Campaign*** *pattern is characteristic of workflows requiring sustained access to more than one Facility or resource, at scale, for years or decades.*

*Discuss if your facility workflow does, or will, require the use of multiple DOE SC facilities, and if it exhibits any of the IRI patterns. A discussion of some of these points will be featured during the in-person review.*

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

Sections 4 through 6 are intended to provide ESnet with a full view of the “who, what, how, and where” of a facility or experiment being represented in the requirements review.

* ***Section 4*** focuses on the “what”, particularly the hard assets that generate or process the data.
* ***Section 5*** focuses on the “how” these assets are used, i.e., the process of the science
* ***Section 6*** focuses on the “where” these assets are located, i.e., the geographic location of all the instruments and facilities that are part of the science effort. Please indicate shared resources where applicable.

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

## 4. 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 facility or experiment, including any major Instruments operated or used; include any plans for major upgrades, new instruments, or similar changes that may be disruptive.*
* *Please describe the facility or experiments computing, storage, and network capabilities.*
* *If you are representing a facility, please describe the resources you make available to your users, or that users may choose to deploy.*
* *Please describe the composition of the data sets produced by the major instruments of the facility or experiment (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)*

## 5. Generalized Process of Science

***How do you use the instrument and facilities above, via the Collaboration space, to produce scientific data?***

*Please describe the way in which the instruments, or the facility as a whole, are and will be used for knowledge discovery, emphasizing the role of networking in enabling the science - where applicable.*

*The aim for this section is not to capture all specific workflows, but we are looking for a generalized workflow of the specific experiment, or of a common user of the facility: this would include a description of the science; methods for data analysis and data reduction; the integration of experimental data with simulation data, or other use cases that may be relevant to help us understand how the science maps to the available technology.*

*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)*

## 6. Remote Science Activities

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

*While we are profiling the operations of an experiment or of a facility, we acknowledge that not all common activities of a general workflow may occur in the same location. If applicable, describe any* ***remote*** *instruments or resources commonly used related to the common process of science. Remote can be defined as not local to the immediate facility operational footprint, and would imply the use of networking technologies to link these different resources. This could include any connections to or between instruments, facilities, people, or data that must be done at a different site.*

*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)*

## 7. 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. If software does not exist yet to accomplish a specific goal or task, please indicate it as an area of need.* ***\*\*Note: If commercial or open source software is being utilized, please indicate this.\*\****

* *Locally or remotely manages data resources.*
* *Facilitates the mobility of data sets from or to remote collaborators (e.g., Globus, ftp, scp, Aspera, etc.)*
* *Processes raw data into final and intermediate formats or data products.*

*Other software use cases that are more functional or operational (e.g., business processes like task trackers, productivity like spreadsheets or presentation tools, communication such as zoom) do not need to be listed, unless they are critical to the process of science. An example would be a tool that facilitates collaboration or workflow, that has now become critical to the overall success of the facility.*

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

## 8. Network and Data Architecture

***What is the network architecture and bandwidth for your experiment or facility?***

*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. We will be separately asking for an overview of the complete laboratory network infrastructure from your site IT professionals, but are specifically seeking information for the smaller experimental or facility footprint.*

*Please describe in detail:*

* *The Local Area Network (LAN) capabilities that connect your 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.*

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

## 9. Emerging Technology Use

***The following subsections will discuss the use of new technologies, and what roles they may play in your scientific process.***

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

### 9.1. 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. This can be experimental to understand if the cloud provides a valuable path forward as a tool in your research, or it could be integrated and operational (e.g., the use of commercially provided GPUs for AI/ML training, or bursting to cloud resources when local resources are not available).*

*Please note: “Cloud” in this case could include research & education computing clouds such as GENI, Cloud Lab, Chameleon, or others along with “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. This question is meant to help understand the external use of resources to a facility over time.*

### 9.2. Artificial Intelligence and Machine Learning

***How has the laboratory adapted to the growing need to supply resources for scientific use of AI and ML in the process of science?***

*Please discuss how the laboratory or facilities have invested in AI/ML technologies. This could be the purchase and operation of on-site capabilities that are available for use, or it can be partnerships with other entities (R&E, commercial, etc) to utilize remote resources.*

### 9.3. Wireless or Distributed LocationScience Activities

***Do you have any wireless capability studies or capability upgrades planned, or needs to extend wireless data coverage to cover new locations? Does any of your work take place outside of a traditional laboratory or facility environment?***

*Please discuss if any aspects of your scientific workflow depends on resources that may be located in an isolated area, e.g., a sensor array or instrument deployed far from terrestrial networking that may need access to cellular or satellite connectivity. Examples include:*

* *Improved building cellular coverage (Distributed Antenna Systems, DAS, or Neutral Host, etc)*
* *Wifi6 or other Wifi upgrades*
* *Use of Starlink or other Low Earth Orbit constellations?*
* *Use of LoRA or other mesh network sensor wireless systems*
* *Other? mmWave, free space optical, etc*

## 10. 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 impact scientific productivity. Examples might include insufficient data transfer performance, insufficient storage system space or performance characteristics, difficulty finding or accessing data in community data repositories, unmet computing needs, or policies that do not facilitate easy access to technology to assist with data mobility.*

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

##

## 11. Data Mobility Endpoints

***How are you sharing research data?***

*If your facility has a Globus (or other form of data transfer software) endpoint for data sharing, please list the hostname(s)/endpoint title(s).* [*ESnet maintains a set of well-tuned test endpoints*](https://fasterdata.es.net/science-dmz/learn-more/2019-2020-data-mobility-workshop-and-exhibition/data-mobility-exhibition-archive/)*, and recommends that facilities evaluate their performance against these resources. ESnet will follow-up during the requirements process to test and evaluate the capabilities of the facility, and suggest tuning steps to improve performance.*

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

## 12. 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, issues with access to or adoption of technology, or policies that may impact scientific productivity, please describe them.*

## 13. Please provide a list of all names/affiliations/emails of any contributors to this report.

#END#

1. March 2021 IRI white paper: <https://www.osti.gov/biblio/1863562> [↑](#footnote-ref-0)
2. IRI Architecture Blueprint Activity (IRI-ABA) reports: <https://www.osti.gov/biblio/1984466> and <https://www.osti.gov/biblio/2008205> [↑](#footnote-ref-1)