

BEFORE YOU BEGIN, PLEASE NOTE:

- ***EPOC recommends collaborating with the IT or networking staff that support your collaboration, facility or lab - particularly questions 6, 7, 8, 9 and 10. If you are unsure who may be able to assist, please email EPOC (epoc@iu.edu).***
- 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.
- Ask any questions about this process directly to the EPOC Team (epoc@iu.edu)

Title of Facility or Research Project, Names/Contact Info of Operator/PI

1. Science Background

Tell us about your science.

Please briefly describe your scientific research or research you support including high level details of the goals of the science, and the department(s) and/or laboratory(s) involved in the research. The narrative should focus on:

- 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:

- 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).

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, and any connections to any other major scientific instruments (e.g.: supercomputers, particle accelerators, tokamaks, genome sequencers, satellite data, computational clusters, storage systems, etc.)
- 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 EPOC 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 major scientific instruments outside of local instruments and facilities (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)

EPOC 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 EPOC (epoc@iu.edu).

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 EPOC to understand the network resources used to move data from the data source location, to the wider facility/campus network, to 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://www.perfsonar.net/>).

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)

EPOC 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 EPOC (epoc@iu.edu).

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)

EPOC 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 EPOC (epoc@iu.edu).

9. Resource Constraints

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

Please describe current and future constraints of the science facility/experiment from the perspective of the data that is produced and shared – particularly as it relates to computation, storage, and networking.

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)

EPOC 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 EPOC (epoc@iu.edu).

10. Parent Organization(s)

What level of support do you receive from your local campus or regional network? Are you partnered on any grant activities for facility upgrades?

Please discuss the relationship that your department, facility, or collaboration has with the overarching organizational structure if your campus and/or regional network. E.g. have you received, or will you pursue, funding via the NSF's Campus Cyberinfrastructure (CC*) program (https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504748&org=NSF&sel_org=NSF)? If you have, cite any relevant partnerships, grant numbers, and length of time expected to be working towards campus upgrades.

EPOC 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 EPOC (epoc@iu.edu).

11. Outstanding Issues

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.

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