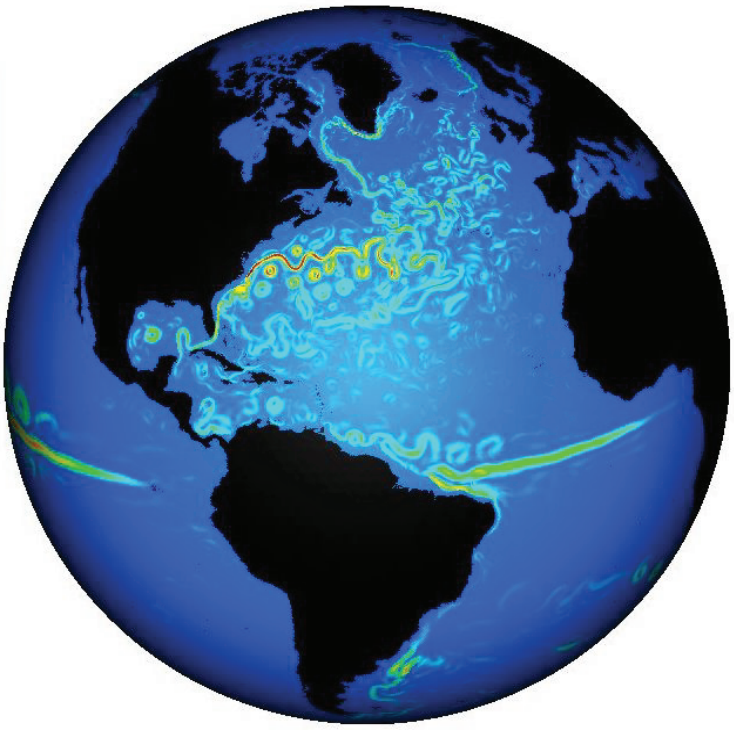
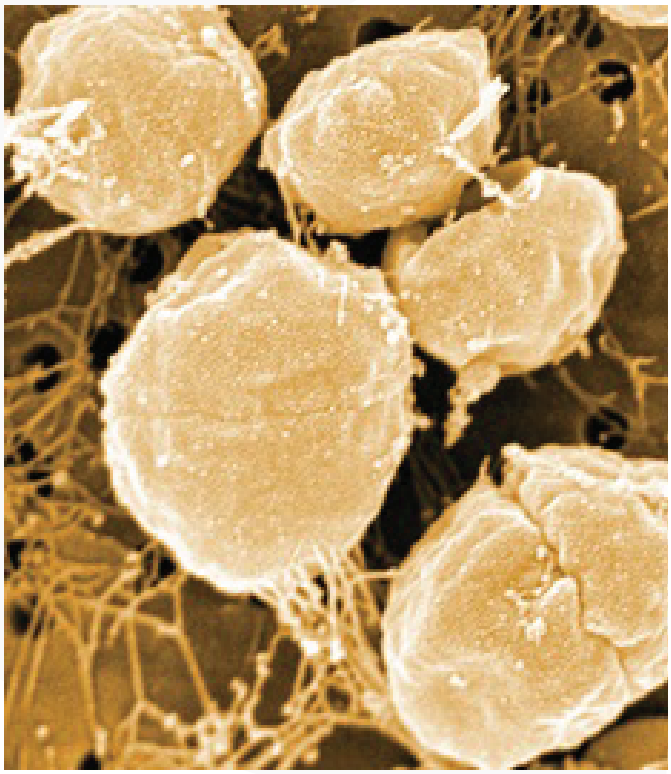
ESnet
ENERGY SCIENCES NETWORK

Biological and Environmenal Research Network Requirements Review

Final Report

September 18-19, 2015





BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science

Disclaimer

Biological and Environmental Research Network Requirements Review Final Report

Office of Biological and Environmental Research, DOE Office of Science
Energy Sciences Network (ESnet)
Germantown, Maryland
September 18–19, 2015

ESnet is funded by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. Vince Dattoria is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

This work was supported by the Directors of the Office of Science, Office of Advanced Scientific Computing Research, Facilities Division, and the Office of Biological and Environmental Research.

This is LBNL report LBNL-1004370.

Cover images: (Left) Model Prediction Across Scales (MPAS) simulation at 15 kilometer resolution. [Source image: http://ascr-discovery.science.doe.gov/2012/09/sphere-of-influence/.] (Right) Scanning electron micrograph of the first sequenced *archea Methanococcus jannaschii*. [Image courtesy of B. Boonyaratanakornkit, D.S. Clark, G. Vrdoljak, University of California, Berkeley.]

Contents

Participants and Contributors

Greg Bell, ESnet, Networking, ESnet Director
David Benton, University of Wisconsin, GLBRC
Shane Canon, NERSC, KBase, NERSC/JGI
David Cowley, PNNL, EMSL
Jody Crisp, ORISE, Logistics
Eli Dart, ESnet, Networking, Review Chair
Vince Dattoria, DOE/SC/ASCR, ESnet Program Manager
Kjiersten Fagnan, LBNL, NERSC/JGI
Ian Foster, ANL, Globus, Data Management, ESGF, ACME, CMIP
Patty Giuntoli, ESnet, Networking
Mary Hester, ESnet, Networking
Bob Hettich, ORNL, Structural Biology
Jay Hnilo, DOE/SC/BER, BER CESD Programs
Adina Howe, Iowa State University, Microbial Communities, Bioinformatics
Robert Jacob, ANL, ACME
Daniel Jacobsen, ORNL, Computational Biology
Lisa Love, NOAA, Network Security
Ramana Madupu, DOE/SC/BER, BER BSSD Programs
Joe Metzger, ESnet, Networking
Giri Palanisamy, ORNL, ARM Facility
Pablo Rabinowicz, DOE/SC/BER, BER BSSD Programs
Lauren Rotman, ESnet, Networking
Robert Sears, NOAA, Networking
Gary Strand, NCAR, UCAR-CA, Data Management
Michael Wehner, LBNL, Large Scale Climate Data Analysis, CASCADE
Dean Williams, LLNL, ESGF, MIPs, ACME, PCMDI, UV-CDAT, Workflow, ICNWG
Jason Zurawski, ESnet, Networking

Report Editors

Eli Dart, ESnet: dart@es.net
Mary Hester, ESnet: mchester@es.net
Jason Zurawski, ESnet: zurawski@es.net

Executive Summary

The Energy Sciences Network (ESnet) is the primary provider of network connectivity for the U.S. Department of Energy (DOE) Office of Science (SC), the single largest supporter of basic research in the physical sciences in the United States. In support of the Office of Science programs, ESnet regularly updates and refreshes its understanding of the networking requirements of the instruments, facilities, scientists, and science programs that it serves. This focus has helped ESnet to be a highly successful enabler of scientific discovery for over 25 years.

In September 2015, ESnet and the Office of Biological and Environmental Research (BER), of the DOE Office of Science, organized a review to characterize the networking requirements of the programs funded by the BER program office. BER advances world-class biological and environmental research programs and scientific user facilities to support DOE's energy, environment, and basic research missions.

Several key findings highlighting the results from the review are noted below.

1. While some fields such as high energy physics and astrophysics have data sets that allow for significant data reduction (e.g., filtering events in high energy physics, or eliminating the dark portions of a telescope image), climate data sets are not easy to reduce in this way. Because of this, climate data analysts need significant volumes of data. In general, data reduction in climate science means eliminating variables from a data set which are not needed for a particular analysis; however, that does not mean those variables are not important or valuable to other climate research and scientific discovery.
2. Some workflows involve the creation of high-value derived data products as a result of the analysis of genomics data from public data repositories. Currently, these workflows require the download of the input data as files from the public repository to a local filesystem, followed by the execution of the analysis code on a computational resource connected to the local filesystem. It would be more efficient if the analysis code could read the input data directly from the network into memory, without relying on copying the raw data to local storage first. The Large Hadron Collider (LHC) experiments are in the process of transitioning to this remote I/O model—genomics could take advantage of this model as well.
3. Computing workflows for metagenomics often require the generation and subsequent use of very large intermediate data sets. The generation of these data sets is computationally costly, and the data sets are quite large–on the scale of tens to hundreds of terabytes. In the ideal case, the intermediate data set would be preserved for a period of months, and would be used in multiple analysis campaigns.
4. Data portal systems are now a significant bottleneck in many data staging workflows. Point-and-click web-based interfaces do not scale to the size and complexity of the data staging tasks required. Some portal systems have machine-consumable application program interfaces (APIs) which allow for automation of data staging, but many do not. Updates to legacy web-based portal systems which allowed for the automation of data staging workflows would be of significant benefit to multiple projects.
5. High-throughput sequencing, microbiology, and computational biology are powerful technologies that can be combined in innovative ways to achieve new understandings of microbial populations. However, it is very difficult to train new researchers and keep skill sets current because the state of the art in these fields is changing so quickly. Documentation of best practices would be a great help for researchers in this space.
6. Several projects and facilities identified the lack of unification of user identities across the DOE facilities complex as an impediment to user productivity. The topic was raised in discussion at the review multiple times in multiple contexts (e.g., Globus data transfer workflows, developers on the Accelerated Climate

Modeling for Energy, ACME, project running code at multiple computational facilities, Atmospheric Radiation Measurement (ARM) facility staff running code at multiple locations).

7. Several facilities are forced to ship physical media (typically USB hard drives) for data ingest and export involving poorly-configured end user systems. Data transfer via portable media is wasteful of valuable human resources, low-performance, and error-prone.
8. According to the BER program managers, one of the significant challenges for the BER Biological Systems Science Division (BSSD) in the coming years will be the integration of multiple heterogeneous data sources and data types. There is significant scientific opportunity if this is done well, but it will be a challenge.

This report expands on these points, and covers other collaborative projects within BER. The report contains a Findings section and documents Case Studies discussed during the Requirements Review.

Findings

Below are the findings for the BER and ESnet Requirements Review held September 17–18, 2015. These points summarize important information gathered during the review.

- Several facilities are forced to ship physical media (typically USB hard drives) for data ingest and export involving poorly-configured end user systems. Data transfer via portable media is wasteful of valuable human resources, low-performance, and error-prone.
- Several Globus users expressed a desire to use the Globus application program interface (API) to automate complex workflows. A workshop or training seminar for the use of the Globus API at DOE facilities would be of use to multiple science communities.
- The Environmental Molecular Sciences Laboratory (EMSL) transfers climate model data to and from other computing centers such as National Center for Atmospheric Research (NCAR) and the DOE ASCR computing facilities. Currently, the data sets are 6–10TB in size, and take multiple days to transfer. It is likely that the performance of this workflow could be significantly improved. Note that the data set size is expected to double within 2–5 years.
- The DOE Systems Biology Knowledgebase (KBase) has a future requirement for guaranteed bandwidth services to support data replication between the Argonne National Laboratory (ANL) and National Energy Research Scientific Computing Center (NERSC) KBase instances.
- In addition to submitting samples to Joint Genome Institute (JGI) for sequencing and subsequent analysis, JGI users submit raw data sets to JGI for analysis. This is currently a small portion of the JGI data workload, but it is expected to grow over time.
- The growth curve of sequencing data at JGI has flattened. New technologies might change this in the coming years.
- Some workflows involve the creation of high-value derived data products as a result of the analysis of genomics data from public data repositories. Currently, these workflows require the download of the input data as files from the public repository to a local filesystem, followed by the execution of the analysis code on a computational resource connected to the local filesystem. It would be more efficient if the analysis code could read the input data directly from the network into memory, without relying on copying the raw data to local storage first. The LHC experiments are in the process of transitioning to this remote I/O model—genomics could take advantage of this model as well.
- Scientists at the BioEnergy Sciences Center (BESC) have found that computing facilities and other resources that are configured to serve traditional simulation workloads such as those found in physics, cosmology, and climate science are not necessarily well-suited to computational biology. BESC scientists are working with Compute and Data Environment for Science (CADES) and Oak Ridge Leadership Computing Facility (OLCF) to develop an environment tailored for biology.
- Computing workflows for metagenomics often require the generation and subsequent use of very large intermediate data sets. The generation of these data sets is computationally costly, and the data sets are quite large–on the scale of tens to hundreds of terabytes. In the ideal case, the intermediate data set would be preserved for a period of months, and would be used in multiple analysis campaigns.
- Data portal systems are now a significant bottleneck in many data staging workflows. Point-and-click web-based interfaces do not scale to the size and complexity of the data staging tasks required. Some portal

systems have machine-consumable APIs which allow for automation of data staging, but many do not. Updates to legacy web-based portal systems which allowed for the automation of data staging workflows would be of significant benefit to multiple projects.

- A unified metadata service which allowed for easily searching across the major metagenomics data archives would be of significant benefit to the metagenomics community. At the moment, users must search across multiple archives and portals which are not well integrated.
- Several projects and facilities identified the lack of unification of user identities across the DOE facilities complex as an impediment to user productivity. The topic was raised in discussion at the review multiple times in multiple contexts (e.g., Globus data transfer workflows, developers on the ACME project running code at multiple computational facilities, ARM facility staff running code at multiple locations).
- Many BER facilities and projects use Globus for data transfer, and in most cases Globus works well (and is a significant improvement over what was in place beforehand).
- The Accelerated Climate Modeling for Energy (ACME) project uses computing resources at multiple sites and facilities. Because of the distributed nature of the computing resources used for ACME, the project needs to transfer data between the different sites as well. There was a discussion of conducting data transfer performance tests between ACME computing and analysis sites before the start of large ACME model runs, which are expected in the summer of 2016.
- While some fields such as high energy physics and astrophysics have data sets that allow for significant data reduction (e.g., filtering events in high energy physics, or eliminating the dark portions of a telescope image), climate data sets are not easy to reduce in this way. Because of this, climate data analysts need significant volumes of data. In general, data reduction in climate science means eliminating variables from a data set which are not needed for a particular analysis; however, that does not mean those variables are not important or valuable to another climate scientist's research.
- There was a discussion of the use of container computing technologies such as NERSC's Shifter for packaging climate data analysis tools to achieve consistent deployment at multiple computing facilities.
- The CASCADE project makes data products available to collaborators (currently 50TB, which is projected to grow as the project purchases more disk space at NERSC). A scalable data service method is needed, because a web portal does not scale. Together ESGF and Globus are likely to be a good fit.
- Large climate model data sets offer significant value for multiple research areas. The Coupled Model Intercomparison Project Phase 5 (CMIP5) data archive is one example, but it is not the only one. There was some discussion on whether and how to save high-frequency (3-hour or 6-hour) data from high-resolution model runs—the data sets are large, but they have significant value for certain types of analysis (e.g., extreme weather analysis under different climate change scenarios).
- The CMIP6 High-Res Model Intercomparison Project (MIP) is expected to contain data from approximately 10 models at 25km resolution. This data set will allow tracking hurricanes in the same way that Prabhat and Wehner et al. tracked extra-tropical cyclones for a recent paper. However, this will require a significant data staging effort from multiple ESGF data centers to a single computing center with sufficient scale to run the analysis.
- Machine learning techniques appear promising for finding and tracking features in climate model data. However, machine learning techniques require large input data sets. There is significant scientific promise if the staging of large-scale data sets can be made routine.
- The Earth System Grid Federation (ESGF) is incorporating Globus in a deeper way in version 2.0 of the ESGF software stack.
- ESGF is moving to a two-tiered system for sites. Modeling centers run models, and may publish some of their own data. However, data centers or "supernodes," will host large replica archives of data from multiple modeling centers on rotating disk. The supernodes will replicate data to/from other supernodes for load sharing and disaster recovery purposes. There are currently five data centers which will be supernodes: the National Computational Infrastructure (Australia), British Atmospheric Data Centre (Great Britain), Deutsches Klimarechenzentrum (Germany), Institut Pierre-Simon Laplace (France), and Program

For Climate Model Diagnosis and Intercomparison (United States). It is expected that in the coming years there will be a supernode in Japan and China.

- Some sites have integrated Software Defined Networking (SDN) with Globus in order to automate some aspects of security policy enforcement for Globus endpoints. It is likely that this technology would be of significant benefit to multiple DOE facilities.
- Globus integration with Google Drive would be of interest to multiple sites. The discussion topic revolved around the ability to create a Globus endpoint associated with a Google Drive folder which might behave in a similar way to a Globus endpoint associated with an Amazon S3 data store.
- According to the BER program managers, one of the significant challenges for BER BSSD in the coming years will be the integration of multiple heterogeneous data sources and data types. There is significant scientific opportunity if this is done well, but it will be a challenge.
- There is significant opportunity for improved data transfer performance between the EMSL and the home institutions of some EMSL users.
- The ESnet knowledge base at `http://fasterdata.es.net/` has been helpful to DOE facility system administrators.
- Institutional firewalls are still a problem at some sites. The Science DMZ model might be helpful in these cases.
- According to Globus data, the data transfer nodes (DTNs) which yield the best data transfer performance are typically the DTNs which have been deployed in alignment with the Science DMZ model.

Action Items

ESnet recorded a set of action items from the BER-ESnet Requirements Review, continuing the ongoing support of collaborations funded by the BER program. ESnet will

- Follow up with the ARM facility regarding the performance problems described in the ARM case study;
- Continue to work with the ACME project on the topic of data transfer performance between computational facilities;
- Share information with the review attendees related to the ESnet test DTNs and their use for testing local Globus endpoints;
- Continue conversations with ESGF and the ACME project to determine the best way to integrate ACME sites into the International Climate Network Working Group (ICNWG); and
- Follow up with EMSL on data transfer performance.

ESnet SC Requirements Review Background and Structure

Funded by the Office of Advanced Scientific Computing Research (ASCR) Facilities Division, ESnet's mission is to operate and maintain a network dedicated to accelerating science discovery. ESnet's mission covers three areas:

1. Working with the DOE SC-funded science community to identify the networking implications of instruments and supercomputers and the evolving process of how science is done.
2. Developing an approach to building a network environment to enable the distributed aspects of SC science and to continuously reassess and update the approach as new requirements become clear.
3. Continuing to anticipate future network capabilities to meet new science requirements with an active program of R&D and advanced development.

Addressing point (1), the requirements of the SC science programs are determined by:

(a) A review of major stakeholders' plans and processes, including the data characteristics of scientific instruments and facilities, in order to investigate what data will be generated by instruments and supercomputers coming online over the next 5–10 years. In addition, the future process of science must be examined: How and where will the new data be analyzed and used? How will the process of doing science change over the next 5–10 years?

(b) Observing current and historical network traffic patterns to determine how trends in network patterns predict future network needs.

The primary mechanism to accomplish (a) is through the SC Network Requirements Reviews, which are organized by ASCR in collaboration with the SC Program Offices. SC conducts two requirements reviews per year, in a cycle that assesses requirements for each of the six program offices every three years. The review reports are published at `http://www.es.net/requirements/`.

The other role of requirements reviews is to help ensure that ESnet and ASCR have a common understanding of the issues that face ESnet and the solutions that it undertakes.

In April 2015, ESnet organized a review in collaboration with the ASCR Program Office to characterize the networking requirements for the facilities and science programs funded by ASCR.

Participants were asked to communicate and document their requirements in a case-study format that included a network-centric narrative describing the science, instruments, and facilities currently used or anticipated for future programs; the network services needed; and how the network is used. Participants considered three timescales on the topics enumerated below: the near-term (immediately and up to two years in the future); the medium-term (two to five years in the future); and the long-term (greater than five years in the future).

More specifically, the structure of a case study was as follows:

- Background—an overview description of the site, facility, or collaboration described in the case study.
- Collaborators—a list or description of key collaborators for the science or facility described in the case study (the list need not be exhaustive).

- Network and Data Architecture—description of the network and/or data architecture for the science or facility. This is meant to understand how data moves in and out of the facility or laboratory focusing on local infrastructure configuration, bandwidth speed(s), hardware, etc.
- Instruments and Facilities—a description of the network, compute, instruments, and storage resources used for the science collaboration/program/project, or a description of the resources made available to the facility users users, or resources that users deploy at the facility.
- Process of Science—a description of the way the instruments and facilities are used for knowledge discovery. Examples might include workflows, data analysis, data reduction, integration of experimental data with simulation data, etc.
- Remote Science Activities—a description of any remote instruments or collaborations, and how this work does or may have an impact on your network traffic.
- Software Infrastructure—a discussion focused on the software used in daily activities of the scientific process including tools that are used to locally or remotely to manage data resources, facilitate the transfer of data sets from or to remote collaborators, or process the raw results into final and intermediate formats.
- Cloud Services—discussion around how cloud services may be used for data analysis, data storage, computing, or other purposes.

The case studies included an open-ended section asking for any unresolved issues, comments or concerns to catch all remaining requirements that may be addressed by ESnet.

Office of Biological and Environmental Research Overview

The Biological and Environmental Research (BER) program supports fundamental research and scientific user facilities to address diverse and critical global challenges. The program seeks to understand how genomic information is translated to functional capabilities, enabling more confident redesign of microbes and plants for sustainable biofuel production, improved carbon storage, or contaminant bioremediation. BER research advances understanding of the roles of Earth's physical and biogeochemical systems (the atmosphere, land, oceans, sea ice, and subsurface) in determining climate so we can predict climate decades or centuries into the future—information needed to plan for future energy and resource needs. Solutions to these challenges are driven by a foundation of scientific knowledge and inquiry in atmospheric chemistry and physics, ecology, biology, and biogeochemistry.

BER research uncovers nature's secrets from the diversity of microbes and plants to understand how biological systems work, how they interact with one another, and how they can be manipulated to harness their processes and products. By starting with the potential encoded by organisms' genomes, BER-funded scientists seek to define the principles that guide the translation of the genetic code into functional proteins and the metabolic/regulatory networks underlying the systems biology of plants and microbes as they respond to and modify their environments. BER integrates discovery- and hypothesis-driven science, technology development, and foundational genomics research into predictive models of biological function for DOE mission solutions. BER plays a unique and vital role in supporting research on atmospheric processes; terrestrial ecosystem processes; subsurface biogeochemical processes involved in nutrient cycling, radionuclide fate and transport, and water cycling; climate change and environmental modeling; and analysis of impacts and interdependencies of climatic change with energy production and use. These investments are coordinated to advance an earth system predictive capability, involving community models open to active participation of the research community. For more than two decades, BER has taken a leadership role to advance an understanding of the physics and dynamics governing clouds, aerosols, and atmospheric greenhouse gases, as these represent the more significant weaknesses of climate prediction systems. BER also supports multidisciplinary climate and environmental research to advance experimental and modeling capabilities necessary to describe the role of the individual (terrestrial, cryospheric, oceanic, and atmospheric) component and system tipping points that may drive sudden change. In tight coordination with its research agenda, BER supports three major national user facilities: the Atmospheric Radiation Measurement (ARM) Climate Research Facility, the Joint Genome Institute (JGI), and the Environmental Molecular Sciences Laboratory (EMSL). Significant investments are provided to community database and model diagnostic systems to support research efforts.

Climate and Environmental Sciences Division

The Climate and Environmental Sciences Division (CESD) focuses on fundamental research that advances a robust, predictive understanding of Earth's climate and environmental systems and informs the development of sustainable solutions to the nation's energy and environmental challenges. As provided by the 2012 CESD Strategic Plan (http://science.energy.gov/ /media/ber/pdf/CESD-StratPlan-2012.pdf), five goals frame the division's programs and investments: (1) Synthesize new process knowledge and innovative computational methods that advance

next-generation, integrated models of the human-Earth system; (2) develop, test, and simulate process-level understanding of atmospheric systems and terrestrial ecosystems, extending from bedrock to the top of the vegetative canopy; (3) advance fundamental understanding of coupled biogeochemical processes in complex subsurface environments to enable systems-level prediction and control; (4) enhance the unique capabilities and impacts of the ARM and EMSL scientific user facilities and other BER community resources to advance the frontiers of climate and environmental science; and (5) identify and address science gaps that limit translation of CESD fundamental science into solutions for DOE's most pressing energy and environmental challenges.

CESD focuses on three research activities, each containing one or more programs and/or linkages to national user facilities. These activities are: (a) the Atmospheric System Research activity, which seeks to understand the physics, chemistry, and dynamics governing clouds, aerosols, and precipitation interactions, with a goal to advance the predictive understanding of the climate system; (2) the Environmental System Science activity, which seeks to advance a robust, predictive understanding of terrestrial surface and subsurface ecosystems, within a domain that extends from the bedrock to the top of the vegetated canopy and from molecular to global scales; and 3) the Climate and Earth System Modeling activity, which seeks to develop high-fidelity community models representing earth system and climate system variabilities and change, with a significant focus on the response of systems to natural and anthropogenic forcing.

The primary programs that actively use ESnet are: (1) the Earth System Modeling (ESM) Program, which develops advanced numerical algorithms to represent the dynamical and biogeophysical elements of the earth system and its components; (2) the Regional and Global Climate Modeling Program, which focuses on understanding the natural and anthropogenic components of regional variability and change, using simulations and diagnostic measures; (3) the EMSL facility, which provides integrated experimental and computational resources for discovery and technological innovation in the environmental molecular sciences to support the needs of DOE and the nation; and (4) the ARM facility, which provides the national and international research community unparalleled infrastructure for obtaining precise observations of key atmospheric phenomena needed for the advancement of atmospheric process understanding and climate models.

ESnet continues to be the primary network provider for data transfer for Coupled Model Intercomparison Projects (CMIPs), which in turn facilitate the analysis and synthesis for the Intergovernmental Panel on Climate Change (IPCC). CMIPs are carried out by utilizing the multiple nodes of the Earth System Grid Federation (ESGF). In addition, numerous multi-lab projects, such as the Climate Science for a Sustainable Energy Future (CSSEF), use ESnet to support data transfer requirements involving the ESGF. As the emphasis on finer spatial resolution for climate and environmental models is combined with more detail on uncertainty quantification associated with model outputs, data transfer requirements become increasingly more important. ESnet is also the primary network provider that enables remote access to EMSL's high-performance computing (HPC) system, numerous mass spectrometry systems, and EMSL's Aurora data-storage archive. EMSL has also established interfaces with the JGI for automated downloading of data. All these developments are significantly increasing EMSL's data storage needs and the associated need for users to access data remotely. ESnet has played and will continue to play an increasingly vital role in enabling the science for DOE climate and environmental research. As data volume increases for both climate models and the observational capabilities in user facilities, CESD expects increasing pressure to assure that the petabytes of data and model output are readily available to the user community through ESnet.

Biological Systems Science Division

The Biological Systems Science Division supports a diverse portfolio of fundamental research and technology development to achieve a predictive, systems-level understanding of complex biological systems to advance DOE missions in energy and the environment. By integrating genome science with advanced computational and experimental approaches, the division seeks to gain a predictive understanding of living systems, from microbes and microbial communities to plants and other whole organisms. This foundational knowledge serves as the basis for the confident redesign of microbes and plants for sustainable biofuel production, improved carbon storage, and contaminant remediation. ESnet is the primary network provider that enables large scale data transfer for the JGI with the National Center for Biotechnology Information (NCBI) and other key stakeholders.

Systems biology research within BSSD's Genomic Science program is aimed at identifying the foundational prin-

ciples that drive biological systems. These principles govern the translation of genetic codes into integrated networks of catalytic proteins, regulatory elements, and metabolite pools underlying the functional processes of organisms. These dynamic interactions of nested subsystems ultimately determine the overall systems biology of plants, microbes, and multi-species communities. The ultimate goal of the Genomic Science program is to achieve sufficient understanding of the fundamental rules and dynamic properties of these systems to develop predictive computational models of biological systems and tools for rational biosystems design. Genomic Science program research also brings the omics-driven tools of modern systems biology to bear on analyzing interactions between organisms that form biological communities and with their surrounding environments. Understanding the relationships between molecular-scale functional biology and ecosystem-scale environmental processes illuminates the basic mechanisms that drive biogeochemical cycling of metals and nutrients, carbon biosequestration, and greenhouse gas emissions in terrestrial ecosystems or bioenergy landscapes.

The major objectives of the Genomic Science program are to:

1. Determine the molecular mechanisms, regulatory elements, and integrated networks needed to understand genome-scale functional properties of microbes, plants, and interactive biological communities
2. Develop -omics experimental capabilities and enabling technologies to achieve dynamic, systems-level understanding of organism and/or community function
3. Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance predictive understanding and manipulation of biological systems

Case Studies
Biological Systems Science Division

Case Study 1

BioEnergy Science Center

1.1 Background

The objective of the Oak Ridge National Laboratory's (ORNL) mass spectrometry work within the BioEnergy Science Center (BESC) proteomics activity focuses on integrative science to develop and exploit high performance mass spectrometric qualitative, and quantitative approaches for the characterization of microbial and plant proteins for enhanced biofuel production. To this end, the demands on state-of-the-art systems biology capabilities remain substantial and quite dynamic. In support of specified yearly milestones within BESC, it has become apparent that the proteomics research effort must maintain significant fluidity to meet both expected and unexpected/emerging program project needs and dramatically varying sample loads in real-time, both in microbial and plant research. This equates to a balance of big sample campaigns, which may involve dozens of samples and 1–2 months of mass spectrometric (MS) measurement time; and focused measurements, which involve purified protein fractions, gel brands, etc.

1.2 Network and Data Architecture

Each high-performance mass spectrometer is equipped with a Windows-based operation computer, which contains all the vendor software for instrument operation and data collection. Dual hard drives (typically 1 TB each) per computer are redundantly arrayed with independent disks. Because vendor software typically lags behind current operating systems, many of these instrumentation units cannot be upgraded or patched with latest software builds or security fixes, as they inhibit (or kill) instrument operation. Thus, we have taken all instruments computers off the ORNL network. Raw MS data files (typically about 2 GB in size) are transferred via portable drives to desktop computers. With all our MS instruments combined, we probably generate up to 15–20 GB of raw MS data per day. The raw data is then processed and uploaded to servers at ORNL or the University of Tennessee, Knoxville (UTK). Uploads at ORNL are typically about 1 Gbps. Thus, there is not an appreciable overhead time for uploading raw MS files. Search results are significantly smaller (tens of megabytes), so downloads are pretty fast as well.

Server hardware accessible for MS proteomics at ORNL include the Viper and the Compute and Data Environment for Science (CADES) systems. Viper consists of a multi-core AMD Opteron processor architecture (2.3 GHz), with long, medium, and large queues that provide up to 64 processors (43 nodes) and 640 cores. The file storage system (Panasas) is dated and is no longer supported by the vendor. When this crashes, it is unclear if Viper will be operational. At present, the work load on Viper is constant and heavy, making it difficult to get reproducible access.

The CADES system at ORNL is an open stack configuration, with processors controlled by various work groups. At present, the Biosciences Division has access to up to 456 processors, with storage up to 750 TB (a plan is underway to upgrade this storage to 1.5 PB). This entire system is much more state of the art than Viper, but also has an active list of users, so access time can be problematic.

At present, most of the data transmissions at ORNL (internal and external) are limited to 1 Gbps. This is not a limitation with the current range and type of MS equipment for proteomics. ORNL does have 100 Gbps access to ESnet.

The primary data workflow at ORNL consists of moving gigabyte-MS files of raw data to desktop computers and local compute servers. Uploads are on the order of minutes to tens of minutes, while searches may take a few days, depending on the sample composition. These somewhat massive raw MS data sets are usually never sent directly to collaborators at other institutions; rather, only the filtered search results (megabytes in size) are the main item of interest for them. Thus, the need to transfer large data sets externally is minimal for current proteome operations.

In order to publish journal articles, it is essential to upload the raw MS data files to appropriate repositories. This is a bit more onerous, especially with respect to negotiating the ORNL firewall system. Uploads here are usually conducted at the 1 Gbps ORNL rate.

1.3 Collaborators

Collaborators for BESC at ORNL include Adam Guss, Jim Elkins, Steve Brown, Tim Tschaplinski, and Jerry Tuskan. Externally (including those from academia), collaborators include Lee Lynd (Dartmouth), Kelly Craven (Noble Foundation), Robert Kelly (NCSU), Mike Adams (UGA), Mike Himmel (NREL), Deb Mohnen (UGA), and Cong Trihn (UTK).

1.4 Instruments and Facilities

1.4.1 Present

For the Proteome Mass Spectrometry and Informatics Facilities at Oak Ridge National Laboratory, the Organic and Biological Mass Spectrometry Group occupies three laboratories (one large open-bay lab with MS instrumentation, plus two sample preparation labs) and another three labs in the Joint Institute for Biological Sciences at ORNL (two MS instrumentation labs and one sample preparation lab). High-throughput instrumentation for protein identification includes seven Thermo Electron mass spectrometers with nano-electrospray ionization sources: one LCQ Deca XP-Plus system, one triple quadrupole instrument, and five LTQ linear ion trap instrument (two with an electron transfer dissociation, ETD, capability). Each of these mass spectrometers is interfaced via electrospray with high-performance liquid chromatography equipment (Dionex Ultimate HPLC) to perform on-line single- and multi-dimensional separations of complex mixtures of peptides. For high-performance measurements, there is a 9.4 Tesla IonSpec Fourier transform ion cyclotron resonance (FT-ICR) instrument equipped with an electrospray and nanospray ion source, and four Thermo Electron LTQ-Orbitrap hybrid mass spectrometers (one with ETD capabilities and two are new LTQ-Orbitrap-Velos-Pro designs) equipped with an Eksigent nanoflow high-performance liquid chromatography (HPLC) system with autosampler as well as electrospray/nanospray sources.

For computing resources, all large-scale proteome informatics including data processing, database searching, quantifications packages, data dissemination (via web-based portals) and data storage is handled in automated UNIX and MS Windows-based proteome informatics platforms. The main data processing is accomplished with a variety of computer cluster systems, either at ORNL (Viper) or University of Tennessee (Newton). Post-search data analysis is largely conducted on desktop, multi-core Windows-based computers. We have also invested in a robust data storage environment consisting of a BlueArc storage system capable of storing up to about 20 TB of data. We have developed a client-server system capable of high throughput, distributed processing for several genomics and proteomics analysis tools. We have developed a comprehensive web site for public dissemination of data related to our publications.

Personal computers with Internet access are available to all investigators for instrument control, data analysis, word processing, etc. Six dedicated dual-processor desktop computers for analysis of protein tandem mass spectrometry data are equipped with MyriMatch, BioWorks, and RELEX, and other software for protein identification through database searching for Windows-based applications and interfacing with our UNIX- based system.

Major Equipment:

- 5 - quadrupole ion traps (LTQ) (2 with ETD and all with 2-dimensional HPLCs and nanospray)
- 1 - 9.4 Tesla FT-ICR-MS (with HPLC, ESI, IRMPD and ECD)
- 4 - LTQ-Orbitraps (1 with ETD and all with 2-dimensional HPLCs and nanospray)
- 1 - Triple quadrupole mass spectrometer (for targeted quantification)
- 8 - 10 HPLCs
- 1- Advion Nanomate automated robotic nanospray
- 1- GELFREE 8100 system

1.4.2 Next 2-5 years

Newer MS equipment, such as the ThermoFisher Q-Exactive and Fusion platforms, are becoming commercially available, but will only modestly increase the file sizes (tens of gigabytes per measurement at most).

1.4.3 Beyond 5 years

It is not apparent that there will be a significant jump in MS technology in the next 5-7 years, but that is difficult to predict.

1.5 Process of Science

1.5.1 Present

Past successes have fueled an increased engagement of proteomics capacities, with several large-scale campaigns each year. In fact, the level of collaboration engagement for proteome work in the last three years has risen significantly over what was needed in the early phases of the BESC project. To meet these broad and dynamic needs, BESC proteome workflows fall into two broad categories:

1. Global proteome characterizations: Engagement of proteome and integrated omics approaches to assist comprehensive understanding and optimization of engineered microbes, and integrated omics to evaluate natural phenotype variants of plants, with a focus on variants that range in lignocellulosic content (in particular, for poplar top lines).
2. Targeted identifications / quantifications: Deployment of a proteome-based total microbial cell density assay to relevant BESC fermentations, a customized MS approach to verify identities and integrities of purified proteins, gel bands, fractions, etc. from selected sample preparation protocols, engagement of targeted proteome approaches for specific quantification determinations of key natural and engineered protein complexes (i.e., natural and artificial cellulosomes).

1.5.2 Next 2-5 years

It is not clear that the current level of effort and MS focus will change significantly over the next 2–5 years. Projects come and go as funding permits, so it is difficult to estimate the nature of the research workflow. MS capabilities at ORNL continue to be heavily engaged, so there is most likely the possibility of adding additional MS hardware, as the current LC-MS/MS measurements are not high throughput. This will generate a modest increase in computational needs and data transfers, but it is not expected to create an unsurmountable bottleneck.

1.6 Software Infrastructure

1.6.1 Spectral assignment by database searching

Peptides are matched to MS/MS spectra using MyriMatch v2.1. For database searching, microbial or plant genomes are translated and annotated into FastA proteome databases, which are then appended with common contaminants, including the sequences for trypsin and α-chymotrypsinogen, then concatenated with reversed entries to assess false-discovery rates (FDR). Search parameters include unlimited miscleavages (with an upper Dalton limit of 10,000) for each specific protease used in the analysis. For trypsin fractions, peptides were required to contain at least one tryptic ends (semi-specific; K or R). Peptide modifications included in each database search included: A static +57.0214 Da on cysteines (carboxamidomethyl by IAA), a dynamic +43.0082 Da on peptide N-termini (carbamylation via urea breakdown), and a dynamic +15.9949 Da on methionine to account for sample-induced peptide oxidation (max dynamic mods = 2).

1.6.2 Bioinformatic tools employed for data analysis and interpretation

IDPicker v. 3.0 is used to assemble identified peptides into proteins and filter the data for subsequent analysis. Metrics for individual sample runs are tabulated by IDPicker after adjusting the filters to maintain FDR at acceptable rates (approximately 0.5% at PSM-level, 1% at peptide-level, and 2-5% at the protein-level) mainly by adjusting the number of assigned spectra per protein on a sample-by-sample basis). Other filters remain constant: 1 spectrum per peptide, 2 distinct peptides per protein, and q-value ≤ 0.02. Assignment frequencies for each sample are assessed and compared across fractions and include a deeper analysis of high-quality peptide spectral assignment by ScanRanker. Data are merged together and statistical distributions are compared across protease fractions. For semi-quantitative analyses and comparisons, peptide and protein matched-ion intensities are calculated for each PSM and tabulated at both peptide and protein levels. Sequence coverage analyses include helical overlap propensities in predicted membrane proteins and protein-level hydrophobicity assessments were aided by TMHMM transmembrane domain prediction and Kyle–Doolittle hydropathy scores. Peptide coverage maps, proteolytic cleavage propensities, and amino acid frequencies are assessed in Microsoft Excel by combining peptide data from IDPicker with Transmembrane Helices Hidden Markov Models (TMHMM). Venn diagrams for peptide-level comparisons of each proteolytic fraction are created using eulerAPE.

1.7 Cloud Services

Minor attempts have been made to utilize Amazon cloud computing on a limited basis, but it has not been explored on a large-scale or sustainable level.

Big questions remain about the future of cloud computing for proteome research, in terms of uploading data, processing data, storing/accessing data, and disseminating data. At present, there is excitement about not maintaining local hardware, but the current practical costs of cloud computing are unacceptable.

Table 1.1: The following table summarizes data needs and networking requirements for BESC.

Key Science Drivers			Anticipated Network Needs	
Instruments, Software, and Facilities	**Process of Science**	**Data Set Size**	**Local-Area Transfer Time**	**Wide-Area Transfer Time**
0-2 years				
· What are the current/new instruments and data sources? ThermoFisher Orbitrap systems. · What is the current/new software used in scientific process? Commercial (Sequest, Mascot, MyriMatch) and customized packages	· Highlights of currrent science process	· What is the size of one data set? (E.g. 5TB/set) 2-4 GB · What is the general range of data set sizes? (E.g. 500GB to 2TB depending on experiment?) 10 MB - 5 GB · What is the data set composed of? (lots of small files in one data set)	· How long does it take to transfer a data set on the local network? Minutes - tens of minutes · How frequent are the transfers? once per day	· How long does it take to transfer a data set offsite? 30-60 min. · How frequent are the transfers? twice per week · Where are the collaborating sites/destination points for the data transfers/data sets? To computer server and back to ORNL
2-5 years				
· What are the planned, new data sources/instruments? ThermoFisher Q-Exactive and · What are the planned/expected software packages? (Unknown)	· What are the foreseeable changes to data flow, science process, etc? Not much, except 2-5X in size and speed of data acquisition	· Size of one data set: 5-20 GB · Range of data set sizes: 1-20 GB · Data set composition: lots of small files in one data set	· How long does it take to transfer a data set on the local network? desire tens of GB/sec · How frequent are the transfers? once per day	· How long does it take to transfer a data set offsite? desire to be < 5 min · How frequent are the transfers? once per day
5+ years				
· Describe any planned new data sources or software packages	· What is the strategic direction for data flow, science process, etc.? More seamless from exptl. Design to data dissmemination. · What is the strategic direction for data flow, science process, etc.? More seamless from exptl. Design to data dissmemination	· Size of one data set (e.g. 5TB/set) 100 GB · Range of data set sizes (e.g. 500GB to 2TB depending on experiment) up to 1 TB		

Case Study 2

Genomics and Environmental Research in Microbial Systems Laboratory

2.1 Background

The Genomics and Environmental Research in Microbial Systems (GERMS) Laboratory at Iowa State University believes that humans are changing the environment that we live in, and we must understand and manage the impacts of global change. Our goal is to provide scientific research that can inform decisions and policy. Specifically, we integrate traditional microbiology approaches with high-throughput sequencing approaches and computational biology as investigative tools to understand natural and engineered microbial populations. Our data is mainly comprised of sequencing data sets (varying technologies), metabolomics, and experimentally associated environmental factors or treatments. The integration of these heterogeneous data sets is critical for our scientific process. Our goal is to use these data sets to understand and manage microbial interactions that impact our lives.

2.2 Network and Data Architecture

Our group's network and data architecture consists primarily of cloud computing resources provided by Argonne National Laboratory (a collaboration associated with DOE BER Award Number SC0010775). Consequently, our local resources consist of personal computers that connect to on-demand cloud resources. Our scientific process requires data transfers from sequencing facilities, collaborators, public data sets, and between multiple cloud servers. Data is moved between data streams via standard protocols (HTTP/S, FTP, and SCP/FTP).

2.2.1 Local Network and Data Architecture

Our group does not currently maintain any local compute infrastructure beyond laptops (4–16 GB memory). Data products greater than 10 GB are not stored on these computers. Data that are stored on these machines are mainly visualization products (e.g., graphs, figures, etc.) from data analyses that are performed on remote resources. Local data is backed up on local hard drives as well as Google Drive. This data is shared with Google Drive and/or version-controlled repositories such as Github.

The Iowa State HPC architecture is available (with a buy-in) as well (see Table 2.1).

Our local network architecture in our campus building has a Cisco Catalyst 3750-X switch stack that is dually attached (1 Gbps or 10 Gbps) to a pair of Cisco Nexus 7706s. The 7706s attach to our core routers over the campus backbone (10 Gbps). The core routers then attach to the border Cisco ASRs (100 Gbps) that then go out to the Internet.

Table 2.1: Iowa State University HPC available resources.

No. of Nodes	Processors per Node	Cores per Node	Memory per Node	Inter-con-nect	Local Scratch Disk	Configuration of Node
256	Two 2.2. GHz 4-Core Intel Opteron 2354	8	8 GB	20 GB IB	150 GB	Normal compute with MPI
60	Two 2.2 GHz 4-Core Intel Opteron 2354	8	8 GB	20 GB IB	150 GB	Hadoop

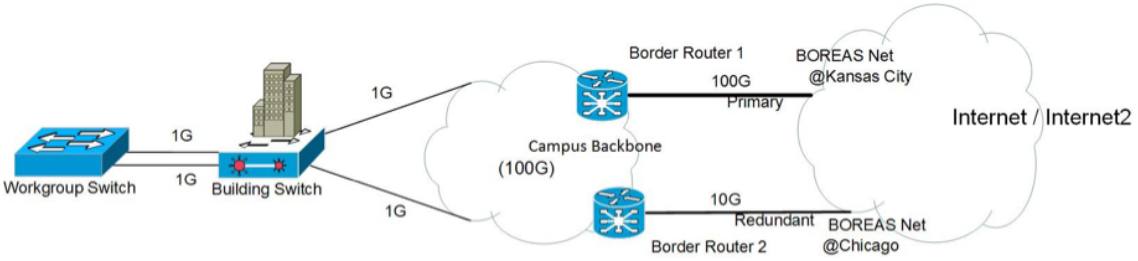


Figure 2.1: ISU local network architecture diagram.

2.2.2 Remote Network and Data Architecture

Virtual machine instances in the Argonne cloud are currently connected via a 1 or 10 Gbps connections. A 10 or 40 Gbps SDN-enabled network is also being evaluated. Future deployments would likely employ a similar network. Network access control lists (Security Groups) and Network Address Translation (NAT) is provided by the Openstack network nodes in a redundant configuration. These network nodes utilize 4x10 gigabit connections upstream to the border router and downstream to the access switches. High-performance data transfers will be enabled using Globus and GridFTP. This will be deployed across both virtual instances and discrete, purpose-built data transfer nodes (DTNs).

2.3 Collaborators

Our collaborators include Argonne National Laboratory ($n = 5$, where n is a Principal Investigator), the Joint Genome Institute ($n = 2$), Pacific Northwest National Laboratory ($n = 2$), and investigators at Iowa State University ($n = 6$), Michigan State University ($n = 3$), the USDA Agricultural Research Service ($n = 3$), the Academy of Sciences of the Czech Republic ($n = 2$), the University of Chicago ($n = 1$), the University of Illinois-Urbana ($n = 1$), and the University of Minnesota ($n = 2$). Data products are regularly requested from and to these collaborators.

2.3.1 Present

Currently, our group regularly uses cloud instances ranging from 22 GB, 8 VCPU ($n = 3$) to 248 GB, 8 virtual CPUs ($n = 1$). We use 65 TB of storage, which are backed-up manually (30% of this storage is data that needs to be backed-up). The high memory instance is used mainly for a key part of our scientific process called genome assembly. This requires that each sequence in a data set be compared to all other sequences in the data set (possibly terabytes in size) and is a computationally intensive process. The remaining instances are used for other parts in our scientific process (see Figure 2.2). In the next two years, we expect the memory requirements for assembly to decrease, though not significantly. We also expect that the numbers of samples and volume of sequences will also scale up, requiring increased capabilities of parallel processing though not necessarily increased memory

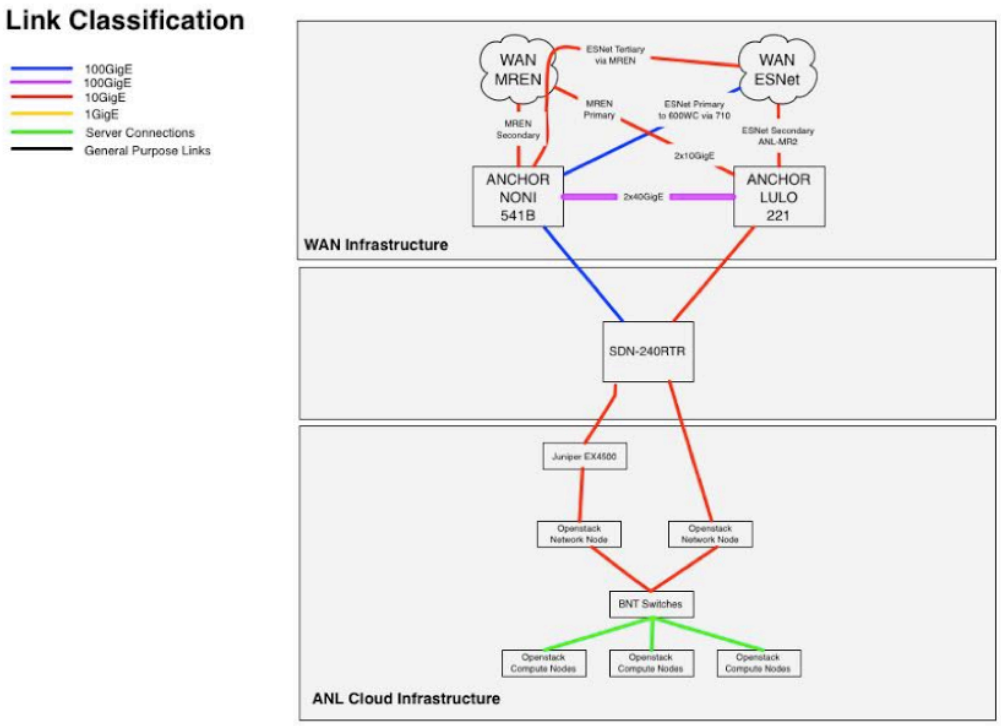


Figure 2.2: ANL WAN and cloud infrastructure.

requirements. We anticipate that storage needs will increase exponentially in the short term—likely doubling every year.

2.3.2 Next 2-5 years

We expect to double the need for resources in the next two to five years, particularly in the number of nodes and cores we require. We anticipate moving from the ANL cloud compute instances to Iowa State HPC, as the funded collaboration with Argonne will conclude. This will require a significant purchase into the ISU HPC for dedicated access and also provide long-term access to these compute resources, which we anticipate being sufficient for the next five years. We expect our required storage space to grow exponentially, likely doubling every one to two years for the next five years (500–1000 TB). Additionally, we anticipate that the number of incoming data streams will increase, extending into accessing and analyzing publicly available data sets. These data sets will most likely be stored on independent servers or public repositories (e.g., NCBI, EBI, MG-RAST). We anticipate that our analysis workflow will be largely automated, and we will need to identify new approaches to enable an automated and efficient release of large data products to the public.

2.3.3 Future

Based on the production of sequencing technologies, it is estimated that worldwide capacity for sequencing currently exceeds 35 TB a year. Though much of this sequencing capacity is currently concentrated at research institutions, hospitals, and large companies (agriculture and biotech), the decreasing costs of this data will make it increasingly democratized and available to independent laboratories in the future. Given this growth, we anticipate that storage will continue to exponentially increase in its requirements, and we will need novel data

compression approaches and/or become increasingly selective of the data we store. We anticipate that the availability of local resources to provide parallel computing opportunities may be restrictive for our research and will likely also use elastic cloud computing for our data processing going forward. We anticipate that the types of heterogeneous data sets that we obtain will increase and include real-time, sensor data to complement molecular approaches. Analyses of these data sets will require obtaining and integrating data from an increasing number of sources (e.g., sensor networks, sequencing facilities, and public databases).

We also anticipate that there will be an increasing number of informative public data sets that will need to be queried, obtained, analyzed, and integrated into our scientific workflow. These data sets will likely be in non-standard formats. At the same time, we will increasingly need to provide our data sets to others with compatible standards. These exchanges of data will require novel methods of data exchange as well as approaches to track the provenance and usage of data to encourage sharing and ensure appropriate credit can be provided. In summary, we anticipate that we will have a need to both push and pull an increasing number and volume of heterogeneous data sets that will require multiple resources, many of which will need to be searchable. For example, we would want to compare local Iowa water microbial communities to water communities in Minnesota. The availability of a federated search that could identify all projects that occurred within constrained Global Positioning System coordinates would be a significant transformative resource for our scientific process. Alternately, we could use such a search constrained to specific sequences to identify their global presence. Michael Schatz provides an insightful perspective that provides a helpful discussion of this vision and its computational requirements.[1]

Another key obstacle for the future of our science, and arguably the main obstacle for our field, is that domain experts are not adequately trained to contribute to and benefit from opportunities of these data sets. Even as novel tools are developed that may provide solutions to many problems, they can only be used by a small number of groups who have adequate training for their integration.

2.4 Process of Science

Our process of science can be summarized in the following steps and shown in Figure 2.3:

- Prepare the genomic representation of experimental microbial communities
- Obtain gene sequences from a sequencing facility
- Extract information from genomic sequences through assembly and/or alignments to known or assembled references
- Increase the quality of these gene sequences, including but not limited to assembly of sequence read fragments that overlap
- Estimate abundances of genes for each sample in experiment
- Integrate abundance, annotation, and metadata or data from collaborators

2.5 Remote Science Activities

Our collaborators at academic universities regularly provide sequencing and sensor data (both real time and independently sampled) that are required to be integrated with our sequencing data sets. The transfer of this data depends on having access to our cloud instances by providing temporary SSH keys. This solution is not ideal, given that it requires significant user training on both ends and access to an instance that cannot be compromised.

[1] http://www.biorxiv.org/content/biorxiv/early/2015/06/02/020289.full.pdf.

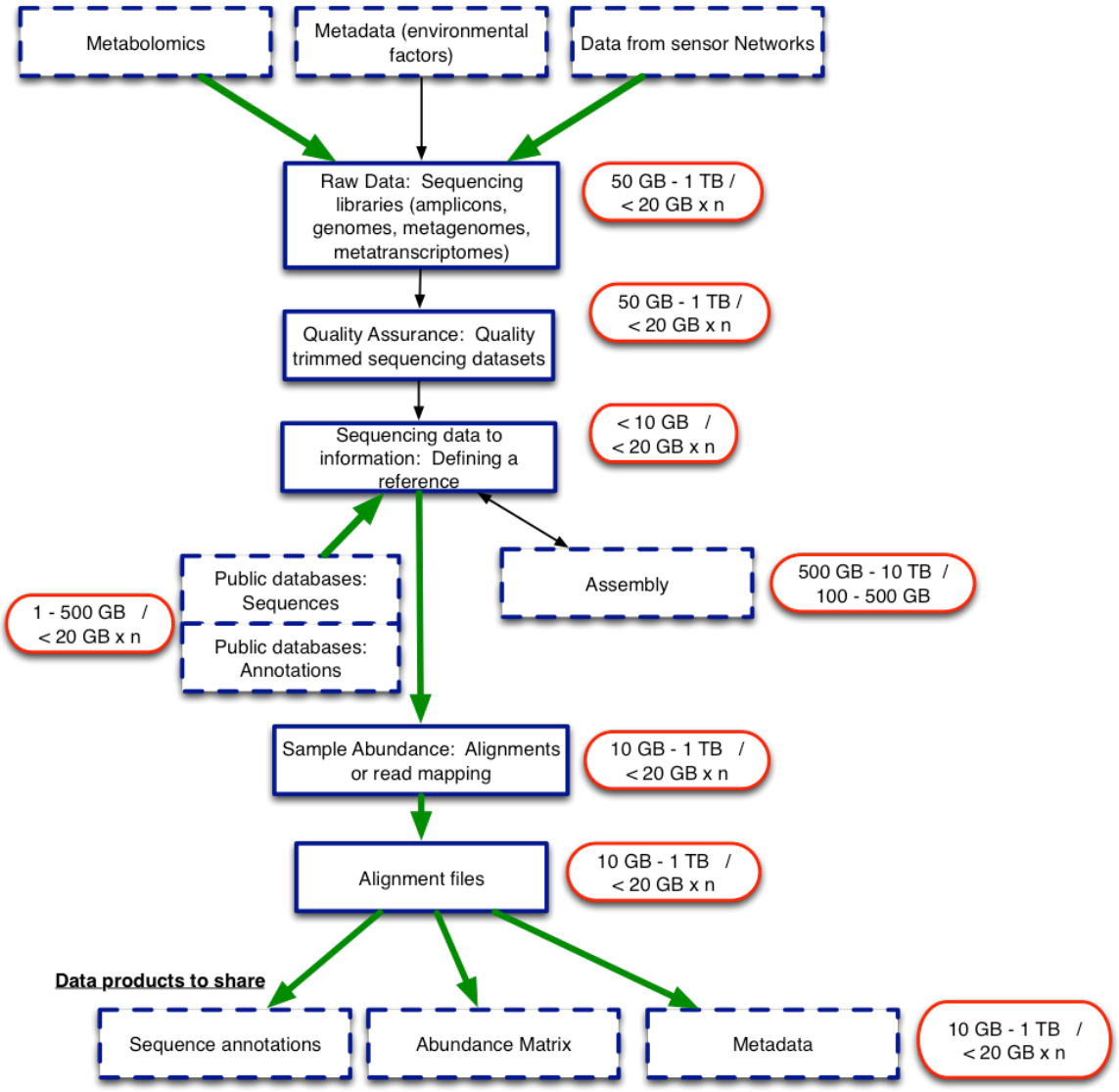


Figure 2.3: GERMS Lab scientific process of sequencing data to information. Green arrows indicate areas where we anticipate significant data transfer challenges in the future. The dotted blue boxes indicate products that are associated with the highest demand to be shared. The red ovals indicate the amount of disk storage / memory times the number of samples required for each step, with the number of nodes, n, expected to double every one to two years.

2.6 Software Infrastructure

Our data transfers currently rely heavily on standard protocols (HTTP/S, FTP, SCP/FTP). When downloading public repositories (from NCBI and MG-RAST), we employ their current APIs but this requires significant software training for students that is consistently and rapidly changing. In limited circumstances, and only on HPC systems, we have used Globus to transfer large data sets, and this was a solution that was facilitated by the local HPC resources. Another regular activity of our scientific process is the compression and/or expansion of large data sets for various software or analyses. Going forward, we are hopeful that there will be software tools that can include online streaming analyses of our data sets so that we will not have to store the data but can begin working with our inferences from these data sets as quickly as possible.

2.7 Cloud Services

Currently, we are nearly exclusively utilizing a private cloud hosted by Argonne National Laboratory. Going forward, we do not anticipate that public cloud services will provide a practical cost solution for the volumes of data we need to process, and consequently, we are anticipating investing in local HPC solutions. However, local HPC solutions currently provide limited access for collaborators, and in this case, we may turn to the cloud services. Our preference is to have a single compute solution, as data transfer and storage becomes increasingly costly with more compute.

2.8 Outstanding Issues

Software and hardware solutions that are not accompanied with accessible documentation or training for our students are challenging to integrate into our workflows. Although my group is increasingly multi-disciplinary, our collaborators may not have similar backgrounds or training, and we are regularly challenged by a need of training with even standard large data transfer protocols in this domain.

Case Study 3

Great Lakes Bioenergy Research Center Sustainability Research: Biogeochemical Responses

3.1 Background

3.1.1 Overview of the Great Lakes Bioenergy Research Center

The production of abundant, clean, and sustainable energy is among the greatest challenges facing society. To help meet this challenge, the interdisciplinary, multi-institutional team of researchers at the Great Lakes Bioenergy Research Center (GLBRC) is working to create a future in which cellulosic biofuels will be a major contributor to the Nation's energy portfolio. The GLBRC mission is to perform the basic research that generates technology to convert cellulosic biomass to ethanol and advanced biofuels. In pursuit of its mission, GLBRC performs fundamental, genome-enabled research that generates the knowledge needed to sustainably produce biofuels and co-products from lignocellulosic plant material.

GLBRC is emphasizing the deployment of productive cropping systems of perennial plants with improved processing traits and the conversion of conventional crop residues and perennial crops to ethanol, advanced biofuels, and valuable co-products. It is addressing two major knowledge gaps that support the goals of the DOE Bioenergy Research Center program:

1. Sustainable production of crops containing desirable biofuel traits, and
2. Efficient conversion of biomass into fuels and chemicals

3.1.2 Sustainable Production of Crops Containing Desirable Biofuel Traits

The economic and environmental sustainability of cellulosic biofuels depends greatly on the choice of feedstock crops: how they are produced, and whether they compete for land with food production. Current agricultural crops were developed mainly for food or fiber production. Optimal sustainable biofuel crops will have different traits than those of crops used solely for food. Consequently, GLBRC is focused on creating crops with traits that improve their value as biofuel feedstocks.

One long-term goal of the center is to understand (1) the attributes and mechanisms responsible for environmental sustainability of biofuel production systems and (2) socioeconomic factors, such as incentives and policy options, that will lead to their acceptance. Most GLBRC field research is carried out at three different scales: (1) small plots for measurement-intensive experiments, such as the Biofuel Cropping System Experiment (BCSE) (Figure 3.1) replicated at the Kellogg Biological Station (KBS) in Michigan and the Arlington Agricultural Research

Station (AARS) in Wisconsin; (2) scale-up sites (Figure 3.3) for field-scale carbon (C) balance experiments using eddy covariance towers (Figure 3.2) at KBS, and (3) extensively dispersed fields for biodiversity investigations across the landscapes of central Michigan and Wisconsin.

In addition, four multi-county areas in Michigan and Wisconsin (two north and two south) are being used for full life-cycle assessments of different cropping scenarios using existing land use, infrastructure, and demographic properties. Data from all three scales are used to perform regional-scale modeling, focusing particularly on the North Central Region of the United States. In 2012, GLBRC initiated a marginal lands experiment (MLE) on plots of seven perennial crops on previously fallow land at three sites along a latitudinal gradient in both Michigan and Wisconsin (Figures 3.4 and 3.5).

Sustainability research in the GLBRC comprises six projects: (1) novel production systems, (2) microbe-plant interactions, (3) biogeochemical processes and responses, (4) biodiversity services, (5) economic services, and (6) biophysical, economic, and life cycle modeling drawing on data generated by the others.

This case study focuses on the data generated by the Biogeochemical Responses research project, with a particular focus on the experiments and measurements conducted in the Michigan State University (MSU) portion of the project. The main objectives of this project are to better understand how cellulosic biofuel cropping systems compare with respect to: (1) nutrient conservation and water use efficiency, which regulate productivity and affect production costs; and (2) global warming impacts, which inform decisions on energy policy. This research focuses on greenhouse gas fluxes and agricultural yields from continuous corn, corn-soybean rotations (with and without cover crops), and six perennial cropping systems: switchgrass, miscanthus, hybrid poplar, mixed native grasses, old fields (successional), and native prairie. A diverse array of data types are collected in this project, including: soil water quality, greenhouse gas emissions, soil carbon and nitrogen, soil water content, and eddy covariance measurements of ecosystem carbon balance (see Figure 3.2 and Section 3.5).

Figure 3.1: Bioenergy Cropping System Experiment (BCSE) depicting eight cropping systems treatments replicated five times in 30 x 40m plots.

3.2 Network and Data Architecture

The KBS network supports 1-gigabit connections at the main buildings and to the GLBRC BCSE agricultural fields. Wireless connectivity is available throughout the research facility and at all of the GLBRC field sites. Figure 3.6 shows the connectivity on the KBS campus.

Figure 3.2: Eddy covariance tower at scale-up fields (see Figure 3.3) used to estimate C balance of annual and perennial cropping systems.

Figure 3.7 shows the fiber connections from KBS to Western Michigan University (WMU) and the connection from WMU to MSU via Grand Rapids. The WMU Grand Rapids-MSU connection utilizes the MERIT (Michigan Educational Research Information Triad) network. The WMU KBS connection is provided by CTS (Climax Telecom Systems). There is a 1-gigabit connection throughout the route.

3.3 Collaborators

The sustainability biogeochemistry research group has a replicate BCSE at the UW's AARS. UW data are stored on a KBS GLBRC database and made available for download (for GLBRC Sustainability researchers) from the GLBRC Sustainability Data Catalog.[1] Data sets are small enough that UW can email files to the KBS Database manager.

The biogeochemical responses project is actively collaborating with all of the other GLBRC sustainability research projects to ensure that GLBRC sustainability research is integrated and synergistic. These projects are located mainly at MSU, KBS, UW, and the University of Maryland. The project is also working with non-GLBRC faculty on soil hydrology and hydrogeology questions that are pertinent to crop water use, including researchers at MSU, the Queensland University of Technology, and UW.

The eddy covariance team has developed several collaborative efforts including modeling integration with the GLBRC sustainability computational modeling group (University of Maryland), data syntheses with the Ameriflux Network, and with researchers in the Biosystems and Agricultural Engineering Department (MSU) to explore the possibility of detecting species composition based on ground-level spectral measurements at mixed prairies.

The project also plans to initiate a collaboration with researchers in GLBRC's deconstruction area (at MSU) to examine N_2O sources in biofuel cropping system soils using isotopomers (i.e., the stable isotope ratios of the two nitrogen atoms in N_2O). As part of this work, the project will take initial isotopomer flux samples in the BCSE, switchgrass N gradient, and rainout shelters for spectroscopic and isotope-ratio mass spectrometry (IRMS) SP analysis, and couple these findings to on-going work with mass fluxes and N process, and metagenomics measurements being made in these experiments.

[1] GLBRC Sustainability Data Catalog can be found at: `http://data.sustainability.glbrc.org/`.

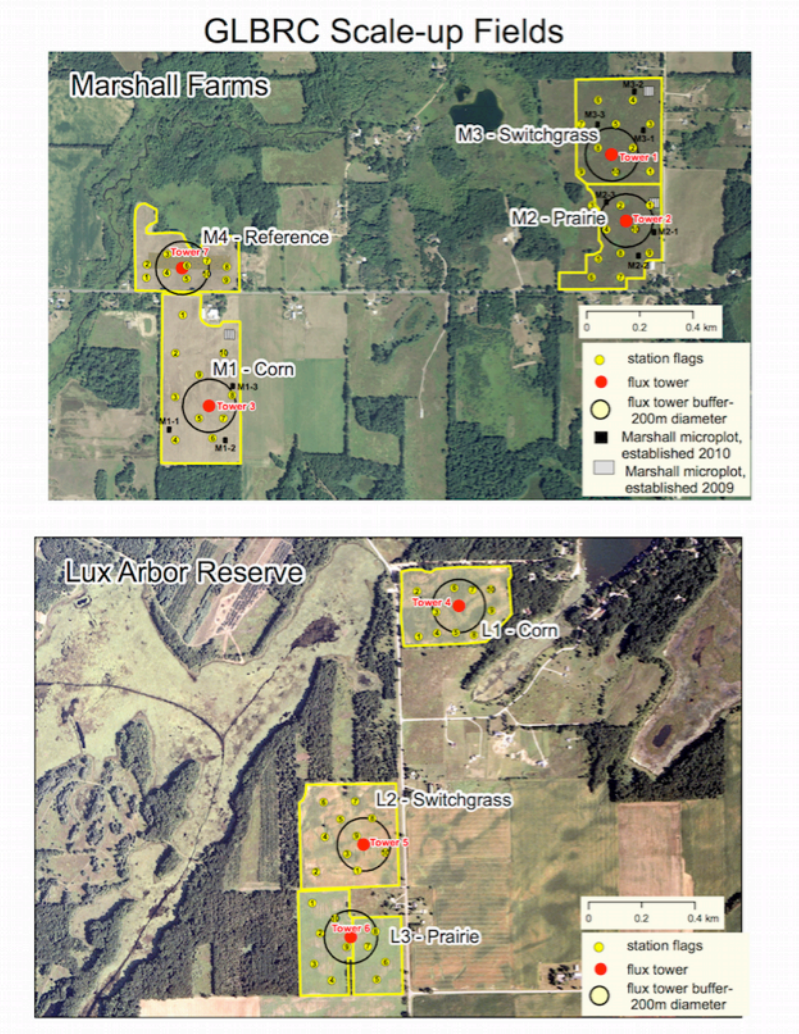


Figure 3.3: Novel cropping systems Scale-up Experiments (Lux Arbor Reserve and Marshall Farms, MI) with 35 to 50-acre fields, 2 sites, 3 cropping systems (continuous corn, switchgrass, restored prairie) at each site and reference. Eddy flux towers continuously measure CO2 and H2O fluxes in and out of plant canopies.

3.4 Instruments and Facilities

3.4.1 Present

Data collected as part of KBS GLBRC research activities are managed locally on PostgreSQL and PostGIS open-source, object-relational, scalable database systems that run in a Linux environment on the database servers. The servers are mirrored on the MSU campus (East Lansing). Postgres has functionality to ensure database integrity and access controls. A local telecom company provides a 10 Mbps backup link. Researchers use specially developed applications that allow them to remotely check their field data and data logger battery status so that instruments have little down time.

A Quantum Cascade Laser (QCL) instrument was recently deployed in the field as a novel means to measure trace gases. This instrument is capable of making continuous trace gas measurements and will produce more and better data in less time, allowing for more in-depth understanding of the patterns and controls of trace gas fluxes. A hotspot to save the data in the field will be installed soon. When the hotspot detects the MSU network, MSUnet, it will load the saved data into the database.

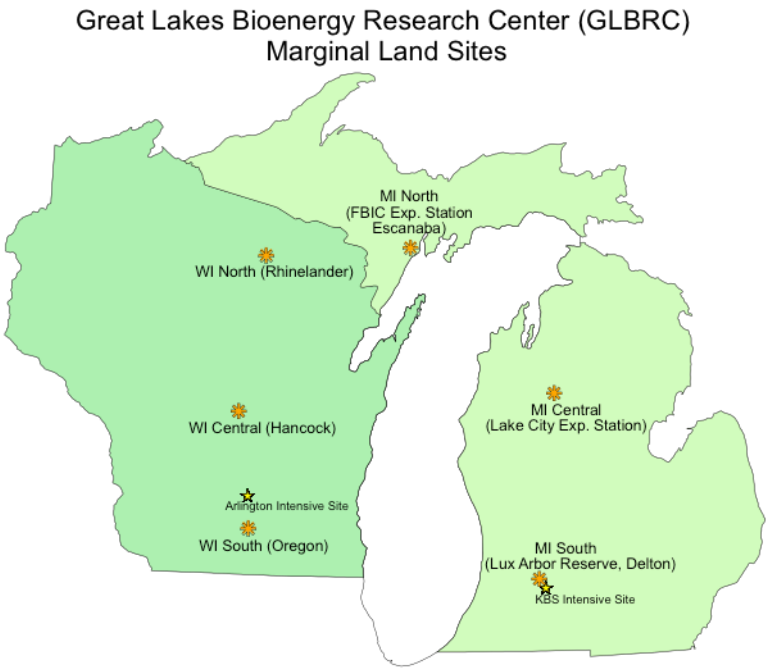


Figure 3.4: Marginal lands experiment (MLE) sites.

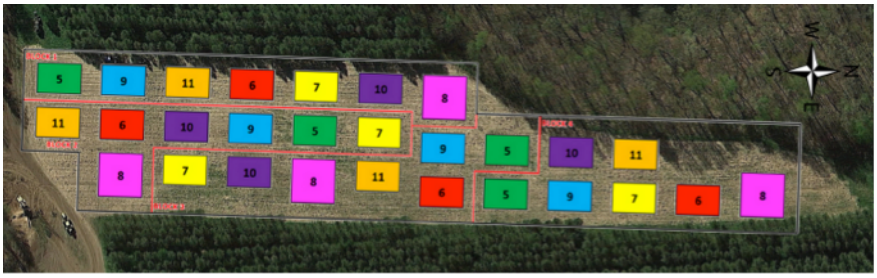


Figure 3.5: Marginal lands experiment MI South site layout (7 perennial cropping systems x 4 replicate blocks: switchgrass, miscanthus, poplar, mixed grasses, successional, and restored prairie).

3.4.2 Next 2–5 years

In the next 2–5 years, the project plans to expand the biogeochemical and hydrological measurements to the GLBRC marginal land experiments, while maintaining some measurements at the main sites for comparison. Ideally similar measurements would be made at the marginal land sites as have been made at KBS and Arlington, although the logistics will be more challenging because of the distances and settings. Some of the BCSE replicates may be converted from cellulosic biofuel crops back to conventional annual crops (corn-soybean rotations) to study the biogeochemical responses and their implications for global warming impact over the entire cropping cycle.

3.5 Processes of Science

3.5.1 BCSE Comparing Candidate Grain and Cellulosic Biofuel Crops

BCSE, which was set up in the first year of GLBRC (2008) at KBS and Arlington, continues to be the centerpiece of GLBRC's biogeochemical responses research. The main experiments are replicated, randomized plots containing

Figure 3.6: Kellogg Biological Station local area network.

a diverse set of biofuel cropping systems including candidate cellulosic crops as well as grain crops (corn and soybean) that are currently used as biofuel feedstocks. Background soil characteristics and spatial variability were documented in detail at the outset and will be re-examined in 2016–2017. The field measurements necessary to understand biogeochemical responses as the experimental plots became established were begun in 2009 and have been maintained to the present.

In addition to the main experiments, switchgrass is grown across a nitrogen fertilization gradient to examine the yield response as well as how nitrate leaching and nitrous oxide emission increase with fertilization. A corn/soybean/wheat rotational crop system across a similar N gradient, crossed with irrigation vs. no irrigation, is maintained at the KBS Long-Term Ecological Research (LTER) site and serves to compare grain crops to switchgrass.

This is designed to be a sustained measurement program that will continue through the life of the project, enabling examination of the short-term as well as more gradual, protracted responses of the soil-plant system to the establishment of new crops and cultivation regimes. Multiple years of observations also encompass the range of climate variability, which is important for crop water availability and phenology. Biogeochemical and water sampling continue outside as well as during the growing season unless precluded by weather conditions. When the crops reach the end of their most productive period (usually 10–15 years after establishment), researchers will begin converting the experimental plots back to conventional corn-soybean production systems. This will allow study of the implications of transitioning out of cellulosic biofuel cropping systems, which is as important

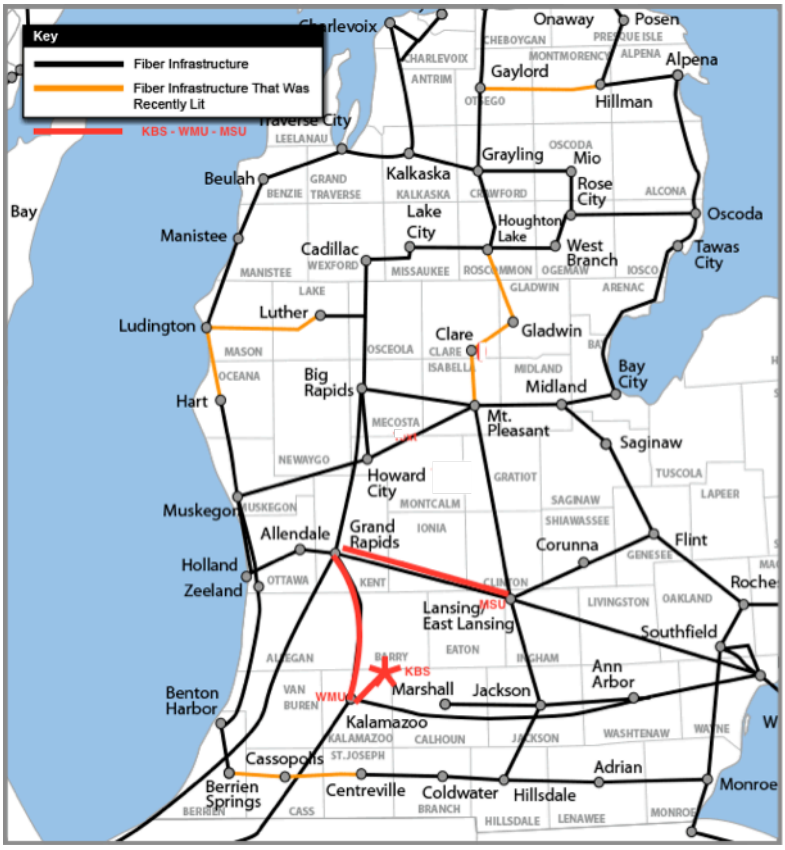


Figure 3.7: Michigan MERIT network with KBS, WMU, MSU connection shown in red.

to a comprehensive assessment of sustainability as the work to date on conversion of croplands or grasslands to biofuel crops.

3.5.2 Exchanges of CO_2 and Water at Whole-field Scales by Eddy Covariance

This research has examined the magnitudes and dynamics of net ecosystem production (NEP) and soil carbon balances. The energy balance is evaluated as well to study implications for radiative forcing as well as to estimate crop evapotranspiration. The net effects of land conversion to, and subsequent management as, biofuel cropping systems on the ecosystem carbon balance can only be assessed over short time scales through eddy-covariance approaches. The six scale-up fields at KBS have been monitored during and after their conversions to corn, switchgrass, and mixed-species restored prairie systems, together with a reference site left in Conservation Reserve Program (CRP) grassland, since January 2009. Three of the converted fields were formerly CRP grasslands for two decades; the other three were used for row crops.

An open-path eddy-covariance (EC) tower (Figure 3.2) and four respiration chambers with root exclusions were installed at each of the six scale-up plots (two replicates of switchgrass, restored prairie, and continuous corn fields with a minimum of 12 ha) and at a native prairie reference site at KBS in December, 2008. The EC towers take continuous measurements of the net exchanges of CO_2, H_2O, and energy as well as microclimatic variables. In support of the eddy-covariance work, regular measurements of canopy reflectance, leaf area, and foliar N have been conducted three to four times per growing season. Root increment cores that quantify the annual root production have been installed since 2009.

3.5.3 Soil Organic Matter Dynamics Following Conversion of Grasslands to Biofuel Crop Production

This research investigates details of soil organic matter (SOM) sequestration of carbon in soils across candidate biofuel cropping systems and as a function of crop management. The field work has been carried out in the BCSE, scale-up fields, extensive sites, and LTER experiments, with the primary goal of understanding processes that regulate soil carbon cycling responses to bioenergy crop production. Specific activities have included the following:

1. Evaluate the impacts of soil carbon process changes on the extent and diversity of soil microbial communities that play key roles in maintaining soil fertility;
2. Examine the physical and chemical stabilization of new carbon inputs under diverse bioenergy crop systems;
3. Evaluate bioenergy crop litter decomposition, its microbial controls, and contributions to soil carbon formation; and
4. Conduct stable isotope tracer experiments to determine the feasibility of maintaining soil carbon in bioenergy crop systems using cover crops.

3.5.4 Non-Leguminous Nitrogen Fixation

Nitrogen (N) supply is key to the sustained productivity of grasses such as switchgrass and Miscanthus, and the possibility exists that symbiotic or otherwise closely associated microbes fix nitrogen from atmospheric N_2, as is the case for legumes. Yet nitrogen fixation has yet to be conclusively demonstrated as important to the nitrogen supply of these species. Stable isotope tracing can reveal nitrogen fixation but is technically challenging. In 2014, the project conducted three $^{15}N_2$ addition experiments: whole-plant incubation in a gas-tight chamber (to measure an N_2-fixation rate), incubations with individual plant parts in vials, and *in situ* $^{15}N_2$ additions. In 2015, the $^{15}N_2$ delivery method was refined by doing trial additions with the inert tracer gas SF_6. Then $^{15}N_2$ gas was added to each of 4 replicate plots twice, once during the early growing season (early June), and once just before flowering (late June). In addition, soil N_2 fixation by free-living microbes was measured four times per year, in three nitrogen fertilizer treatments, at both KBS and Arlington.

Biogeochemical Responses Experiments Major Data Types					
Data type	How measured	Measurement frequency	Location	Data volume	Delivery to database
Soil moisture profiles	Time Domain Reflectometry (TDR)	Continuous	KBS + UW intensive sites	Not signif.	Automatic flow to data server
Greenhouse gas fluxes	Static chamber; lab gas chromatorgraphy	Biweekly	KBS + UW intensive sites	Not signif.	Automatic flow to data server
Greenhouse gas fluxes	Automated gas chambers; field gas chromatorgraphy	Several times per day, year-round	KBS Intensive Site	Not signif.	Automatic flow to data server
CO_2 and H_2O fluxes	Eddy covariance	Continuous	KBS Scale-up sites	About 60MB/day	Automatic flow to server
Soil inorganic nitrogen; soil organic carbon and nitrogen	Soil sampling; lab analysis by Autoanalyzer or Elemental Analyzer.	Monthly (inorganic) and annually (organic)	All sites	Not signif.	Periodically; at least annually
Surface soil moisture	Field sensors and loggers	Continuous	KBS Scale-up; UW intensive sites	Not signif.	Periodically; at least annually
Soil water nitrate concentration	Suction soil samplers; ion chromatography or autoanalyzer	Biweekly	KBS + UW intensive sites	Not signif.	Periodically; at least annually

3.6 Remote Science Activities

University of Wisconsin researchers on this project load raw gas chromatograph data into a KBS web application to visualize, clean, and convert the raw trace gas data to gas flux data.

The project's research does not use any major scientific instruments with large data flow at remote locations.

3.7 Software Infrastructure

Presently, researchers work with specially developed applications that allow them to remotely check their field data and data logger battery power so that instruments have little down time. Eddy flux data is analyzed with the EdiRe package (University of Edinburgh, v 1.5.0.32, 2012). Data collected as part of KBS GLBRC research activities are managed locally on PostgreSQL and PostGIS open-source, object-relational, scalable database systems that run in a Linux environment on our database servers. The servers are mirrored on the MSU campus (East Lansing). Postgres has functionality to ensure database integrity and access controls.

3.8 Cloud Services

Amazon S3 is utilized for the sustainability research publication database.

3.9 Outstanding Issues

The marginal lands experimental sites are remote from existing data collection sites.

Site	Distance from (miles)	
	KBS	Madison
MI South (Lux Arbor, Delton)	8	
MI Central (Lake City)	158	
MI North (Escanaba)	435	244
WI South (Lancaster)		81
WI Central (Hancock)		84
WI North (Rhinelander)		205

Case Study 4

Joint Genome Institute

4.1 Background

The Joint Genome Institute (JGI) is a raw data generator as well as a large repository of genomic data. The JGI runs several sequencers nearly 24X7 and these systems send their data to National Energy Research Scientific Computing (NERSC) Center in nearly real time. External collaborators access JGI data through several web portals.

Scientists around the world submit applications to have their data processed and sequenced at the JGI. There is also a growing synthetic biology group that produces a relatively tiny amount of data (tens of gigabytes per year).

In addition to sequencing, the JGI now offers metabolomics analysis and has installed two mass spectrometer machines. Data from mass spectrometers are also transferred to NERSC in nearly real time.

4.2 Network and Data Architecture

The JGI is now connected to the Lawrence Berkeley National Laboratory's network, LBLnet. It has a 2x10 Gbps connection to NERSC and 1 Gbps Ethernet connection throughout the facility. All of the data generated at the center is accessed through web portals that are housed at NERSC. Most data is kept in HPSS, but the JGI also has 7.1 PB of Global Parallel File System (GPFS) storage and some web servers mount these file systems. The JGI also has its own Globus endpoint that they have architected to work with their single-sign-on service. NERSC runs this endpoint on one of the data transfer nodes that are designed for wide-area network transfers outside of the NERSC facility.

4.3 Collaborators

Collaborating facilities include:

- NERSC—runs the computational infrastructure for scientific computing at the JGI.
- Hudson Alpha Institute—plant collaboration; JGI funds several FTEs and Hudson Alpha utilizes JGI's sequencing and compute infrastructure.
- KBase—integrated data transfer; KBase has access to a large amount of JGI's data.
- EMSL—another DOE facility with some sequencing and mass spectrometry capabilities. There were eight new joint projects initiated in 2015.

- Web portal traffic—17,000 new users registered for JGI web portal access in 2015, with more than 700,000 unique visitors in 2015.
- Three DOE Bioenergy Research Centers: JBEI, BESC, and GLBRC—JGI performs sequencing and provides data to the centers to further their research in biofuels.
- Emerging Technology Opportunity Program—9 projects funded at University of Washington, Stanford University, MIT, University of Vienna, University of California, Berkeley, University of Arizona, Broad Institute, PNNL, ORNL.

4.4 Instruments and Facilities

- Present
  - 140 terabases of genomic sequence generated on behalf of the BER community;
  - 17 total sequencers generate approximately 5TB/day, and 8 mass spectrometry systems generate approximately 100GB/day;
  - 8400 core cluster, 72 nodes that have more than 256 GB of memory, and one 2 TB node;
  - 7.1 PB of IBM GPFS storage, and 4PB of tape storage in HPSS;
  - Hundreds of terabytes of data downloaded by external users.
- Next 2-5 years
  - Growth of computing cluster is expected to remain flat and will be located in the new Computational Research and Theory (CRT) at LBNL, so their compute and data will be connected to other facilities through ESnet.
  - Sequencing is estimated to stay constant with a potentially smaller data footprint because of increased read length and technology improvements.
  - Web portals will continue to be a critical resource for collaborators and the broader scientific community, and petabytes are expected to be downloaded per year.
- Beyond 5 years
  - New technologies like Oxford Nanopore have the potential to revolutionize how sequence data is generated.
  - JGI may ingest more external data, but will always have strength in scientific research tied to developing a deep understanding of genomics data.
  - Web portals will continue to be a critical resource for collaborators and the broader scientific community; analysis and compute being collocated will limit the need for lots of data downloads.

4.5 Process of Science

- Present
  - Sequencers generate around 5 TB per day; mass spectrometers generates 100 GB per day, and data is streamed to NERSC.
  - User downloads generate a larger load on the network and this varies from gigabytes per day to terabytes per day. Requests are usually limited by the network connectivity of the user doing the download unless they have Globus.
- Next 2-5 years

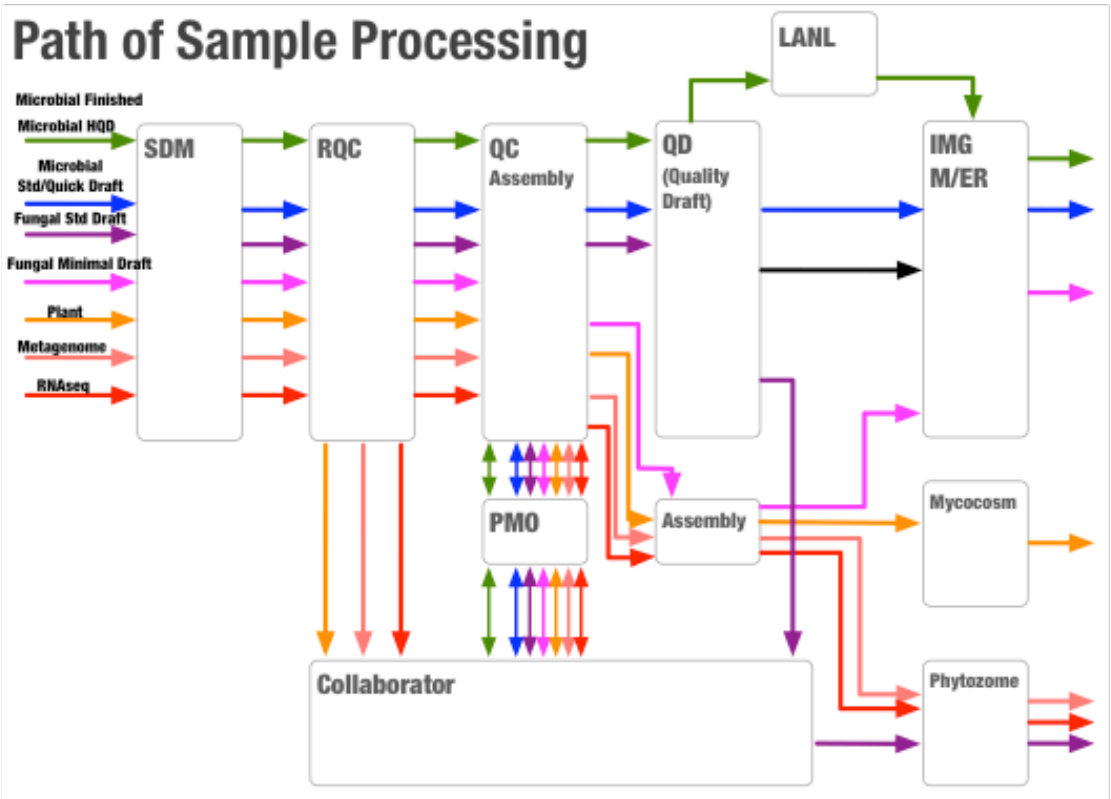


Figure 4.1: Each path represents one product the JGI provides to external collaborators. Each rectangular block shows analysis or interaction with the end users. Most of the external collaboration is managed through the Program Management Office (PMO) the end points Mycocosm, Genome Portal, IMG M/ER and Phytozome are the main JGI web portals.

  - Sequencing will remain flat or decrease; data generated by sequencers may decrease depending on advances in technology.
  - Number of mass spectrometry systems may increase depending on demands from the scientific community; small amounts of data, but new analytical tools similar to OpenMSI[1] will allow scientists to ask deeper questions.

- Beyond 5 years
  - The JGI has a wealth of genomic data that will always be of interest to scientists around the world, but we envision a future where the scientists do not need to download data to do their analyses.
  - There will be global mirrors for data collocated with computational tools.
  - Technology like Oxford Nanopore will enable larger numbers of scientists to contribute data to sites like the JGI, so we will ingest more data from external sources.

4.6 Remote Science Activities

The JGI is connected to NERSC through ESnet. Remote users primarily interact with JGI data through web portals, though some send large metagenomic data sets on the order of terabytes via external hard drives. The analytical tools available to remote users through web portals leverage the compute infrastructure at NERSC to complete some analysis. Older data is stored in HPSS and retrieved through the JAMO system and then made available to users for download.

[1] Information about OpenMSI can be found at: http://crd.lbl.gov/news-and-publications/news/2013/openmsi-a-science-gateway-to-sort-through-bio-imaging-s-big-datasets/.

4.7 Software Infrastructure

- Present
  - The JGI has over 400 software packages installed on their cluster Genepool for analysis; each JGI program has its own software stack because different tools are optimal for different organisms.
  - There is no standard workflow management system—primarily Perl and Python glue code between third-party executables.
  - For sequencer software, Illumina analysis done on the machine, PacBio provides the SMRTanalysis and SMRTportal software used by JGI analysts.
  - Major file formats include ASCII, BAM, SAM, CRAM, FASTA, FASTQ, and many more. Different formats are needed for different types of analysis. Remote users want both the raw data and the analysis that has been done at the JGI (QA/QC, assembly and annotation). See Figure 4.1 for the flow of the various products produced by the JGI through different analysis steps.
  - JGI uses the JGI Archive and Metadata Organizer (JAMO, a tool developed in-house) for data management and provenance.
- Next 2-5 years
  - Software stack likely to become even more varied as tools/heuristics improve.
  - Standard workflow management system built using Docker containers allowing workflows to run anywhere, not just the JGI's cluster.
  - Sequencer software will continue to be developed and maintained by the sequencer companies.
- Beyond 5 years
  - Hopefully we will see some consensus on the best tools for the job and the "best" tools for assembly and similarity search will be adapted for many-core architectures.
  - There will be even more web-based analytical tools that scientists around the world will be able to use.
  - Web portals at the JGI will be modern and able to take advantage of the latest advances in web technology.

4.8 Cloud Services

It is unclear whether or not the JGI will take advantage of the cloud infrastructure at this time. We are currently investigating the feasibility of hosting web services in the cloud instead of at NERSC, but foresee some issues with moving petabytes of data between NERSC and the cloud infrastructure (mostly due to the costs of data storage and access in Amazon Web Services and others). We should have a better idea of what this plan will look like and a full assessment from the JGI at the end of FY16. This analysis is coupled with an effort to examine more scalable database infrastructures.

4.9 Outstanding Issues

We are not currently limited by our network infrastructure; though it will be nice to have JGI, KBase, and NERSC collocated at LBNL. We are watching the increased usage of the data portals and downloads carefully to determine additional resources that may be necessary to support this usage in the future.

Table 4.1: The following table summarizes data needs and networking requirements for the JGI.

Key Science Drivers			Anticipated Network Needs	
Instruments, Software, and Facilities	**Process of Science**	**Data Set Size**	**Local-Area Transfer Time**	**Wide-Area Transfer Time**
0-2 years				
Illumina, PacBio, Oxford Nanopore, Mass Spec · What is the current/new software used in scientific process? **Illumina RTA, basecall, bcl2fastq; PacBio SMRT Analysis**	**Samples are provided to the JGI by external collaborators; DNA is extracted; DNA is sequenced and data is sent to NERSC for subsequent storage and analysis; external collaborators access data through web portals when analysis is completed.**	**Average data set size 541GB** · What is the general range of data set sizes? **16GB to 3.5TB depending on experiment and platform** · What is the data set composed of? **100s to millions of files depending on experiment and platform.**	· How long does it take to transfer a data set on the local network? N/A · How frequent are the transfers? (E.g. three times per day? N/A	· How long does it take to transfer a data set offsite? Total current transfer from all platforms/experiments is **~70GB/hour, individual experiments will transfer in between 2 and 400 hours, depending on experiment and platform** · How frequent are the transfers? **Most of the bandwidth is 7/24 streaming.** · Where are the collaborating sites/destination points for the data transfers/data sets? **NERSC:/global/seqfs; remote users**
2-5 years				
· What are the planned, new data sources/instruments? **Mass Spec, updates in Illumina platforms (HiSeq 4000, additional NextSeq)**	· What are the foreseeable changes to data flow, science process, etc? **Faster sequencers, higher density flowcells/smrt cells/etc, high transfer rate by reduction of reliance on intermediate some intermidate data, improved basecalling/sequence consensus calling software**	· Size of one data set: **16GB to 500GB depending on experiment and platform** · What is the general range of data set sizes? **8GB to 500GB depending on experiment and platform** · What is the data set composed of? **100s to 100s of thousands of files depending on experiment and platform.**	· How long does it take to transfer a data set on the local network? N/A · How frequent are the transfers? (E.g. three times per day? N/A	· How long does it take to transfer a data set offsite? **Total current transfer from all platforms/experiments to JGI cache is ~100GB/hour, individual experiments will transfer in between 2 and 100 hours, depending on experiment and platform** · How frequent are the transfers? **Most of the bandwidth is 7/24 streaming; web downloads are discrete proceesses limited by the end users connectivity** · Where are the collaborating sites/destination points for the data transfers/data sets? **NERSC:/global/seqfs; remote scientists**
5+ years				
	· What is the strategic direction for data flow, science process, etc.? **We would like to enable more analysis through web portals that can leverage the NERSC supercomputing infrastructure. JGI will develop a scalable database infrastructure to facilitate fast queries to answer scientific questions**	· Size of one data set: **16GB to 500GB depending on experiment and platform** · What is the general range of data set sizes? **8GB to 500GB depending on experiment and platform** · What is the data set composed of? **100s to 100s of thousands of files depending on experiment and platform.**	· How long does it take to transfer a data set on the local network? N/A · How frequent are the transfers? N/A	· How long does it take to transfer a data set offsite? **Total current transfer from all platforms/experiments is ~150GB/hour, individual experiments will transfer in between 2 and 50 hours, depending on experiment and platform** · How frequent are the transfers? **Most of the bandwidth is 7/24 streaming; web downloads are discrete proceesses limited by the end users connectivity we expect less whole data set downloads in the future and instead just smaller result sets** · Where are the collaborating sites/destination points for the data transfers/data sets? **NERSC:/global/seqfs, HPSS; remote scientists**

Case Study 5

Department of Energy Systems Biology Knowledgebase

5.1 Background

The Department of Energy Systems Biology Knowledgebase (KBase) is a software and data platform designed to meet the grand challenge of systems biology: predicting and designing biological function. KBase integrates data, tools, and their associated interfaces into one unified, scalable environment, so users do not need to access them from numerous sources or learn multiple systems in order to perform sophisticated systems biology analyses. Users can perform large-scale analyses and combine multiple lines of evidence to model plant and microbial physiology and community dynamics. KBase is the first large-scale bioinformatics system that enables users to upload their own data, analyze it (along with collaborator and public data), build increasingly realistic models, and share and publish their workflows and conclusions. KBase aims to provide a knowledgebase: an integrated environment where knowledge and insights are created and multiplied.

Underlying the KBase platform is a service-oriented architecture that runs across a distributed set of resources located at Argonne National Laboratory (ANL) and Lawrence Berkeley National Laboratory (LBNL). The two sites mirror one another both in data and services. KBase takes advantages of existing connectivity to ESnet at both sites. KBase uses its connectivity to 1) replicate data between the two sites 2) enable users to upload and download data to the KBase system.

5.2 Network and Data Architecture

KBase is connected to ESnet directly at the two major sites (ANL and LBNL). Both sites are effectively connected at 10 Gbps, but the way in which the connect to ESnet is slightly different at the two sites.

At LBNL, KBase is housed with NERSC. This is currently at the Oakland Scientific Facility but will move in late 2015 to the new CRT Facility which has just completed construction. Currently a dedicated KBase router (Juniper QFX3500S) is connected via 10Gb to a NERSC-owned Alcatel-based 100Gb router which is connected to ESnet via a 100-gigabit link. The connection through the NERSC router is virtual local-area-network-based (i.e., the NERSC router does not perform any Layer-3 routing). KBase servers are connected to the Juniper via a Mellanox 40-gigabit Ethernet switch. Currently, the Juniper router has a minimal set of access control lists (ACLs) defined. The Juniper QFX3500s has four 40-gigabit interfaces which are not currently being used, but could be used to increase the bandwidth in the future.

5.3 Collaborators

The KBase project has members from four National Laboratories and a number of universities. They include: ANL; Brookhaven National Laboratory (BNL); Cold Spring Harbor Laboratory; Hope College; LBNL; and Oak Ridge National Laboratory (ORNL). Users of the system span the Nation and the globe.

An important partner facility is the JGI (see Section 5.6). User can use the JGI portal to request data sets to be directly uploaded from the JGI into KBase. JGI resources are located in the NERSC computing facility where the Berkeley resources are also housed. Work is underway to provide similar mechanisms for data generated by the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory.

5.4 Instruments and Facilities

KBase has three classes of resources it makes use of for its services and analyses. They include dedicated hardware (purchased via KBase), the Argonne private cloud (Magellan), and the allocated ASCR compute facilities at ALCF, OLCF, and NERSC.

Table 5.1 summarizes the computational and storage resources currently used by KBase. The network connectivity for the dedicated services is described in Section 5.2.

Table 5.1: Current KBase Resources

	Cores	Memory	Storage
Dedicated Servers	492	4TB	424TB
Argonne Cloud Allocation	1756	5TB	60TB
HPC Allocations	2M core hours		

For the next 2–5 years, KBase has allocated a portion of its budget to maintain and refresh the dedicated hardware. Based on that allocation we expect the dedicated hardware to scale to roughly 3500 cores and 1.5 PB by 2020. In this time-frame we would expect that the connectivity for KBase would be upgraded to at least 40 or 100 Gbps based on ingest rates.

KBase is an SFA on a 3-year cycle. The plans beyond five years are currently not defined. However, if the SFA were to continue, it is likely that KBase would continue to maintain and refresh its server and storage hardware to support continued growth.

5.5 Process of Science

The KBase project is building novel analysis and modeling techniques for biological data, focusing on microbes, plants, and microbial communities, as well as a service-oriented architecture that delivers analysis and modeling services to users. Users either upload their own data sets or make use of data sets already loaded into KBase, and apply KBase operations to these data sets. Developers can also develop new analysis and modeling approaches and integrate them into KBase. The goal here is to provide a common infrastructure for large-scale biological data analysis and model creation and refinement. Improvements developed through these processes will be rolled out for KBase users over time.

5.6 Remote Science Activities

Since KBase is a web-based science platform, all of its users are remote. Users primarily interact with the system through a “Narrative Interface”. This interface is based on the popular IPython/Jupyter platform with significant customizations done by KBase. Users can upload data into the system through this interface, conduct analysis, and download analyzed data. Uploaded data sets can vary significantly in size and quantity. Currently, users

typically upload microbial data sets on the order 100 MB in size. Eukaryotic organisms (plants and fungi) can be significantly larger, on the order of tens to hundreds of gigabytes. Metagenomc data sets can be even larger exceeding 1 TB in some cases. The platform currently lacks strong support for these larger data sets, but the project plans to improve support in the coming year for eukaryotes, and the following year for metagenomes. In addition to genetic data, KBase will be adding support for expression-related data, such as RNA, in the near future. These data sets can also be large in size (around 10–100 GB).

In addition to user-uploaded data, KBase supports the direct import of data from the JGI and NCBI (public data sets). Support is currently limited to isolate microbes, but the project plans to expand support to other data sets.

Finally, KBase has allocations or access to Director's allocations with the ASCR computing facilities. To date KBase has not made heavy use of these resources, but has efforts to run large-scale pre-computing jobs, assemblies, and other specialized jobs. This could lead to more flow between KBase systems and the ASCR centers. However, the KBase hardware is collocated at ALCF at ANL and NERSC at LBNL, so their may be little impact on ESnet.

5.7 Software Infrastructure

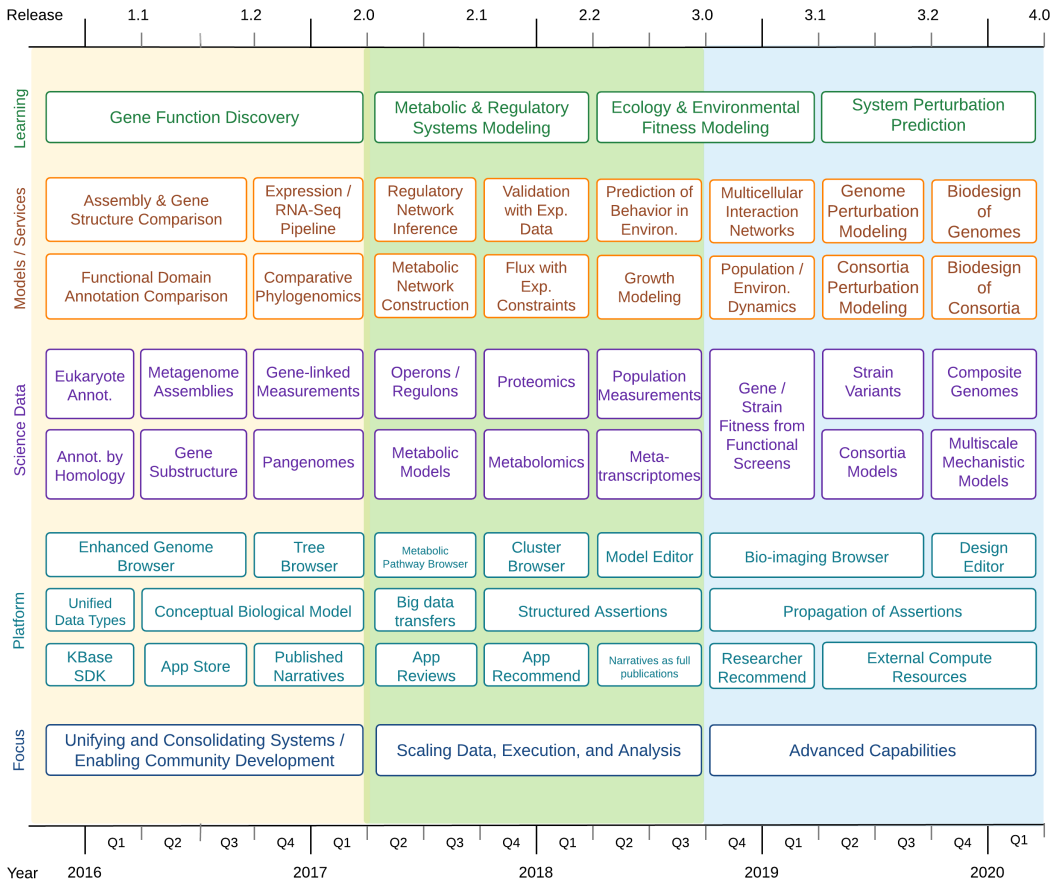


Figure 5.1: The KBase Roadmap through 2020

KBase is a essentially a software development project. Figure 5.1 shows the KBase roadmap. This summarizes the major plans for the KBase platform over the next 4.5 years. In addition to the core KBase platform (e.g., data services, job execution services, Narrative and user interface components), KBase contains various Science Services

that include both JSON-RPC services as well as asynchronous analysis jobs that execute on backend resources (both dedicated and leveraged).

Presently, KBase operates several data systems to store reference data and user data. KBase uses a mix of off-the-shelf data stores like MongoDB and custom software, like Shock and the KBase Workspace. Data is primarily transferred via HTTP(S). KBase does operate Globus endpoints, but these endpoints are not useable by KBase users at present.

Over the next several years, KBase plans to continue to enhance its data stores to meet expected scaling demands. In addition, KBase will be adding support for bulk data uploads. This will likely lead to an increase in data ingest. Finally, the roadmap includes plans for new data types and analysis which will drive data ingest.

Beyond 5 years, the project plans are currently not defined.

5.8 Cloud Services

KBase does not currently make use of any commercial cloud services with the exception of standard cloud-based software tools (such as Github, Docker, and Slack). KBase currently uses the private Magellan Cloud at Argonne. It is possible that in the future KBase may utilize public clouds for handling certain burst activities, but does not have any specific plans to do so at this time.

5.9 Outstanding Issues

Most of the current limitations in uploading large data sets are due to known constraints within the KBase software. The KBase roadmap already includes plans to improve support for these data types. As the limitations are lifted, KBase may encounter new challenges, but the network is not currently imposing any serious barriers on KBase.

Table 5.2: The following table summarizes data needs and networking requirements for KBase.

Key Science Drivers			Anticipated Network Needs	
Instruments, Software, and Facilities	**Process of Science**	**Data Set Size**	**Local-Area Transfer Time**	**Wide-Area Transfer Time**
0-2 years				
· Dedicated Hardware at Argonne National Lab and Lawrence Berkeley National Lab · Data from the JGI and EMSL (in addition to user uploaded data) · KBase currently supports analysis of Sequence data and expression data sets.	· Users upload data or import from other sources (e.g. JGI, EMSL, NCBI).	· Data sizes varify from 10 MB (microbe) to > TB (metagenome) · Anticipate uploads growing to 10TB per month · Formats include FASTA/FASTQ, SBML, GenBank. Future formats include image data (including multi-modal)	· Internally KBase uses 10 Gb/40Gb - So transfers are up to 1 GB/s	· Sites connected at 10Gb allowing for transfers speeds of >100 MB/s. Speeds are limited more by the underlying data structure (lots of small files). · Data is continuously copied between the sites. This is done hourly as new objects are created in the data stores. · Data can be uploaded from users which span the globe. Large data sets may come from JGI, NCBI, and, eventually, EMSL.
2-5 years				
· KBase plans to continously expand its dedicated hardware · The KBase Platform will continue to expand and add functinality including support for new data types such as bio-imaging data.	· KBase will add support for more analysis tools, data types, and workflows.	· Data sizes of similar scale up to 10 TB · Uploads growing to 100TB per month by 2020	· In this time-frame KBase may adopt 100Gb for internal connectivity.	· KBase sites connected at 100Gb or greater. Continue to leveage connectivity at ANL and LBNL/NERSC. · Continuous replication of new data between sites. · Same sources plus potentially new sources from DOE Light-Sources or other facilities.
5+ years				
· Current plans extend only to 5 years, but hardware and software would continue to expand.	· Current plans extend only to 5 years, but KBase would continue to add new methods and support for new data types.	· New instruments will likely generate larger data set sizes (i.e. 10-100 TB)	· In this time-frame KBase may adopt >100Gb for internal connectivity.	· KBase sites connected at 400Gb or greater.

Case Study 6

Plant Microbial Interfaces and the Bioenergy Science Center

6.1 Background

Plant Microbial Interfaces (PMI) and the Bioenergy Science Center (BESC) generate a range of omics datatypes including genomics, transcriptomics, proteomics, glycomics and metabolomics from both plants and microbes. In addition, there is an increasing amount of meta-omics data being created and planned. We are using this data to build large, integrated descriptive network models encompassing evolutionary perspectives (co-evolution networks), genome-wide association study (GWAS) networks, GWAS-profile networks, phenotype correlation networks and co-expression networks of transcripts, proteins and metabolites. These integrated descriptive models are proving to be very useful for both hypothesis generation as well as contextual interpretation of experimental results. They are also giving us new insights into evolutionary mechanisms as well as genome structure and organization.

The (meta)genomics and (meta)transcriptomics data is usually generated at the JGI. This year, we have transferred over 1000 Populus genomes, hundreds of bacterial genomes and roughly 600 transcriptomes from the JGI. All told we have transferred roughly 125 TB over ESnet in the past 8 months. Later this calendar year, we will transfer 1000 Populus transcriptomes from the JGI as soon as they have been generated. In the coming year we anticipate the need to transfer several hundred more plant genomes as well as hundreds of plant transcriptomes. In addition, we are anticipating that we will need to transfer dozens to hundreds of metagenomes. Pyrolysis molecular-beam mass spectrometry and sugar release data is transferred from National Renewable Energy Laboratory (NREL) and glycomics data is transferred from the Center for Complex Carbohydrates Research (CCRC) at the University of Georgia. Furthermore, we download all publicly available microbial genome data from the NCBI and other public repositories, and transcriptome data sets from the NCBI Gene Expression Omnibus (GEO) or the sequence read archive (SRA). We are also receiving some data directly from the Noble Foundation.

6.2 Network and Data Architecture

All of the data has been transferred via Globus to the ORNL CADES DTNs. Once the data has been transferred to the CADES DTNs it is either transferred to a preprocessing cluster within the CADES DMZ and/or to the DTNs of the OLCF for use on Eos or Titan.

6.3 Collaborators

The primary collaborators and facilities that we interact with are the JGI, NREL, CCRC, the Noble Foundation, NCBI, CADES and OLCF.

6.4 Instruments and Facilities

- Present
  - CADES
    - DTNs
    - Preprocessing cluster
      - 1 petabyte of storage
      - 300 compute cores
    - Cray Urika XA (Map/Reduce Hadoop)
    - Cray Urika GD (Graph Discovery Appliance)
    - SGI 3 Tbyte Memory SMP platform
  - OLCF
    - DTNs
    - Eos
      - 20,000 compute cores
    - Titan
      - 300,000 compute cores
      - 18,000 GPUs
- Next 2-5 years
  - CADES
    - DTNs
    - Preprocessing cluster
      - Petabytes of storage
      - 5000 compute cores
      - Small GPU footprint for development
    - Cray Urika XA (Map/Reduce Hadoop)
    - Cray Urika GD (Graph Discovery Appliance)
    - SGI 12 Tbyte Memory SMP platform
  - OLCF
    - DTNs
    - Eos
      - 20,000 compute cores
    - Summit

        - 5 times the computational performance of Titan
        - Compute cores
        - GPUs
- Beyond 5 years
  - CADES
    - Continued expansion of storage, preprocessing, SMP, and data analytics architectures
  - OLCF
    - Exascale computing infrastructure

6.5 Process of Science

Presently, we are using the computational ecosystem available across CADES and OLCF in order to do very large-scale genome and omic data mining to detect co-evolutionary patterns in complex biological systems. These integrated co-evolutionary patterns are modeled as networks and then used, in combination with experimental data sets, in order to generate hypotheses (that can be tested experimentally) and to gain insights into the systems under study. Currently, our focus is on Populus and microbes relevant to BESC and PMI. This requires: 1) preprocessing platforms where raw data undergoes QC and reference mapping; 2) large memory platforms for individual and agglomerative genome and metagenome assembly and genome variant detection; 3) HPC clusters to look for statistical associations with phenotypes and genome variant correlations across 1084 Populus genomes and hundreds of microbial genomes to be modeled as very large networks; 4) map/reduce or large-memory SMP platforms to analyze the resulting large networks.

For the next 2–5 years, similar approaches will be taken using evolution and omics data to understand complex systems. However, the number of plant species analyzed in this manner will increase from one to dozens of species, the number of microbial genomes analyzed this way will expand to 100,000. Currently our descriptive modeling efforts scale in an n^2 fashion. Furthermore, we have developed more sophisticated hyper-network models that will scale in an n^m fashion where m may be 3, 4, 5 or higher. All of these activities will have dramatic impacts on the amount of data that we need to transfer to ORNL, the amount of storage and compute power to analyze the data, and the amount of results that we will need to store and disseminate.

Beyond 5 years, these needs will continue to increase as more data becomes available and we develop new, more sophisticated methods and modeling approaches.

6.6 Remote Science Activities

As described above, we import data generated at the JGI sequencing facilities, NREL, NCBI, Noble Foundation, and CCRC. The largest volume of data comes from the JGI and NCBI. Data are then utilized on the CADES and OLCF high performance computing environments at ORNL.

6.7 Software Infrastructure

Currently, we use a combinations of community code and in-house code for:

- Genome/Transcriptome assemblers (reference mapping and de-novo assembly),
- Genome variant detection,
- Genome wide association studies,
- Transcriptome, proteome, metabolome, glycome statistical and correlation analysis,

- Genome-based evolutionary pattern finding and model building,
- Gene family construction, and
- Network construction and analysis.

In the next 2–5 years, we expect this to be similar to the activities that we are doing at present, but on a much larger scale and with new sets of tools that are developed both by the community and in-house with the goal of starting to link descriptive models to process-based models.

The software infrastructure beyond the next five years will be similar to the current infrastructure but with an expanded focus on more sophisticated descriptive models. The goal will be to link these models to create and parameterize more sophisticated processes and predictive models, eventually linking evolutionary models to ecological- and climate-scale models.

6.8 Cloud Services

At present, the only cloud-like services used are in the CADES and OLCF environments at ORNL. ORNL may develop cloud architectures that we find useful in the future, and we will use them on an as needed basis.

6.9 Outstanding Issues

At present the data transfer process via Globus is functional but somewhat lacking in robustness. Large-scale data transfers have to be set up manually, which often take many manual interventions as the transfers crash and fail to restart. The size of data transfers are often many terabytes in a single transfer which can take quite some time—faster transfer speeds would of course be appreciated.

We are increasingly receiving data from collaborators who are generating data at commercial facilities. These data transfers are currently being done by Dropbox or similar mechanisms, and are becoming increasingly problematic as the data volumes are getting larger.

Table 6.1: The following table summarizes data needs and networking requirements for PMI.

Key Science Drivers			Anticipated Network Needs	
Instruments, Software, and Facilities	**Process of Science**	**Data Set Size**	**Local-Area Transfer Time**	**Wide-Area Transfer Time**
0-2 years				
· What are the current/new instruments and data sources? Genomics, transcriptomics, proteomics, metabolomics, glycomics · What is the current/new software used in scientific process? Combinations of community code and in-house code for: • Genome/Transcriptome assemblers (reference mapping and de-novo assembly) • Genome variant detection • Genome wide association studies • Transcriptome, proteome, metabolome, glycome statistical and correlation analysis • Genome-based evolutionary pattern finding and model building • Gene family construction • Network construction and analysis	· Highlights of currrent science process: We are using the computational ecosystem available across CADES and OLCF in order to do very large-scale genome mining in order to detect co-evolutionary patterns in complex biological systems. These co-evolutionary patterns are modeled as networks are then used, in combination with experimental datasets to generate hypotheses (that can be tested experimentally) and to gain insights into the systems under study. At present our focus is on Populus and microbes relevant to BESC and PMI. This requires: 1) preprocessing platforms where raw data undergoes QC and reference mapping; 2) large memory platforms for individual and agglomerative genome and metagenome assembly and genome variant detection; 3) HPC clusters to look for statistical associations with phenotypes and genome variant correlations across 1084 Populus genomes and hundreds of microbial genomes to be modeled as very large networks; 4) Map/Reduce or large memory SMP platforms to analyze the resulting large networks.	Datasets range from gigabytes to 50 TB and can be composed of thousands of files.	Large datasets can take 1 to 3 days to transfer internally. Smaller data sets can take from minutes to hours.	Datasets can take from hours to days to weeks to transfer. Most transfers are done in batch fashion. Datasets are transferred in an ad hoc fashion as datasets become available or as project need data. Nightly transfer of microbial genomes are done from various public repositories. Collaborating sites/destinationsare JGI, NREL, CCRC, the Noble Foundation, NCBI, CADES and OLCF.
2-5 years				
· What are the planned, new data sources/instruments? · What are the planned/expected software packages? Similar datasets but the number of species will increase dramatically Combinations of community code and in-house code for: • Genome/Transcriptome assemblers (reference mapping and de-novo assembly) • Genome variant detection • Genome wide association studies • Transcriptome, proteome, metabolome, glycome statistical and correlation analysis • Genome-based evolutionary pattern finding and model building • Gene family construction • Network construction and analysis	· What are the foreseeable changes to data flow, science process, etc? The amount of data will continue to increase and the complexity of the models will also increase. Thus the number of calculatons and the scale of the results will also continue to increase.	Datasets range from gigabytes to 50 TB and can be composed of thousands of files.	Large datasets can take 1 to 3 days to transfer interannlly. Smaller data sets can take from minutes to hours.	Datasets can take from hours to days to weeks to transfer. Most transfers are done in bacth fashion. Datasets are transferred in an ad hoc fashion as datasets become available or as project need data. Nightly transfer of microbial genomes are done from various public repositories. Collaborating sites/destinations are JGI, NREL, CCRC, the Noble Foundation, Dartmounth, Duke, Penn State Univ, Univ of Noth Texas, INRA (France) NCBI, CADES and OLCF.
5+ years				
· Describe any planned new data sources or software packages JGI and public data will continue to expand as well as those datasets produced directly by our collaborators. Software in the formof community code and in-house code will continue to evolve.	· What is the strategic direction for data flow, science process, etc.? The strategic direction is for our descriptive models to increase in scale and eventually become community resources. Our descriptive models will also begin to inform and parameterize predictive models at physiologics, ecological and climate scales.	Datasets range from gigabytes to 50 TB and can be composed of thousands of files.	Large datasets can take 1 to 3 days to transfer interannlly. Smaller data sets can take from minutes to hours.	Datasets can take from hours to days to weeks to transfer. Most transfers are done in batch fashion. Datasets are transferred in an ad hoc fashion as datasets become available or as project need data. Nightly transfer of microbial genomes are done from various public repositories. Collaborating sites/destinations are JGI, NREL, CCRC, the Noble Foundation, Dartmounth, Duke, Penn State Univ, Univ of Noth Texas, INRA (France) NCBI, CADES and OLCF.

Case Studies
Climate and Environmental Sciences Division

Case Study 7

Accelerated Climate Modeling for Energy

7.1 Background

Understanding climate change is one of the most important scientific problems of our time. The major theoretical tool for studying the climate is the coupled climate model or coupled Earth system model. These models solve the basic equations describing the fluid flow of the atmosphere, ocean, sea-ice, and land-ice; as well as important physical processes affecting heat, moisture and momentum in the atmosphere, land surface, ocean, sea-ice and land-ice. Due to the number of operations and grid points needed for modeling these solutions, climate modeling is a challenging problem in high-performance computing (HPC).

The Accelerated Climate Modeling for Energy (ACME) project is sponsored by the Earth System Modeling (ESM) program within DOE's BER. ACME is a collaboration among eight national laboratories and six partner institutions to develop the most complete, leading-edge climate and Earth system models and apply these models to challenging and demanding climate-change research imperatives. ACME is the only major national modeling project designed to address DOE mission needs and to efficiently utilize DOE leadership computing resources now and in the future.

While the project's capabilities will address the critical science questions, its modeling system and related capabilities will be flexible for the DOE research community to address mission-specific climate change applications as illuminated the report U.S. Energy Sector Vulnerabilities to Climate Change and Extreme Weather.

7.2 Network and Data Architecture

In a project with as many collaborators as ACME, one institution's network and data architecture may not provide the complete picture but many of the Office of Science laboratories have similar capabilities. ANL is connected to the outside world with 10 Gbps links to Internet2 and ESnet, as well as a 100Gbps to ESnet's 100Gbps testbed network. Several of ACME's ANL staff have joint appointments with the University of Chicago, Argonne Computation Institute (CI) and some have offices on the University of Chicago campus. Both ANL and the University of Chicago participate in the Illinois Wired/Wireless Infrastructure for Research and Education (I-WIRE) project, which links the sites by a dedicated network infrastructure funded by the state of Illinois. In addition, I-WIRE connects both institutions to StarLight, an international network facility, as well as to various research institutions in Illinois. As a comparison, LLNL connections include ESnet, the dynamic science network, and has access to the ALICE grid system, The Open Science Grid, and a wide variety of other resource services.

7.3 Collaborators

ACME makes use of ASCR Computing Facilities at ORNL (OLCF), ANL (ALCF), and LBNL (NERSC). Over 100 personnel are involved in the project with effort ranging from 35% to 100% of their time. The labs involved are: LLNL, ORNL, LBNL, PNNL, SNL, LANL, ANL and BNL. Personnel are not distributed uniformly across the labs. Smaller groups number around 5–10 personnel, while larger groups are 15–25 personnel. The project is run by a council with one member per lab and chaired by the Principal Investigator, David Bader (LLNL). The 6 non-lab partners include NCAR, Scripps Institute of Oceanography, University of Maryland, New York Polytechnical Institute, University of California-Irvine and the private company Kitware. Tasks range from developing the model, running simulations on ALCF, OLCF, and NERSC, analyzing results and developing related software such as Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT), and build and test infrastructure.

7.4 Instruments and Facilities

Currently, ACME makes use of Titan at OLCF, Mira at ALCF, and Edison at NERSC. ACME output data from experiments is stored on the disk and tape systems of each of those facilities.

While major simulations are done at the ALCF, OLCF and NERSC, ACME also makes use of internal lab clusters for code development, testing and analysis (roughly one per ACME lab partner.) Use of a lab's cluster is usually confined to the personnel at the lab. One exception is Blues at ANL that allows external collaborators with a relatively painless account request process and is used for model development and testing. The major reason for using local clusters and workstations for analysis is that special software used by climate researchers is sometimes not available or not configured correctly to run on LCF analysis systems such as Cooley.

In the next 2–5 years, we will continue to make use of pre-exascale systems at the ALCF, OLCF and NERSC and NERSC. Lab internal systems will be single-petaFLOP machines and can be used for additional simulations as well as model development, testing and analysis.

Beyond 5 years, ACME will plan to use exascale systems at the LCF's and multi-petaFLOP systems at individual labs.

7.5 Process of Science

Climate modeling is very much a data-centric process. An important point to make is that the climate model does not directly calculate the climate. Instead, it outputs years of simulated global weather—the time series analysis made from that output defines the climate. For example, an "average temperature" at a location (or over a region) is usually calculated from 30 years of data. The model must first calculate 30 years of global weather, which can take days to weeks depending on the resolution, before the data analysis can determine the climate.

Climate modeling is always starved for compute cycles, and so if there are N facilities ACME is eligible to use for its planned experiments, then model code development and experiment planning will try to use all N facilities. A single climate model experiment may have one part of the simulation done on one platform and a second part on another platform because of the allocation ACME acquired or local queue contention. A scientist will often need to compare one experiment with a previous one done at a different site or against observations stored at a third (non-DOE) site. The network demands come mostly from having to co-locate the data to analyze the experiment because the crucial analysis tools currently all assume the data is on locally mounted disks, so input/output (I/O) becomes the bottleneck. Scientists spend a lot of effort thinking about where data is physically located instead of focusing only on what variables from which experiments should be examined.

A climate model's output data in files where each file contains the information for hundreds of variables at a specific time or average over a period of time. But analysis usually focuses on only a handful of variables and typically more data resides on disk than is necessary. Transposing the model output to be one file per variable for multiple times could help remove the pressure on spinning disk.

Presently for the process of science, model simulations are performed at the OLCF, ALCF or NERSC. Some results are analyzed locally. Data is then transferred from two or more facilities to local cluster/workstation for further analysis.

In the next 2–5 years, we expect the process of science to evolve to do more of the analysis locally with data transfers minimized by using more intelligent analysis tools. Model simulations will still be performed at the OLCF, ALCF NERSC and one lab's local cluster.

Beyond 5 years, model analysis performed *in situ* with cloud-like analysis as a service and sharing of analysis through mechanisms similar to Galaxy or K-base.

7.6 Remote Science Activities

All ACME researchers make use of remote facilities: the ALCF, OLCF, NERSC and cloud-based software. Even researchers located at ORNL, ANL and LBNL will need to access data or compute engines at the other sites because need data at that location or their problem is better suited to that lab's architecture and platforms.

7.7 Software Infrastructure

Focusing only on the simulations and not the software used for model development, presently the Earth System Grid Federation (ESGF) is used for the formal publication of data results. Globus is used for transferring data between computing sites or from computing sites to local lab clusters/workstations (scp may also be used for the later). Data reduction including simple time averaging, spatial sub-sampling and variable extraction is performed locally by the netCDF Operators command line tools. Additional analysis and visualization is done with UV-CDAT, NCL, python or MatLab.

In the next 2–5 years, the tools set is not expected to change but there will hopefully be less need to transfer data as analysis tools become more standardized and installed at the the ALCF, OLCF, NERSC.

Beyond 5 years, the tools will need to run in cloud-like environments on exascale systems with significant data reduction/compression before outputting data to spinning disk. It will never be possible to do all analysis *in situ* because we learn by comparing results across experiments. Climate models allow the exploration of various "what if" scenarios allowed through different sets of parameters and input files, meaning that all experiments can not be planned in advance.

7.8 Cloud Services and Outstanding Issues

ACME depends crucially on cloud-based services for its day-to-day function. Most bandwidth requirements come from using Globus to transfer model output between ALCF, OLCF, NERSC, and local clusters or workstations. For wiki and task tracking, we are using cloud-based instances such as Atlassian tool's Confluence and JIRA hosted at atlassian.net. Page load time from this site directly affects productivity on the project. Our code repository is on github.com (github.com/ACME-Climate/). Github only hosts the code during the editing and development process on OLCF and ALCF or local clusters. While a distributed version control system like git is tolerant of network latencies or outages, many day-to-day git commands, as well as the initial "clone" of a repository to a local disk, must go over the network and interact with github.com. ACME is using web-based video/audio teleconferencing from gotomeeting.com.

We would like to use more cloud-based systems in the near future. Our continuous integration platform is a Jenkins instance hosted at Sandia National Laboratories (SNL). Because of high security in the Sandia network, not all information and functionality of Jenkins is available to ACME researchers outside of the NNSA labs and the communication required to run tests on the ALCF and OLCF systems require special security exemptions approved individually by Sandia's network staff. A cloud-based Jenkins service would alleviate some of these issues. Jenkins submits tests to run on the ALCF, OLCF, NERSC, local Sandia machines and Blues. Looking further ahead, using

a cloud platform for running tests would be useful but may prove difficult given the strict software stack and performance requirements (for nightly testing) of ACME.

Case Study 8

The Atmospheric Radiation Measurement Program

8.1 Background

The Atmospheric Radiation Measurement (ARM) Climate Research Facility operates field research sites around the world for global climate change research. Three primary locations, the Southern Great Plains Mega-site, North Slope of Alaska Mega-site plus aircraft and the portable ARM Mobile Facilities, are heavily instrumented to collect massive amounts of atmospheric measurements needed to create data files. Scientists use these data to study the effects and interactions of sunlight, clouds, and radiant energy, as well as interdisciplinary research involving hydrology, ecology, and weather forecasting. As part of this effort, ARM scientists and infrastructure staff provide value-added processing to the data files to create new data streams called value-added products. Software tools are then provided to help open and analyze these products, which are available for discovery and delivery from our archive.

The ARM facility is currently undergoing a reconfiguration that is designed to accelerate the application of ARM observations and data processing for the understanding of key atmospheric processes and the representation of these processes in global climate models. This enhanced impact on the research community will be achieved by:

- Enhancing ARM observations and measurement strategies to enable the routine operation of high-resolution models and to optimize the use of ARM data for the evaluation of these models;
- Undertaking the routine operation of high-resolution models at ARM sites; and
- Developing data products and analysis tools that enable the evaluation of models using ARM data.

The reconfiguration of the ARM facility does not alter the ARM mission; however, it does involve a number of changes to align all aspects of the facility with the new strategy through an integrated vision where ARM observations will be used to advance climate models. This vision also includes developing and improving interactions between the ARM facility and the research community to make full use of this next-generation strategy.

8.1.1 Mission and Vision

The ARM Climate Research Facility, a DOE scientific user facility, provides the climate research community with strategically located *in situ* and remote sensing observatories designed to improve the understanding and representation, in climate and earth system models, of clouds and aerosols as well as their interactions and coupling with the Earth's surface. In addition, ARM's mission is to provide a detailed and accurate description of the Earth's atmosphere in diverse climate regimes to resolve the uncertainties in climate and the Earth system models toward the development of sustainable solutions for the Nation's energy and environmental challenges.

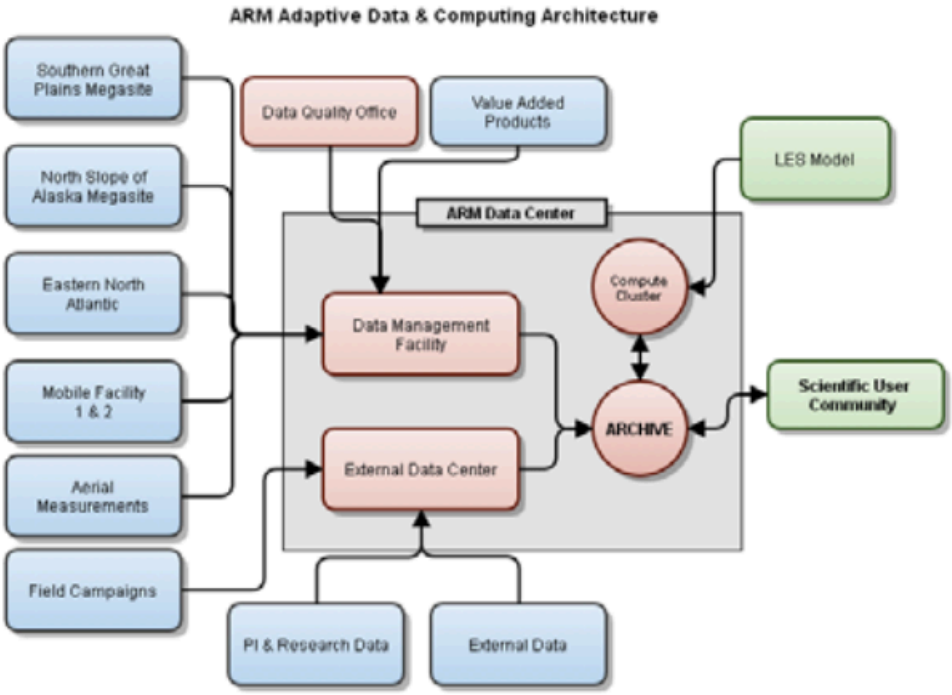


Figure 8.1: ARM adaptive data and computing architecture.

8.2 Use Cases

The uses cases presented below will describe the data flow from the Mega-sites to the ARM data center, which will provide flexible data delivery options for big data orders such as the ARM Large Eddy Simulation (LES) model data and ARM radar data.

8.2.1 Use Case 1: ARM Data Flow

The ARM data are generated from instruments and loaded directly onto standalone computers, virtual machines, and data loggers. Data is collected to a central collection system (collector) at each measurement facility. Most sites have a single location and send data back to a single collector. Other sites have multiple locations and send data back to the central collector from either the instrument itself or a remotely deployed collector. The data are sent via microwave radios or over virtual private network (VPN) tunnels on digital subscriber line (DSL) or cellular connections. Once data has been collected to the primary central collector, it is sent to the Data Management Facility (DMF) for processing. All sites communicate through the VPN Server Network (VSN) at ANL. All traffic to and from the sites goes through ANL. All traffic between ANL and the measurement facilities is encrypted on a VPN tunnel. As data flows from from ANL to the DMF, which is part of the ARM Data Center at ORNL, it is not encrypted, but it is sent over ESnet. The data transfers themselves are not encrypted. A new version of data transfer software is being developed that will encrypt data in transit.

Data Collection

At all measurement facilities, data is collected via FTP on a local network. No FTP traffic leaves a site when collecting data. FTP traffic for data collections is never on a public network.

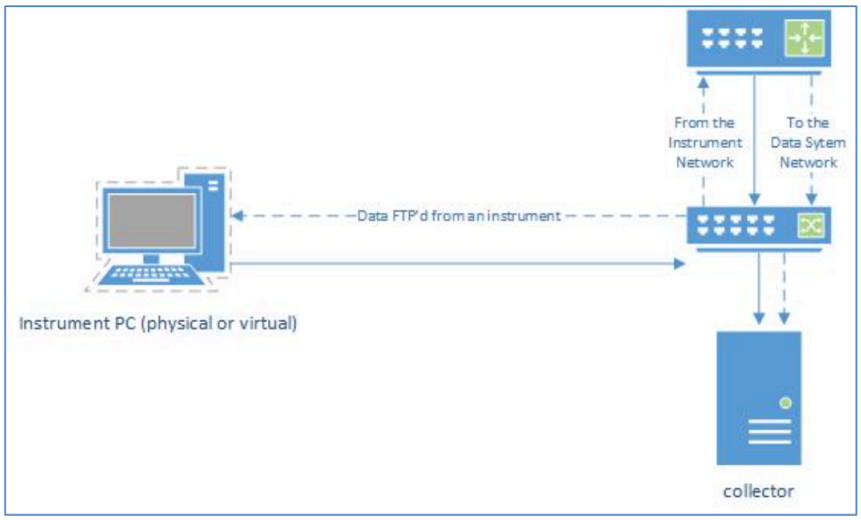


Figure 8.2

Data Flow

Data is transferred using an in-house developed system known as site transfer. Site transfer allows for multiple FTP sessions to run concurrently. It also provides for the creation of file manifests and facilitates cleanup once the data is successfully transferred to the DMF. This is especially useful on highly latent network links, such as satellite connections. It is anticipated that ARM will have an updated site transfer software by CY16 that utilizes HTTPS instead of FTP. For sites that do not have the network capacity to send all of the data over the network, those data are written to disk and shipped to the DMF.

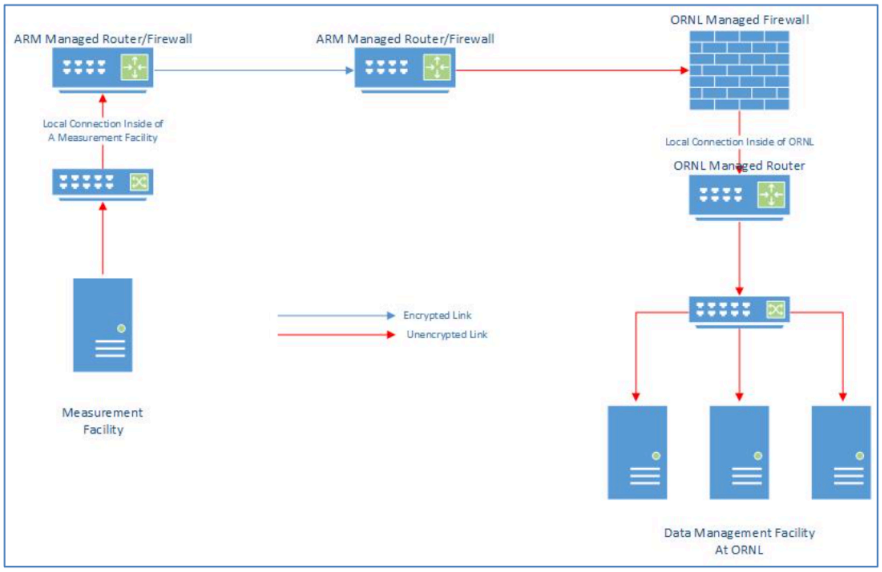


Figure 8.3

Data Volumes

It is anticipated that ARM's data volumes will increase tremendously in FY16 as a result of new radar systems and scanning strategies. As it stands, most measurement facilities do not possess enough bandwidth to send all of the data over the link. With the anticipated increase in data volume this coming year, a lot more data will be sent

on disk. Note that there are plans for fiber to be installed on the North Slope of Alaska by spring of 2017[1]. ARM will be able to utilize this infrastructure and, hopefully, be able to send all data from the Alaska sites over the network. Currently, there is not a lot of additional available bandwidth, if any, at the sites. The links are run near capacity to facilitate maximum data flow and still allow for remote access for maintenance of the instruments and data system infrastructure.

ARM Data Volumes and Bandwidth per Site				
Site	**Current Data Volume (MB/mo)**	**Anticipated Data Volume (MB/mo)**	**Current Bandwidth**	**Anticipated Bandwidth in FY16**
SGP (Billings, OK)	5,983,400	36,564,468	100 Mbps	1 Gbps
NSA (Barrpw, AK)	3,500,000	11,981,771	5 Mbps satellite	same
ENA (Graciosa Island, Azores)	12,040,000	24,528,823	100 Mbps	same
AMF1 (Manaus, Brazil)	1,800,000	11,420,404	1 Mbps satellite	same
AMF2 (Antarctica)	11,221,000	21,239,767	2 Mbps satellite	same
AMF3 (Oliktok Point, AK)	12,720,000	24,716,176	1.5 Mbps Satellite	same
VSN (ANL, Lemont, IL)	47,264,400	130,451,409	1 Gbps	10 Gbps

8.2.2 Use Case 2: Support Large ARM Data Requests

ARM collects and archives different types of data: regular instrument data streams, processed data, Value Added Products (VAPs), special collections from Principal Investigators (PIs), field campaigns, and data from external sources. Information about ARM instruments, measurements and data products is compiled and presented via the ARM web site[2] ARM data can be discovered using the ARM Archive Data Discovery tool, which is available from the ARM Data Archive page.[3] As part of the data discovery, users can refine the search to ARM data of interest by using faceted keywords grouped in instrument and measurement categories. Additional information such as data plots, data citation (digital object identifiers, DOIs), time grid, and Data Quality Report (DQR) also aids the data selection process. Figure 8.4 shows the workflow of a typical search of ARM data using the arm data discovery tool.

For typical user requests, ARM data retrieval process extracts the user requested data from the Archive, packages the data, places the results on an ARM FTP server, and then notifies the user with an email containing the FTP link to the requested data. The transactions within the data request processing occurs across systems within a dedicated ARM 10 Gigabit network which is a "research enclave" at ORNL with direct access to the primary Oak Ridge firewall. The ARM network uses the ORNL firewall and the security management plan of ORNL. As needed, the data transactions for user data requests also communicates with the Oak Ridge installation of HPSS which is located in the National Center for Computational Sciences (NCCS) network domain. The data transfers between the ARM network and HPSS may include multiple concurrent file transfers to one or more systems. Each HPSS file transfer can use up to 16 concurrent threads to move data blocks from HPSS to ARM servers. Therefore, numerous, concurrent network connections may be in use to move large volumes of data between ARM systems and HPSS.

Currently ARM utilizes Globus/GridFTP for moving large data (multiple terabytes bytes) in and out of ARM radar data processing clusters; ARM Data Center also provides this option for some special multi-terabyte data requests.

As explained in the use case 1, any data requests such as large radar data, ARM will provide Globus/GridFTP service to transfer the data to the user. ARM will explore adding Globus Online as an option for downloading any large data from the ARM Data Center.

[1] `http://www.adn.com/article/20150510/arctic-spanning-fiber-optic-project-moves-ahead-alaska`
[2] www.arm.gov
[3] www.archive.arm.gov

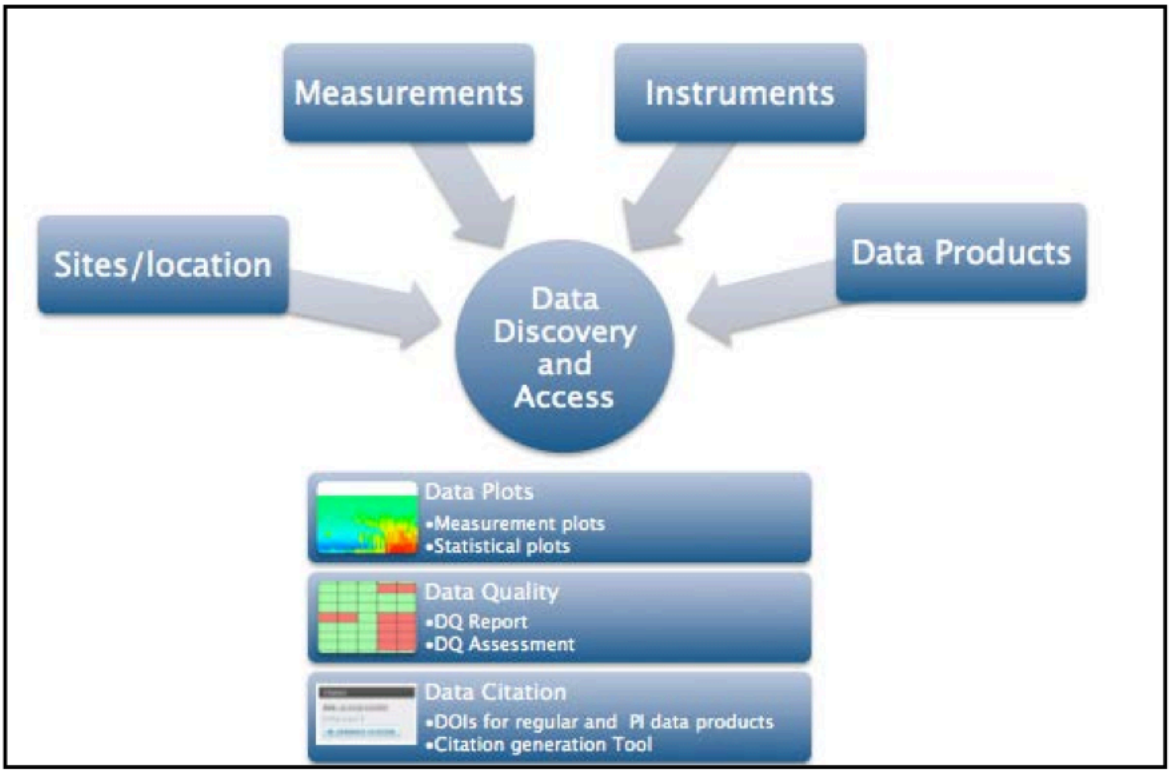


Figure 8.4

The ARM LES model output will also be archived using the existing ORNL HPSS resource starting later in FY 2016. We anticipate that users will be requesting multiple terabytes of the simulation data. In some instances, the LES output will be summarized and packaged within ORNL resources before distribution to users. In other instances, data analysis and visualization tools will be available to the user community to evaluate the LES output. When large volumes of LES output need to be transferred to the user's own computing resources, we hope these data will be made accessible using Globus/GridFTP. We are anticipating many of these data transfers to user will utilize the ESnet infrastructure.

8.2.3 Use case 3: Write optimization for Remote users

Previous experience with large data transfers to or from user destinations have indicated that optimization of routing between ESnet and other networks may need some optimization. A detailed evaluation was done on data transfers between Harvard and ORNL. This evaluation indicated that most of the network transfer was routed over Internet II until it got "lost" on a slow link somewhere in Georgia and then transferred to ESnet. An alternate route could have been a short connection between Harvard and MIT which has direct access to ESnet. In another case, the route from the NASA Ames (Moffet Field airport) in California to ORNL was evaluated. This route started with a short-term use of a private ISP. After observing very slow transfer rates to ORNL, the route was evaluated and the analysis revealed that a good bit of the landscape was covered by a variety private (and somewhat slow) ISP until the connection was made to ESnet within a few hundred miles of ORNL. Could connections for very large data transfers be optimized to use ESnet to cover most of the route?

Case Study 9

Calibrated and Systematic Characterization, Attribution, and Detection of Extremes Scientific Focus Area

9.1 Background

The Calibrated and Systematic Characterization, Attribution, and Detection of Extremes (CASCADE) Scientific Focus Area (SFA) is a three-year multi-disciplinary project at the Berkeley Lab to investigate changes in extreme weather as the overall climate warms. This large SFA brings together physical climate scientists, statisticians, and computer scientists to devise analyses of large model and observational data sets. Case studies for this SFA are divided into two broad categories. The first involves analyses of model output produced by the international community. Typically this is from the Coupled Model Intercomparison Project (CMIP5). The second involves analyses from model simulations produced by the CASCADE team members themselves. These data sets are more uniform than CMIP5, but may contain many more individual simulations or be of significantly higher resolution.

9.2 Network and Data Architecture

Almost all of our calculations are performed at NERSC. Either this is because we are using the HPC compute resources or we need the disk space. The project has purchased 100 TB of project storage at NERSC with plans to purchase another 400 TB.

9.3 Collaborators

CASCADE has one collaborator at most of these following institutions:

- NERSC (Berkeley, California)
- Stonybrook University (Long Island, New York)
- NCAR (Boulder, Colorado)
- University of Capetown, South Africa
- University of New South Wales (Sydney, Australia)
- University of California at Berkeley

- Meteorological Research Institute (Tokyo, Japan)
- Seoul, South Korea
- The MetOffice (Exeter, UK)
- University of Reading (UK)
- University of California, Davis (Davis, California)
- Colorado State University (Fort Collins, Colorado)
- And many others

9.4 Instruments and Facilities

CASCADE uses NERSC for all computing

9.5 Process of Science

To describe CASCADE's process of science, we will cover two case studies divided by the different requirements for the CMIP5 project, and other CASCADE research.

9.5.1 Case Study 1: Large-scale CMIP5 Data Transfer

The scientific goal for CASCADE's CMIP5 work was to track simulated storms of a particular type, known as Extra-tropical cyclones (ETC) in simulations of the past and future climate. Understanding the potential for future changes in the frequency and intensity of this class of storms are of critical national and international importance. Identifying and tracking storms requires the usage of sub-daily model output, which typically has a much higher frequency of data output than most climate model studies, resulting in much larger input data sets for analyses. Part of the CMIP5 protocols specified retention of selected variables at a six-hour frequency. Although the number of individual model realizations that had such data saved is less than ideal, the output from twenty-five models was available for the historical period. Fewer models were available for the future and other requisite scenarios, but enough was available to make interesting projections of the future and perform detection and attribution analyses. For the tracking part of the study, we used the highly parallel Toolkit for Extreme Climate Analyses (TECA) operating on mean sea level pressure (psl) and 850mb wind speeds (obtained from the component wind vectors, u_a and v_a). Data was retrieved via the ESGF via standard techniques over the course of three months with the help of ESnet. Figure 9.1 [4] summarizes data transfer performance from the various worldwide data centers to the data transfer nodes at NERSC. Performance varies from unacceptable to merely painfully slow. After the tedious collection of the data sets, it was transferred en masse from NERSC to ANL via Globus for a massive 700K processor run of TECA on the Mira IBM Blue Gene machine [6]. That transfer of data was made in two pieces of approximately 3e13 Bytes in about 2 days each.

Our CMIP5 case study was likely the largest data transfer of climate model output for a single study. However, other larger data transfers have been made at the large data centers for backup archival purposes.

In the next 2–5 years, we expect that CMIP6 project will start to come online. Initial contributions to the database will be of only moderate increases in resolution. We might expect them to be at 100km resolution, which is close to the highest resolutions in CMIP5. By the end of this period, the high-resolution MIP part of the CMIP6 database will be populated. These are far more interesting simulations for storm tracking studies, as tropical cyclones are permitted at these resolutions. We would expect that Case Study 2 analyses would be possible in a multi-model sense (see Section 9.5.2).

Beyond 5 years, all project projections are purely speculative. At the least, we would hope that ensemble sizes are increased to at least 10, if not higher. This would be an increase from the 1–3 PB that CMIP5 data currently has for high-frequency experiments.

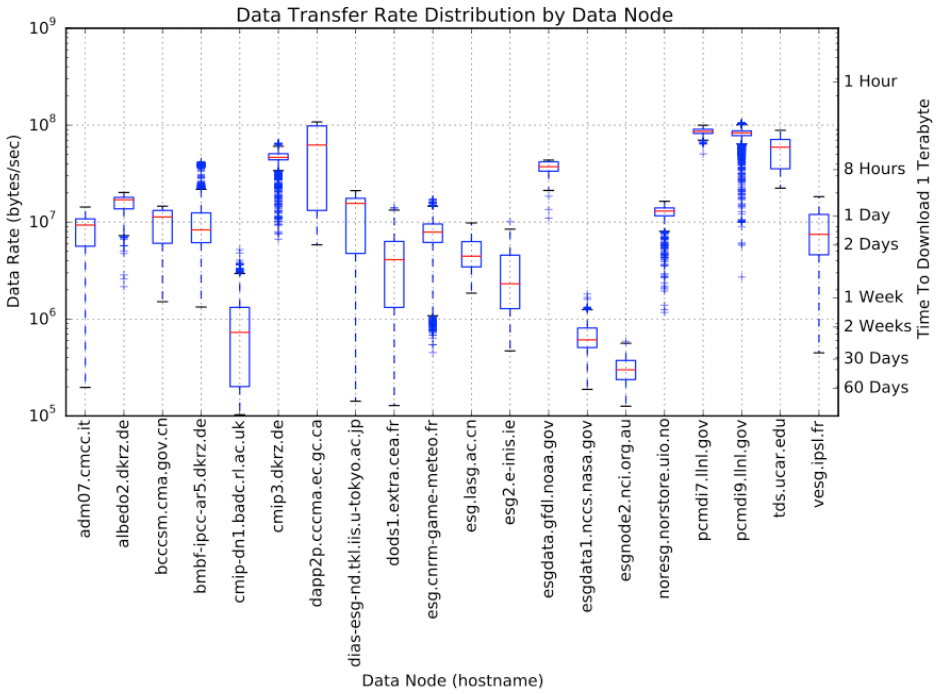


Figure 9.1: Data transfer rate performance statistics for ESGF data nodes plotted on a logarithmic scale.

9.5.2 Case Study 2: Locally Produced Data

The TECA code, being written in C++ with the Message Passing Interface (MPI), has proven to be portable to all of the NERSC and LCF platforms. Hence, data transfers are far less onerous in this case than in Section 9.5.1. The largest CASCADE analyses currently are the tracking of hurricanes in output from the 25km version of the Community Atmospheric Model (CAM). This involves eight variables drawn from the three hourly instantaneous model output. A single year of the requisite TECA input is about 500GB with typically analyses being of simulations of one to two decades. Usually, jobs on the NERSC computers are broken up to about 20K processors to reduce queue wait time, while still maintaining reasonable throughput. CASCADE has purchased 100 TB of disk space on the project disk system and will likely purchase another 400 TB. However, TECA is about an order of magnitude faster on the LUSTRE scratch systems than on the non-parallel project file system. Hence bottlenecks from data transfers result in one of two ways. 1) transfer from tape or 2) transfer from project to scratch. Typically we find that 2 is slower than 1 but we have not performed detailed measurements.

We note other issues associated with the locally produced CASCADE data sets, which are currently problematic. The project has many international collaborators at both national laboratories and universities. While we have published much of the data on the ESGF was down for about seven months. We have provided a back door via an HTML link, which is much more popular with our collaborators (see `http://portal.nersc.gov/c20c/`) than ESGF. However, we do not feel that this is a viable option for transfers of high-frequency output, especially at high resolution. In lieu of some type of Globus-enabled option, we typically grant our collaborators limited access to our computational resources at NERSC.

Finally, we want to mention some of our future ambitions within CASCADE involving much larger data sets. Supervised machine learning techniques, which replicates the hurricane tracking of the TECA map-reduce algorithm at the cost of requiring more data, have already been demonstrated by Prabhat's group at NERSC. This work opens up the possibility of tracking storms and features that physically based tracking algorithms are not yet developed for or cannot be developed for. Further afield, unsupervised machine learning techniques offer the promise of discovering unknown weather features. This is likely to involve processing variables of three spatial dimension as opposed to the two-dimensional fields required by TECA. As model fidelity increases in both the horizontal and vertical dimensions, enabled by the march to exascale, data requirements will increase by several orders of magnitude in the next decade.

Overall, we expect that two factors will lead to increases in data set sizes. First, there will be longer and more simulations at 25km—century scale is likely. Second, a doubling in horizontal resolution is also likely. This will increase data sets by a factor of 4. Also, the machine learning analyses described above will likely become routine, leading to larger analyses than are currently done.

Long term, exascale computing will lead to cloud system resolving global models. This development will cause data sets to be at least two orders of magnitude larger, although three orders of magnitude is more likely when considering the vertical dimension. The amount of information of such model outputs is considerable, the challenge will be in extracting it.

9.6 Outstanding Issues

- The size of a productive working set on a filesystem varies a lot, but in many cases it is larger than our filesystem quota. Currently Wehner has 60 TB on NERSC's Hopper scratch for hurricane tracking analyses. Usually we stage this sort of simulation in 10 TB chunks. The ability to increase our working set size on the filesystem (without staging to/from tape) would make us more productive.
- The CASCADE group is in the business of producing derived data sets, so productive and timely access to the raw data is critical and often challenging. We are interested in more effectively supplying such derived products. However, usually derived data sets are smaller than the raw data.
- Repeating the work noted in Section 9.5.1 at the scale described is not something we plan again soon, due to the human resource requirements. Easier access to raw data which reduced the human effort required to assemble large-scale data sets would result in greater scientific use of the large CMIP5 data sets.

Case Study 10

Environmental Molecular Sciences Laboratory

10.1 Background

The Environmental Molecular Sciences Laboratory (EMSL) is a DOE BER National Scientific User Facility with a vision to pioneer discoveries and effectively mobilize the scientific community to provide molecular science foundations for BER research priorities and our Nation's critical biological, environmental, and energy challenges. These scientific challenges include:

- Gaining systems-level knowledge of genomes to functional translations in cells to underpin a predictive understanding or redesign of metabolic processes for sustainable bioenergy and environmental purposes;
- Understanding fundamental molecular-scale properties of natural and anthropogenic inputs to improve predictions of key atmospheric and environmental processes; and
- Designing and characterizing new catalytic materials for improved energy storage and conversion (including biomass) processes to make clean, affordable, and abundant energy a reality.

EMSL users and scientists are conducting groundbreaking research to address scientific challenges. By connecting the scientific user community through our suites of scientific expertise, state-of-the-art equipment, and mission ready facilities, EMSL provides a creative environment that supports problem-solving beyond what is possible in a typical university, industrial, or even single national laboratory setting.

10.2 Network and Data Architecture

EMSL's internal network core is built around a fiber infrastructure that provides connectivity to a standard 8-port network switch in each office and laboratory space. The capability exists to directly attach gigabit and ten-gigabit instrumentation and computational resources directly to the EMSL core network. Isolated instrumentation networks are created using the building fiber to interconnect lab spaces throughout the EMSL building, and in some cases extending into other buildings. The EMSL core network has multiple connections into the PNNL core network through redundant fiber paths.

PNNL currently has data transfer nodes (dtn.pnl.gov) attached to its Secure Collaboration Zone (SCZ) perimeter network. The dtn.pnl.gov system is connected via 10 Gigabit Ethernet to the Internet, and Infiniband attached to a 4 PB Lustre storage cloud—it supports 1Gbps data transfers. The storage cloud has multiple internal mount points, and is available to the Olympus supercomputer via the same Infiniband interconnects. EMSL also has its Aurora archive 10 Gigabit Ethernet connection to the SCZ network providing up to 1Gbps data transfer capability to other labs. The SCZ has perfSONAR/NDT testing points attached at 10 Gigabit Ethernet (ndt-scz.pnl.gov and speedtest.pnl.gov). The SCZ utilizes a host-based firewall model where Port Scan Attack Detector (PSAD) is

used in conjunction with iptables to detect and block attackers with little performance degradation on individual hosts.

PNNL has not started to engineer any DTNs on the 100 Gbps connection, and more importantly, we have very few data transfers that utilize more than 100 Mbps streams. PNNL does not consistently exercise the existing 10 Gbps capability.

The Aurora storage archive at EMSL is increasingly being used as a central store for EMSL data. It contains 7.3 PB of user data, up from 4.5 PB three years ago. In the last year, EMSL has produced about 22 TB of data weekly.

10.3 Local Instruments and Facilities

EMSL consists of multiple experimental capabilities. Each EMSL capability operates a set of scientific instruments on behalf of EMSL users. The capabilities include:

- Cell Isolation and Systems Analysis (CISA)
- Deposition and Microfabrication (DM)
- Mass Spectrometry
- Microscopy
- Molecular Science Computing
- Nuclear Magnetic Resonance (NMR) and Electron Paramagnetic Resonance
- Spectroscopy and Diffraction
- Subsurface Flow and Transport

We will focus on the capabilities with significant networking needs below.

The Cell Isolation and Systems Analysis (CISA) capability provides technologies and expertise to study individual cells and cell communities or tissues at the molecular level. Live cells or organelles can be isolated from complex populations, including environmental microbial communities or plant tissues for analyses spanning quantitative live cell fluorescence imaging with single molecule sensitivity, super resolution fluorescence and atomic force microscopy, and transcriptomic analyses using next-generation sequencing technologies. Together with proteomics, metabolomics, and electron and ion microscopy, these capabilities provide the foundation for attaining a molecular-level understanding of individual cells and cell community dynamics; and function to support biofuel research, understand the role of biological systems in carbon cycling, and enable research in biodefense and other national needs.

Mass Spectrometry enables high-throughput, high-resolution analysis of complex mixtures. These resources are applied to a broad range of scientific problems from proteomics studies with applications to environmental microbial and plant communities and human health to aerosol particle characterization, as well as fundamental studies of ion-surface collisions and preparatory mass spectrometry using ion soft-landing. Instruments include Fourier transforms (FT) mass spectrometers, including Tesla Fourier-Transform Ion Cyclotron Resonances (FTICRs), Orbitraps and LTQ-Orbitraps, linear ion traps, triple-quadrupole spectrometers, ion mobility spectrometry (IMS)—time-of-flight (TOF) spectrometers, HPLCs, a field-deployable second-generation single-particle laser-ablation time-of-flight mass spectrometer, and an ion soft-landing deposition instrument.

Microscopy has a wide variety of sophisticated microscopy instruments, including electron microscopes, optical microscopes, ion microscopes, scanning probe microscopes, and computer-controlled microscopes for automated particle analysis. These tools are used to image a range of sample types with nanoscale and even atomic resolution with applications to surface, environmental, biogeochemical, atmospheric, and biological science. Each of the state-of-the-art instruments and customized capabilities is equipped with features for specific applications. Instruments include electron microscopes with tomography, cryo, scanning, spectroscopy, and high-resolution capabilities, an NMR microscope, a dual Raman confocal microscope, optical microscopes, single-molecule fluorescence tools, spectroscopy tools with visible, near-, mid-, and far-infrared capabilities, atomic force microscopy, and scanning probe microscopes.

Spectroscopy and Diffraction has a suite of spectroscopy and diffraction instruments in EMSL, which allow users to study solid-, liquid-, and gas-phase sample structure and composition with remarkable resolution. Ideal for integrated studies, spectrometers and diffractometers are easily coupled with EMSL's computational and modeling capabilities, allowing EMSL users to apply a multifaceted research approach for experimental data interpretation and to gain a fundamental understanding of scientific problems. Instruments include electron-, Mössbauer-, and secondary ion mass-spectrometers, as well as atom-probe tomography. Optical spectroscopy tools include confocal-Raman, Fourier transform infrared, time-resolved fluorescence, and second harmonic generation capabilities, and multiple X-ray diffraction instruments.

Molecular Science Computing provides an integrated production computing environment supporting a wide range of computational activities in environmental molecular research, archive storage, scientific expertise, and the NWChem computational chemistry software suite. Systems include a 1440 node supercomputer (with a peak of 2.54 PetaFLOPS, consisting of dual 8-core Intel "Sandy Bridge" processors, 184 TB of memory, a 3 PB global file system), and a 15.7 PB hierarchical archive storage system.

EMSL expects to procure the new HPCS-5 HPC system(s) for delivery in mid calendar year 2017. The system is expected to increase EMSL's capability to support multiscale modeling of earth, environmental, and biological systems. In addition to scientific computing the supercomputer will increasingly be used to provide real-time analysis of the experimental data streaming off the scientific instruments. In addition, in the next five years the archive storage system will likely be upgraded to exceed 20 PB of data storage.

The common denominator in the next 2–5 years is increased resolution and increased data rates coming off of the instruments. New data management policies and processes will improve EMSL's ability to make unique data available to the scientific community. This should make EMSL a supplier of petabytes of data to ESnet's users. It is anticipated that the archived proteomics data will be accessed with increasing frequency as its use in gene annotation becomes more common. Thus the volume of accessed data should increase by a factor of at least 2–5 in the next several years.

10.3.1 Process of Science

EMSL's capabilities are available to researchers through a peer-reviewed proposal process, at no cost, if research results are published in the open literature. Users access the facility to use one or more capabilities, and work with EMSL's expert staff to gain insight and knowledge into their scientific problem. A large majority of instruments at EMSL offer remote access, although some require hands-on work and assistance from scientific experts from EMSL.

Data is generated by most instruments, and usually processed either automatically or manually before delivery to a user. It is typically sent to the user's home institution through email or FTP, or by media such as a CD and thumb drives when necessary due to bandwidth limitations. In extreme cases, hard drives are shipped to the user's home institution, owing to the quantity of data, and the uncertainty of reasonable bandwidth between EMSL and the home institution.

New generations of ultrafast and high-resolution electron microscopes are driving new data growth. The first of the new generation, the Dynamic Transmission Electron Microscope (DTEM) will produce approximately a terabyte of raw data per day. Post processing will need to be applied to the raw data to capture images of interest, specifically those of chemical reactions in progress. The data of interest will be some fraction (yet to be determined) of the raw data.

The CISA capability operates two SOLiD Sequencers, generating 48–128 TB per year. All of the sequence data gets shipped offsite on hard drives, due to the 1–2 TB of each data set.

EMSL is just beginning to operate a brand new 21 FTICR mass spectrometer known as the High Resolution Mass Accuracy Capability (HRMAC). It enables "top-down" proteomics of large complex proteins and protein complexes as well as identification of complex carbon molecules in soils and the atmosphere. HRMAC is expected to generate approximately 1 TB of data per day initially, growing to 5 TB per day in five years.

10.4 Remote Instruments and Facilities

EMSL's users have remote access the Cascade supercomputer, the Aurora data storage archive, and use remote tools to access many instruments remotely, saving time and travel costs. Remote instrument access does not drive a need for wide-area network bandwidth. Reliability of these connections has not been reported as a problem.

EMSL has multiple and redundant connections to ESNET (and the internet) via 10 Gigabit links through PNNL to Seattle (Primary) and Boise (Failover). The network was engineered with reliability and performance in mind: should one of the 10 Gigabit links suffer an incident that disrupts primary service, traffic is automatically failed-over to the redundant link.

In the 2–5 year timeframe the next-generation supercomputing capability will become available to the EMSL user community. Increased size and number of raw data sets will make it more difficult for users to move the data to their home institutions for analysis, which drives the need for increased access to remote analyses and visualization capabilities at EMSL. If the aforementioned DTEM capability is seen as successful by BER, it may lead to an Ultrafast Microscopy Capability (UMC) being developed in 5–7 years. The UMC capability is expected to generate 2–3 orders of magnitude more raw data (around 100–1000 TB/day) than DTEM. How much of that data may be kept and processed will depend upon bottlenecks in computing and networking technology at that time.

Multiscale modeling of earth and environmental systems will likely drive transfer of instrument data from experimental capabilities like ARM, and from other HPC sites. Use cases and likely requirements in terms of wide-area network bandwidth have not yet been established.

10.4.1 Process of Science

Remote data sources are of increasing importance as EMSL develops more of a focus on team-based science and integration of data from multiple sources. We expect needs for remote data to be driven by science such as measurement and integration of multi-omic data and multiscale modeling of terrestrial and atmospheric systems, where data produced by other institutions will be transported to EMSL for integration, analysis, and visualization. Likely remote endpoints in this scenario include the ARM facility, JGI, the Joint BioEnergy Institute (JBEI), and other BER-funded Bioenergy Research Centers, and KBASE. EMSL and JGI have already established new interfaces for automated downloads from JGI to EMSL.

The MyEMSL data management system will soon become a focal point for data transfer and collaboration activities by EMSL users and their collaborators. MyEMSL stores data acquired from EMSL's experimental and computational instruments and the output of analysis software, together with relevant metadata. MyEMSL will allow users a simple way to find, retrieve, visualize, and analyze their stored data. The data management system will provide a simple and consistent interface for the EMSL staff operating the instruments and for the users accessing and sharing their data. It will also provide EMSL's capability to make research data public. Furthermore it is expected to be a catalyst for the creation of new data and discoveries derived from existing data.

EMSL expects that the combination of MyEMSL, higher quality image data, and systems biology to drive more hosting of data sets by EMSL for access by the external scientific community. There will be increased interest in remote collaboration in which data is posted for shared access, and collaborators can share information about the data in real time.

10.5 EMSL Beyond 5 Years

EMSL plans to increase its scientific impact by focusing on critical science challenges in biology, the environment, and energy. These science themes help define and direct development of key capabilities and collections of user projects that can have significant impacts on important areas of environmental molecular science that are critical to DOE, BER, and the Nation. As EMSL's user research expands and matures, new and enhanced capabilities will be developed. Additionally, existing systems will be modified to support the needs of the user community. Beyond

5 years, we should see growth in the body of biological data that EMSL and PNNL will maintain for searches and analysis. The increased focus on omics, multiscale modeling, and improved imaging capabilities will significantly increase the data volumes for complex samples analyzed by EMSL.

A new generation of mass spectrometers for proteomics applications are being developed that should increase sample throughput and data output by multiple orders of magnitude. Access to the massive sets of data generated by these new instruments will significantly increase network requirements. Much of this data will be transferred in from offsite, combined with existing data, curated, and shared back with the scientific community.

Strong integration of data from multiple scientific domains to allow users to address systems-level problems will require EMSL to manage and integrate multi-petabyte-scale data sets. This will require the development of complex workflows accessing data generated and stored at EMSL with data from other user facility and research laboratories, which will significantly impact network requirements.

10.6 Collaboration tools

EMSL provides a set of collaboration tools for users and their collaborators via `http://emslhub.emsl.pnl.gov`. MyEMSL will provide shared file access to authorized users, and public access to data that has been released. All of the above will rely on a standard set of protocols including HTTP, ssh, FTP, VNC.

10.7 Data, Workflow, Middleware Tools and Services

EMSL has yet to determine what middleware and services might be required. Interest has been growing in external private and public clouds, but use cases for cloud computing in these scientific fields have not yet been identified.

10.8 Outstanding Issues

EMSL frequently needs to ship physical copies of media to users when data sizes exceed a few gigabytes. More often than not, this is due to lack of bandwidth or storage resources at the user's home institution. There are also cases where good tools do not exist to transfer large data sets. MyEMSL is expected to resolve the lack of robust file transfer tools.

Table 10.1: The following table summarizes data needs and networking requirements for EMSL.

Key Science Drivers			Anticipated Network Needs	
Science Instruments and Facilities	**Process of Science**	**Data Set Size**	**LAN Transfer Time needed**	**WAN Transfer Time needed**
Near Term (0-2 years)				
Broad suite of scientific instruments including: · High Resolution Mass-Accuracy Capability (HRMAC) · Cascade 2.54 PFlop supercomputer · SOLiD Sequencer · Dynamic Transmission Electron Microscope (DTEM)	· Primarily onsite access to instruments · Remote access to Cascade computer, Aurora archive, and remote instrument operation · New MyEMSL data management system with search · EMSLHub with collaboration tools, pilot workflow and analysis capabilities · Ad hoc transfer of genomic data to EMSL of 25 GB – 50 TB a few times per year	· Data volume of 3-5 Tbyte/day ingested into EMSL archive, primarily from : · DMS: 2 – 8 Tbyte/month · SOLiD Sequencers: 1-2 TB every 10-12 days · HRMAC: 1 Tbyte/day DTEM: 1 Tbyte/day	· 5 Tbyte/day continuous, 24x7	· 200 GB per month at 1 Gbit/sec outgoing · Data are transferred mostly to users' home institutions · Up to 50 TB terabyte ad hoc data transfers
2-5 years				
Existing instruments plus: · Next-generation HPCS-5A and 5B replace Cascade supercomputer	· Enhanced integration of data from multiple instruments · Collaborative data access and analysis	Estimated aggregate data generation rate of 20 Tbyte/day with growth from: · HRMAC: 5 Tbyte/day	· 20 Tbyte/day continuous, 24x7	· 150 Tbyte/month at 1 Gbit/sec to user's home institutions · 50-100 Tbyte/month at 1 Gbit/sec from ARM, JGI, JBEI, KBASE
5+ years				
Existing instruments plus: · HPCS-6 HPC system(s) to replace HPCS-5 systems · Ultrafast Microscopy Capability (UMC)	· Strong integration of data across capabilities · Comprehensive problem-solving environment on top of MyEMSL	Estimated aggregate data generation rate of 200 Tbyte/day with growth from: · UMC: Estimated 100 Tbyte/day	· 200 Tbyte/day continuous, 24x7	· 1 Pbyte/month at 10 Gbit/sec to user's home institutions · 100 – 200 Tbyte/month at 10 Gbit/sec from ARM, JGI, JBEI, KBASE

Case Study 11

The Earth System Grid Federation

11.1 Background

This document describes three use cases for DOE's BER CESD projects supporting Model Intercomparison Projects (MIPs) and satellite and instrument observations. More specifically, these use cases address the requirements of the international Sixth Phased Coupled Model Intercomparison Project (CMIP6), Observations for MIPs (Obs4MIPs), and DOE's Accelerated Climate Model for Energy (ACME). Focusing on these national and international inter-agency projects will cover a broad spectrum of needs for additional projects such as the other 70+ MIPs and other observational data efforts, such as the ARM observational facility and related observational data sets from the Carbon Dioxide Information Analysis Center (CDIAC). An additional use case was added to cover the European and U.S. statistical and dynamical down-scaling project, known as the COordinated Regional climate Downscaling (CORDEX), which is used to determine regional climate changes.

These projects and other communities systematically rely on the Earth System Grid Federation (ESGF) software infrastructure now and in the distant future to securely manage, serve, and manipulate their data. ESGF is a multi-agency, international collaboration whose purpose is to develop the software infrastructure needed to facilitate and empower the study of climate change on a global scale. ESGF's architecture employs a system of geographically distributed peer nodes that are independently administered yet united by common federation protocols and application programming interfaces. The cornerstones of its interoperability are the peer-to-peer messaging, which is continuously exchanged among all nodes in the federation; a shared architecture for search and discovery; and a security infrastructure based on industry standards.

11.2 Context

The primary ESGF archives needed for CMIP6 will be distributed between several data centers, on different continents, using different storage architectures consisting of tape storage and/or rotating disks. As shown in Figure 11.1, the data centers include the United States, the Lawrence Livermore National Laboratory; Australia, the National Computational Infrastructure; Germany, the Deutsches Forschungsnetz; Great Britain, the Centre for Environmental Data Archival; and the Netherlands, the Royal Netherlands Meteorological Institute. Also shown are future data centers which include China, Beijing Normal University; France, the Institut Pierre Simon Laplace; and Japan, the University of Tokyo. These CMIP6 data centers will be connected to each other and to dozens of climate modeling centers via multiple high-speed research networks shown in Figure 11.1 (i.e., ESnet, Internet2, Jisc, SURFnet, DFN, AARNet, and GÉANT).

Managing and accessing the petabytes of data will be achieved through the use of the ESGF software stack. Through ESGF, modeling centers will retain ownership of their data and allow community access and replication of their data to the CMIP6 data centers. ESGF will use the high-speed Internet and the Globus/GridFTP software to move large amounts of data from one location to another (i.e., between modeling and data centers and to the end user). In this paradigm, more popular or requested data will be replicated at the primary ESGF data

Figure 11.1: The network connections and collaborations made through the network groups help climate and computational scientists manage and disseminate petabytes of modeling and observational data, which traverse more than 13,000 miles of networks, spanning two oceans and three continents, with more connections planned, including, IPSL (France), NASA (US), NOAA (US), and universities in China and Japan.

centers for greater data accessibility and manipulation. Changing of the original data sets can only occur at the modeling centers. If data does change, controlled versions of that data will be replicated at the primary ESGF data centers.

The World Climate Research Programme (WCRP), the leading international body for coordinating and facilitating climate research that oversees CMIP and the other 70+ MIPs, have endorsed ESGF as the standard for managing and distributing data for projects which it supports. Therefore, the use case for CMIP hold true for most (if not all) of the other MIPs and climate research intercomparison data efforts.

The DOE ACME project is designed to create and operate a test bed for advancing Earth system model development and has among the most varied data management needs. ACME scientists will perform many short model runs with rapid turnaround during the model development phase. This involves more computational demands for uncertainty quantification and optimization work for model refinement and massive data runs at leading DOE supercomputer centers, shown in Figure 11.1 (i.e., NERSC, OLCF, and ALCF). For this effort, the full array of ESGF features is needed for model production and analysis. As with the connection to primary ESGF data centers, the DOE supercomputer centers are connected via ESnet's network and the petabytes of data are managed and manipulated via the ESGF software stack. For model validation and verification, ARM and NASA's Distributed Active Archive Centers (containing observation and reanalysis data) are also served through ESGF for DOE model diagnostics efforts.

The United States alone has many climate modeling effort (i.e., Community Earth System Model, CESM; Goddard Institute for Space Studies, GISS; Model for Prediction Across Scales, MPAS; Climate Limited-area Modelling, CLM; CAM, etc.) and many of them rely on ESGF to server their model results to the community. ACME is a good representative of a model development use case.

Currently, there are over 20 web portals (or gateways) for registering and accessing ESGF data and services and over 70 nodes for data and software provision are presently in use. With these portals and many other software libraries, packages, and sub-components are integrated in an infrastructure (or data ecosystem) that enables real-time comparison of model output to observational measurements. This controlled environment minimizes or eliminates tedious activities associated with climate research. The ESGF data ecosystem of components contain:

- Critical Complex Data Generating Systems that generate petabytes of data from sophisticated technology sources, ranging from high-end supercomputers, clusters, and computer servers to sensitive environmental detectors, lab analyses, and orbiting satellites;
- Data Collection and Management, which collects, stores, and organizes data for easy user discovery and accessibility;
- Data Analytics for pattern discovery, structure identification, dimension reduction, image processing, machine learning, and exploratory visualization;

- Data-Intensive Computing for describing applications that are input/output (I/O) bound and enabling large and complex data manipulations; and
- Decision and Control for decision control and knowledge discovery.

11.3 Specific Projects Use Cases

11.3.1 Use Case 1: CMIP and obs4MIPs

The climate community has worked for the past decade on concerted, worldwide modeling activities led by the Working Group on Coupled Modeling (WGCM), sponsored by the WCRP, and leading to successive reports by the Intergovernmental Panel on Climate Change (IPCC). The Fifth Assessment Report (IPCC-AR5, CMIP5), released in September 2013, was the latest report. The Sixth Assessment Report (IPCC-AR6, CMIP6) is scheduled for release in 2019. These activities involve tens of modeling groups in as many countries, running the same prescribed set of climate change scenarios on the most advanced supercomputers and producing several petabytes of output containing hundreds of physical variables spanning tens and hundreds of years. Over the years, the successive CMIP experiments have produced the following archive sizes:

- CMIP1 (1995): 1 GB
- CMIP2 (2001): 500 GB
- CMIP3 (2007): 35 TB
- CMIP5 (2013): 1.8 PB

The sixth assessment is project to be 50 times larger than CMIP5:

- CMIP6 (2019): 1.8 PB * 50 = 90 PB (projected)

These data sets sizes range from megabytes to terabytes with individual file sizes averaging 250 MB and are held at distributed locations around the globe. With increased resolution for CMIP6 output, file sizes are expected to increase 1 TB. As with CMIP5 all data for future ESGF projects are expected to be discoverable, downloadable, and analyzable as if they are stored in a single archive, with efficient and reliable access mechanisms that span political and institutional boundaries.

The same ESGF infrastructure also allows scientists to access and compare observational data sets from multiple sources, including, for example, NASA Earth Observing System (EOS) satellites such as those found in obs4MIPs. These observations, often collected and made available in real-time or near real-time, are typically stored in different formats and must be post-processed to be converted to a format (e.g., netCDF-CF) that allows easy comparison with model output. The need for providing data products on demand, as well as value-added products, adds another dimension to the capability demands. Finally, science results must be applied at multiple scales (global, regional, and local) and made available to different communities (scientists, policy makers, instructors, farmers, and industry).

Because of climate research high visibility and direct impact on political decisions that govern human activities, the end-to-end scientific data investigation must be completely transparent, collaborative, and reproducible. Scientists must be given the environment and tools for exchanging ideas and verifying results with colleagues in opposite time zones, investigating metadata, tracking provenance, annotating results, and collaborating in developing analysis applications and algorithms. This virtual collaboration environment that facilitates and advances scientific discovery is precisely the data ecosystem environment that inspires and motivate the ESGF project.

11.3.2 Use Case 2: ACME

Based on their detailed workflow requirements, general use cases within DOE ACME can be separated into three distinct categories. The first is the process for developing a new capability within the model, which requires many small runs with rapid turnaround of the workflow steps, significant interaction with software tools, and automated

testing and version control. The structure of the output varies and needs to be easily accessible through short-term, local archives. Plotting and analysis need to be more interactive, nimble, and extensible for the user as development proceeds.

Secondly, exploratory use-cases and their workflows involve numerous and varied length and spatial-scale simulations with single or multiple components activated, potentially using ensembles for uncertainty quantification and optimization to explore parameter space and model fidelity. Output is shared within small groups of project scientists both locally and externally using short- to medium-term archiving. Interactive, web-based, visualization tools are required to incorporate High-Performance Computing (HPC) information that is especially useful at this stage for diagnosing issues before full production runs begin. Provenance is also necessary to record testing and evaluation steps required for paper and data publishing in development-focused journals.

Thirdly, production runs of the model comprise the most substantial and diverse set of use cases. Collections of ensembles are performed over months and may be transferred to multiple staff as they proceed. Large jobs are queued on ASCR computing facilities (i.e., NERSC, OLCF, and ALCF) systems where large data sets are created, and complete provenance and archiving infrastructure is required for data publishing to other collaborators and eventual public release. The data generated at these computing centers will be federated between the sites for backup and ease of use.

11.3.3 Use Case 3: CORDEX

Regional climate downscaling (RCD) techniques, which include both dynamic and statistical approaches, are being increasingly used to provide higher-resolution climate information than is available directly from contemporary global climate models (GCMs). The techniques available, their applications, and the community using them are broad and varied, and it is a growing area. It is important, however, that these techniques and the results they produce be applied appropriately and that their strengths and weaknesses are understood. This requires a better evaluation and quantification of the performance of the different techniques for application to specific problems. A coordinated, international effort to objectively assess and compare various RCD techniques, built on experience gained in the global modeling community, is providing a means to evaluate RCD performance, to illustrate benefits and shortcomings of different approaches, and to provide a more solid scientific basis for impact assessments and other uses of downscaled climate information. WCRP views regional downscaling as both an important research topic and an opportunity to engage a broader community of climate scientists in its activities. CORDEX has served as a catalyst for achieving this goal.

One of the ESGF's successes has been in getting data out to the community in a coordinated manner, using a single and documented format and file structure. It has been decided that CORDEX will use the same ESGF infrastructure as CMIP. Thus, the same facility now exists for CORDEX data. The IS-ENES2 (European) community took responsibility for implementing several adjustments to the process:

- Data reference syntax (DRS) has been adapted for dynamical downscaling;
- Attribute service (data access and term of use) is operated by Linḱoping University;
- Versioning of data sets has been done at the variable level; and
- Quality control (by DKRZ) has been done prior to the publication.

11.4 Modeling and Data Center Requirements

Producing and/or maintaining high volumes of scientific data within an HPC and high-performance storage environment present unique production and operating challenges. Often, the only realistic choice for long-term storage and backup are robotic tape drives within a hierarchical management system. Access constraints (security policies and firewall restrictions) set by modeling, HPC centers, or data centers can lead to such inefficient behaviors. Redundant efforts waste considerable resources and significantly slow scientific productivity. Additionally, some applications require that large volumes of data be staged across low-bandwidth networks simply to access relatively small amounts of data. Finally, when data usage changes, or storage devices are upgraded,

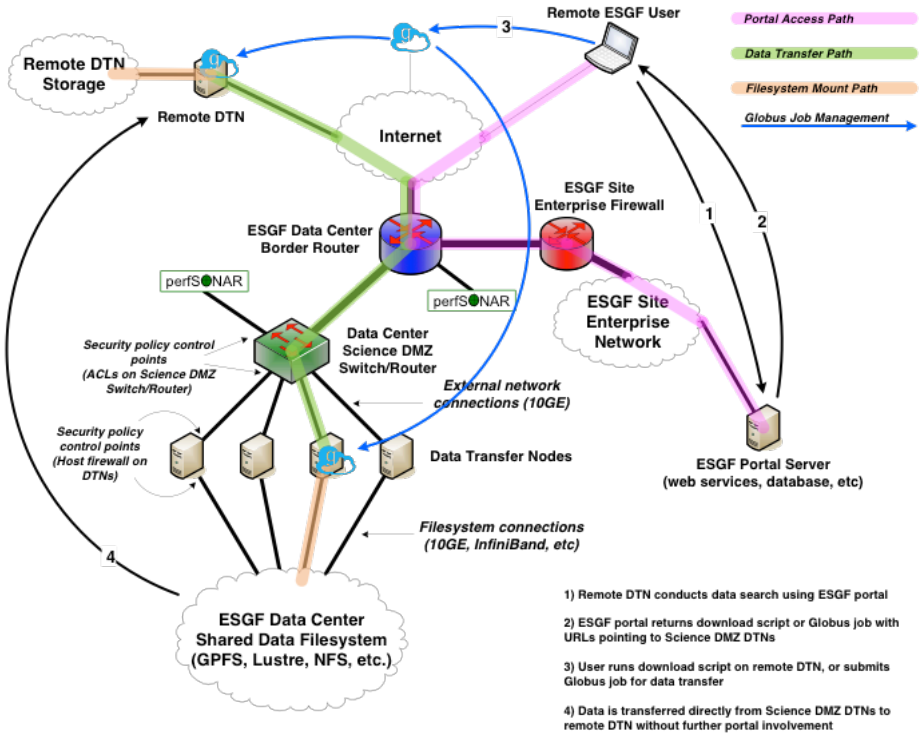


Figure 11.2: Future standardized ESGF node and Internet workings and connections.

large data sets may need to be reorganized to take advantage of the new configuration, making the entire process expensive and extremely time consuming for the centers or operating facilities. For these and other reasons, there is a need to develop a more intelligent and integrated data ecosystem across the ESGF that provides services that anticipates usage, storage, disruptive technologies, and better communication and network connections between the centers and facilities. For backup purposes and greater accessibility, there is a need to move petabytes of data between the worldwide-federated centers that are supported by ESGF. At present, the tested disk-to-disk sustained throughput is 4 Gbps, which is approximately delivering data at a transfer rate of 1 PB a month. However, for data replication of tens to hundreds of petabytes at any given sites, the long-term goal is a sustained disk-to-disk performance of 1 PB a day. With proper tuning and improved networks, as shown in Figure 11.2, we hope to achieve this goal before 2019 CMIP6 production runs.

Case Study 12

A Globus Perspective on BER Research Data Management

12.1 Background

We describe data management requirements that the Globus team at ANL and the University of Chicago has encountered across a range of BER-related facilities and projects; highlight, from across those facilities and projects, common themes and requirements for data transfer and collaborative data sharing that motivate networking needs; and describe how facilities and projects use Globus research data management services to address many of the requirements that we identify.

Globus leverages modern software-as-a-service (SaaS) methods to provide reliable and secure data transfer, sharing, publication, and discovery services. Its Web 2.0-based interfaces have proven popular with both facility administrators and end-user researchers and engineers, allowing the former to outsource time-consuming operations and support tasks to a reliable third party, and permitting the latter to work with large data with ease. Globus services are deployed at most DOE laboratories, form part of the ESnet-designed Science DMZ concept, and are used in major BER projects such as ACME, ARM, and KBase.

This case study presents the perspective of a solution provider rather than a single facility, instrument, or project. As such, it provides a composite picture of use cases and trends in data management across numerous campuses and facilities.

We make four main points:

- High-speed, reliable, and extremely easy to use data transfer, sharing, and publication have become vital to a wide variety of BER science projects.
- Globus services have proved highly useful in terms of reducing barriers to the use of advanced networks, improving performance and reliability of data transfers, and facilitating data sharing and publication.
- The SaaS approach used by Globus has proved effective both for facilities, developers of software tools, and individual researchers. It reduces demands on system administrators, software developers, and researchers, while also increasing usability, reliability, and performance.
- Globus services are widely deployed and used, but there remain many opportunities to use them to a far greater extent.

12.2 Network and Data Architecture

Globus services are used by hundreds of research institutions across the United States and internationally. More than 8,000 computer and storage systems have been configured as Globus endpoints by installing the simple

Globus Connect software that allows them to participate in the Globus data transfer and sharing network. Registered users use web, command line, or application programming interfaces (APIs) to request data transfers between Globus endpoints and to configure data sharing on endpoints. In every cases, access is controlled based on secure authentication and authorization. The Globus software uses the high-performance GridFTP protocol to accelerate transfers between a wide variety of storage systems, including hierarchical storage management systems, and over a wide range of networks. Integrity checking is performed on all transfers by default and encryption can be applied if desired.

The Globus software that manages transfers is implemented as SaaS on the Amazon cloud. This SaaS approach provides both the Globus team and the administrators of individual endpoints and facilities with a high degree of visibility into data transfer patterns and achieved performance. In addition, we work closely with many groups who are applying Globus services in their science, for example by advising them on how best to configure Globus endpoints or how to integrate Globus APIs into their applications. We thus have experience with a wide range of network and data architectures. We draw some general conclusions in the following.

We consistently see the highest performance being achieved at facilities and campuses that have **adopted the ESnet Science DMZ model**, with its dedicated data transfer nodes, friction-free network access to storage, and appropriately configured Globus software. In contrast, sites that lack dedicated and capable data transfer nodes, operate aggressive firewall security policies, and/or fail to provide Globus services to their users often provide end-to-end transfer experience that is not practical for science requirements.

Our experience also emphasizes that **networking issues do not stop with the border router**. Many users suffer from last mile problems within a facility or campus. Connectivity to researcher-used machines such as local department clusters or desktops must be considered if we are to truly address the end-to-end data transfer issues that users contend with today. For example, at ANL's Advanced Photon Source (APS, used by many BER-related projects), despite sustained effort to improve internal networking, access to remote sites remains relatively poor. Some APS beamlines have established Globus transfer endpoints closer to the acquisition machine in order to permit high-speed data movement. Such last-mile connections must be viewed as an essential service if researchers are to be able move data efficiently to other systems or outside the facility. Similar concerns arise at some ESGF sites, with local network configuration issues hindering high-speed access to important climate data.

One potential solution to such problems is to create **data staging services** to which data that is intended for external distribution can be transferred. For example, ANL has established the Petrel system, with 1.7 PB storage, for this purpose[1]: see Section 12.4.1. This system is used by both APS and climate projects to locate data that is intended for external access.

We also find that **international network connectivity is important** for key efforts supported by BER, such as Earth System Grid Federation [7]. International network connectivity with key sites on the federation is critical. This is important for user access to distributed archives, and for administrative purpose of geographic replication of assets for localized efficient access and disaster recovery.

Later sections of this case study highlight some solutions adopted by facilities that have resulted in high-performance data access infrastructure while staying true to the security tenets required by the facilities.

12.3 Collaborators

We work directly or indirectly with many thousands of researchers and facilities. Globus services are in use at thousands of sites across the United States and worldwide. More than 25,000 registered users have used the Globus service to process more than 20 billion files and to transfer more than 110 petabytes since late 2010. More than 8,000 storage and computer systems run Globus endpoints. Figures 12.1, 12.2, and 12.3 show usage *across DOE labs*. Usage is growing steadily. (These numbers only encompass data transferred with the Globus transfer service; total Globus GridFTP data transfer is more than a petabyte per day.)

[1] `http://petrel.alcf.anl.gov`

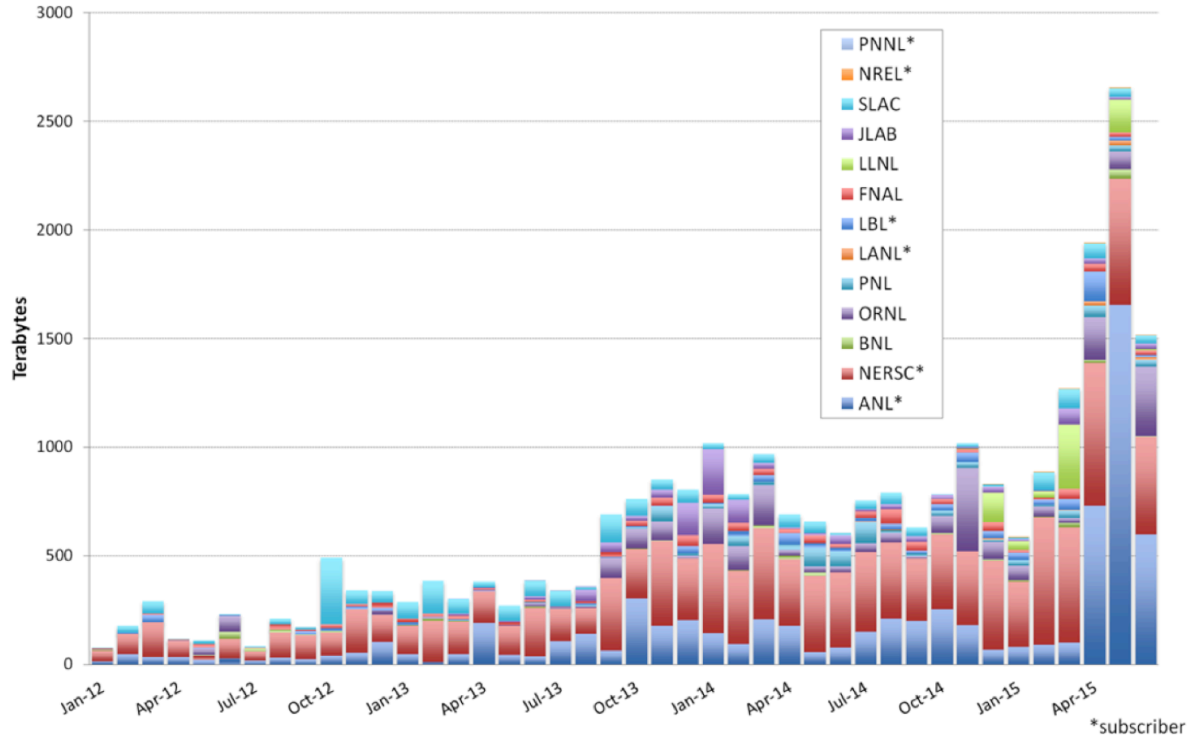


Figure 12.1: Globus transfer service usage at DOE labs, as measured in total bytes transferred per month.

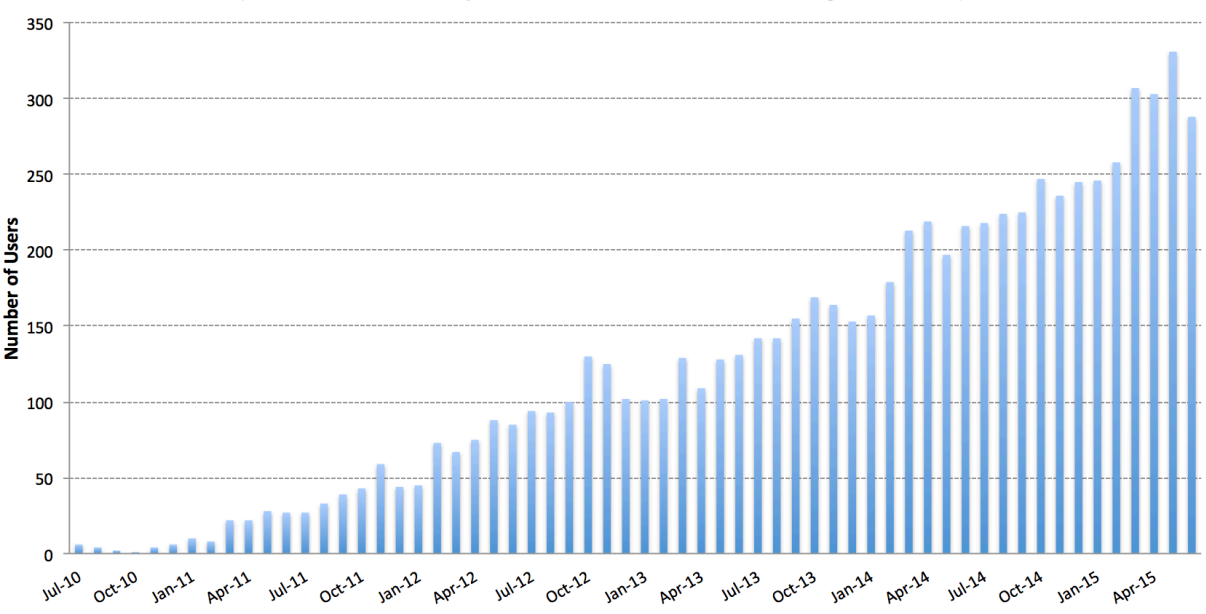


Figure 12.2: Cumulative data transferred via Globus in and out of DOE facilities.

12.4 Instruments and Facilities

We review use cases from a range of BER-relevant facilities and science projects.

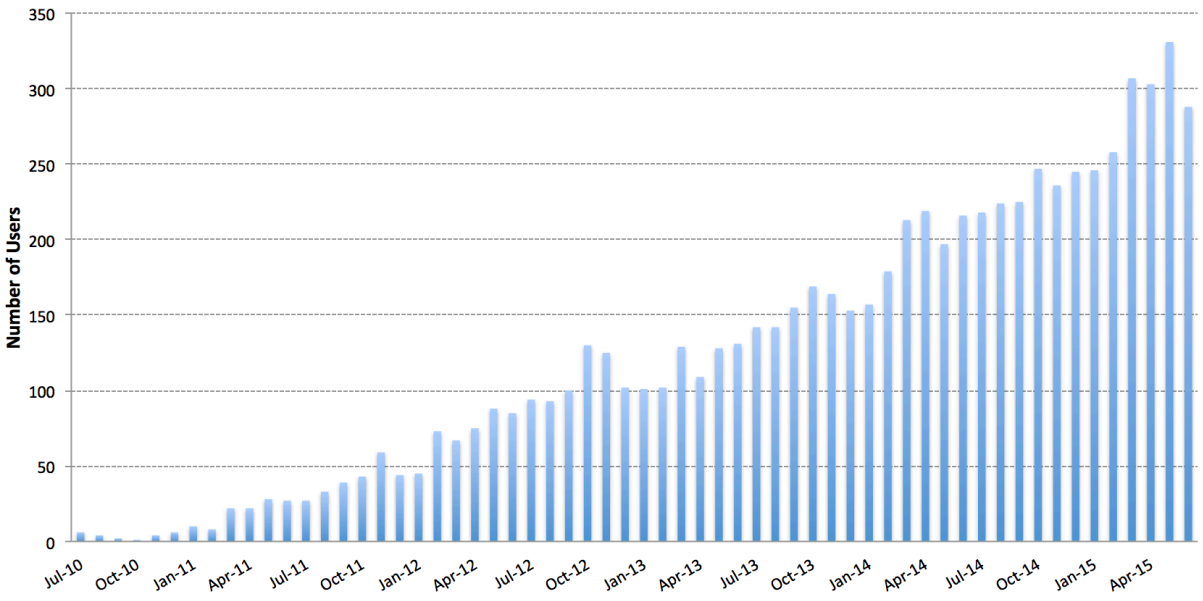


Figure 12.3: Number of users who access Globus endpoints at DOE facilities per month.

12.4.1 Present

The researchers with whom we work use a wide variety of instruments and facilities. The quality of the end-to-end data transfer path to/from these instruments facilities varies widely. Some have high-speed ESnet connections, powerful data transfer nodes, well-configured Science DMZs, and well-configured Globus endpoint software, so that users can move data to/from a facility at high speed. Many other facilities are lacking in one or more of these respects, compromising achieved performance and/or usability.

Security policies are a frequent obstacle to effective data movement to/from instruments and facilities. BER facilities have varying levels of security requirements, motivating the need for supporting different levels of assurance for secure data access. For example, some sites require the use of two-hop systems in which data must be staged to a separate server before it can be transferred out or shared. Others run software-defined network solutions that allow bypass in firewalls. Yet others require read-only mounts of externally accessible file systems.

Friction-free network paths can be engineered in such cases, but require effort and care if we are to allow data transfer and sharing in meaningful timelines with good usability for end users. For example, Los Alamos National Laboratory (LANL) has integrated their network bypass policy engine with Globus transfer management capabilities (management console) so that a Globus transfer's data channel can bypass their packet inspection firewall while the transfer is active; the control channel goes through their standard firewall inspection. This approach allows data transfers to leverage available network bandwidth while adhering to the high levels of assurance and security measure that LANL needs.

The Petrel data server at ANL represents another approach to problems of both security and data access. The APS and ALCF require a local account to access their storage services. For this and other reasons, ALCF and Globus set up a 1.7 PB data store, Petrel (`http://petrel.alcf.anl.gov`). This service uses Globus identity and groups, and data transfer and sharing, to allow users to share data with collaborators. The pilot has about10 groups using it to share their data, including Argonne climate scientists. In the last year 600 TB of data has been moved in and out of the storage system using Globus.

12.4.2 Next 2–5 Years

We expect to see the widespread development and use of large *discovery engines* [5] (see Figure 12.4): facilities that integrate big storage and compute resources to enable the aggregation and analysis of large quantities of

data from different sources. KBase is an early example of such a system. We expect to see similar systems arise in environmental science and other fields.

The development of these systems will be motivated by the need to colocate data and computing to facilitate rapid and collaborative analysis. However, they will inevitably increase demands on networks.

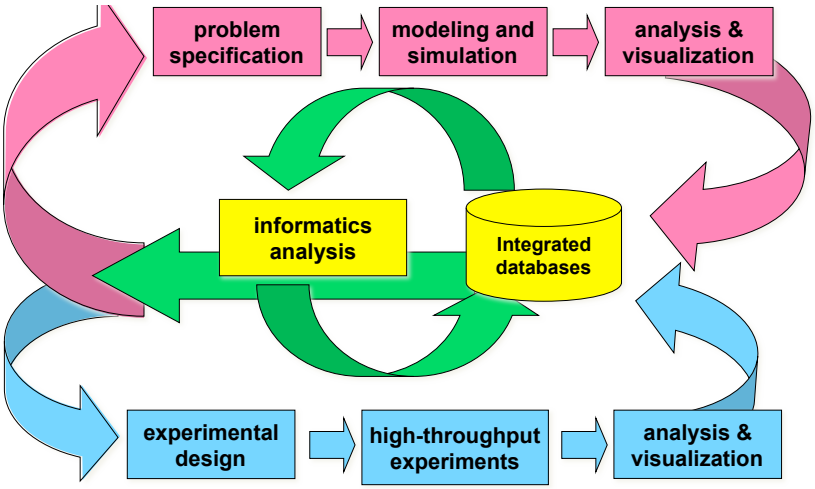


Figure 12.4: Discovery engine schematic. [Figure courtesy of Rick Stevens.]

12.4.3 Beyond 5 Years

Large-scale automated science processes will increase data volumes yet further.

12.5 Process of Science

12.5.1 Present

We outline two sets of use cases that encompass important aspects of the current state of the art.

Data movement and sharing

Rapidly increasing data volumes lead some to argue that data movement is no longer feasible or desirable. Our experience, however, is that use cases that require large-scale data movement are becoming more rather than less important for science. We see a wide range of rationales for data movement. For example, data may need to be transferred from a generation site to a storage site; from a generation or storage site to an analysis location that has available computing power or specialized software; from a storage site to another storage site for backup or redundancy; or from one storage site to another for integration with other data. In each case, reliable, secure, and high-speed data movement can increase the pace and reliability of research, and enable new research methodologies not otherwise possible.

A related set of use cases involve sharing data with collaborators, as for example when a researcher produces a data set that must then be shared with other members of a research team for further analysis.

Wide-area network performance can certainly be an obstacle to these use cases. However, it would be a mistake to think that high-speed wide-area networks are all that is required. Other factors can also cause problems: for example, local network configuration, security requirements, local storage system configuration, and lack of suitable data transfer tools. Integrated solutions such as Globus that address security, reliability, bookkeeping,

transfer optimization, and sharing management in an easy-to-use package can make a big difference to research effectiveness.

Near-real time data movement and sharing

We see a strong use case for moving data as it is generated, whether at scientific instruments or by simulations, to analysis machines and collaborators in near-real time. Inevitably analysis machines and collaborators may be located remotely as well as locally. This requirement adds a quality attribute to the previous requirement in terms of timed delivery.

One motivation for moving data to analysis machines in near-real time is to provide feedback on the data collected which can be used to make changes and adjustments to the data collection oricess, thus improving the quality of the data collected and ensuring success of the time spent at the facility. An exemplar of this was work done by researchers at APS, PNNL, and Globus to enable such analysis for an experiment conducted at APS at PNNL. Erin Miller, a scientist from PNNL conducted an experiment at APS and used Globus to move data back to PNNL for analysis and visualization. Such an endeavor requires high-performant networks between facilities, data transfer nodes at the border of each facility to allow secure data movement, and resilient transfer tools to leverage the network.

Workflow leveraging data at facilities

Integrating use of compute at these facilities within a workflow that involves data discovery and staging also poses unique challenges that are discussed in later sections of this case study.

Data distribution and replication

Many projects host data assets and provide two important but distinct use cases that dictate networking requirements: user access to data subsets, restricted user access to derived data, and data replication. Examples of such projects include ESGF, ACME, and ARM.

The first use case involves providing user access to data subset. A user will typically want to search for and discover data set of interest and then download the data, often to a local machine or other institutional machine. Access is typically read-only, and the user machines to which data is downloaded may not have good bandwidth, so downloads can be slow. This use case motivates the need for highly simple and usable tools that make it trivial for a user to obtain the data. (A slight variation on this use case is where users want to move a significant portion of the data to other well-connected sites for further analysis. That variant emphasizes a need for well-tuned networks and tools that can leverage the network.) ESGF with multi-petabytes of data holdings across international distributed site is an exemplar of this use case.

ESGF has integrated Globus as an option for user data download. In so doing, it leverage the platform aspects of the transfer service, where the Globus capability for identity management and data transfer are integrated into the ESGF portal giving users a seamless user experience while using the managed data transfer tool. Recently Globus download was also added to the ESGF COG front-end, a new system adopted by ESGF as a front end portal. Work done in collaboration with NASA Jet Propulsion Laboratory (JPL) also leverages Globus sharing for publicly accessible data sets, thus reducing the number of user logins in the end-to-end download process. This also bring third-party transfers to the ESGF ecosystem, supporting the variation on the end user download use case to transfer data to other campus and facility systems that are Globus enabled.

A second major use cases is when projects allow users to request some custom data analysis and then allow access to the derived data that results from the analysis. In such cases, automation of analysis runs and sharing of the result set with the requesting user are key. It can also be important to track user download of the derived data, so as to determine when it can be safely deleted. NCAR's Research Data Archive (`http://rda.ucar.edu`) is a project that has successfully demonstrated an end-to-end solution for this use case by integrating their portal with Globus services, which they use for federated identity, data transfer, and data sharing. With scale of data assets over 17 PB of data processed for over 4000 custom user requests last year, the inclusion of Globus for data

management has allowed the archive to leverage a specialized service for secure data sharing, while providing users an intuitive and managed data delivery solution.

Lastly replication of these data sets across sites that are often well connected is a crucial requirement for many such data archive projects. These are administrator-driven operations that are repeated at some interval where potentially large data assets (for example petabytes of data in the case of ESGF) is replicated across sites. This operation requires write access on the remote side, often synchronization using checksums to move only data that has been added or changed and in some cases involve international sites. Networking requirements and tolerance for successful completion of transfer is quite different from the end user data movement in this case

The automation of data transfers to support workflows is a key part of the science process. Most often researchers use experimental facilities collect data or compute facilities to generate data, and that becomes the first step in a workflow process. First step usually is the validation of the data and there is a spectrum of use cases covering various time thresholds for this analysis. For example, certain beamlines at APS, including those that do X-ray Photon Correlation Spectroscopy (XPCS) require real-time analysis and feedback that takes order of minutes. On the other hand, climate model runs by ACME will use data for the first five years to run diagnostics to see the quality of the run and determine if it will be continued or not.

In all these cases, data needs to move from source to where the analysis is done, often as soon as a file is written. This requires good tools for automation that also do credential management for data access and secure transfer. Globus is used in some deployments where the programmatic interface (REST APIs) or the scriptable command line interface is integrated with the workflow and used to move the data.

Once initial analysis is done, often further steps of the workflow are used to create derived or value-added products, that are then shared with collaborators. While the data in this step is usually smaller in size, the need to share it introduces requirements around moving it to a location where it can stored, backed-up, fine-grained access control can be setup for access, and has a high-performance network to support multiple user access.

Once the data has been processed, publishing the data to make it available to the community is often a next step. Publishing of data involves placing the data in storage that has preservation capabilities matching requirements (few years to long-term archive), associating metadata with the published data, curation as needed, and attaching a persistent identifier to the data set that can be used for reference and discovery. In some projects like the ones that use ESGF, publication is a key aspect of the system, and the discovery of data is facilitated by maintaining an index of the published artifacts. As described in Section 12.7, Globus publication and discovery services support this use case.

12.5.2 Next 2–5 Years

We expect to see many more data distribution and analysis services established at DOE facilities and elsewhere, as DOE as a whole and BER in particular engages works increasingly with “big data” and with methodologies that involve integration of data from different sources and of different types.

Demand for Globus services as a means of transferring, sharing, and publishing data will increase rapidly. We also see an increased need for services that allow users to organize, manage, discover, and manipulate large quantities of diverse distributed data.

12.5.3 Beyond 5 Years

We expect the more distant future to feature increasingly complex data management and sharing structures that will emphasize the need for far more sophisticated research data management capabilities.

As facilities grow to exascale, we expect data volumes to increase significantly. For example, file sizes that have largely remained constant over a decade are expected to significantly increase such that tools for data management, including those that support movement of such files, need to cope with files that are on the order of terabytes. Innovative solutions to address this, such as improvements to protocols to support efficient striping of files, and mechanisms for scaling network usage with the number of data transfer nodes and their capacity, are critical to support science in an exascale environment.

12.6 Remote Science Activities

Science is collaborative in nature and most if not all projects have an element of remote science driven by multi-institution collaboration for leveraging specialized resources (such as instruments, supercomputers) and specialized skills (core facilities, collaborators with expertise). All requirements presented in this case study cover the remote science aspect.

12.7 Software Infrastructure

This case study is concerned primarily with the use of Globus services and thus it is in this section that we focus our attention.

12.7.1 Present

Globus is widely deployed at many institutions, both national labs and campuses, and in some cases the default service to move data. As a hosted service, Globus delivers high-performance managed data transfer, sharing, and publication capabilities that integrate seamlessly with a site or cluster's authentication system.

Globus transfer [1] automatically tunes transfers to take into account characteristics of both the transfer endpoints and the data being moved. For example, it configures transfers differently when moving many small files vs. a few big files. The service also delivers critical capabilities for site administrators who want to deploy and host a service to enable their storage for use with Globus. These capabilities include trivial setup and configuration of Globus endpoints; the ability to configure concurrency and parallelism preferences and limits; SaaS-based tools for monitoring all transfers in and out of their system (e.g., see Figure 12.5); and the ability cancel or pause/resume tasks on their system.

Globus sharing [2] gives the end user the ability to share folders with their collaborators directly off their storage system. Users simply select the folder(s) that they want to share and provide the email address of the person they want to share. Globus notifies the other user, and gives them read or write permissions as specified by the owner. The shared data is not replicated or copied elsewhere, thus making it a scalable solution for sharing data sets of any size with partners and collaborators. As described on other, this capability is already in use by several projects to securely share data: RDA and ESGF are examples of such use. LBNL has enabled sharing for some projects, allowing scientists to leverage storage at the lab to store the data and Globus to facilitate controlled sharing of the data.

Globus has become a vital part of the data management infrastructure across various DOE facilities: as shown in Figures 12.1–12.3, we see steady growth in use across DOE facilities, both in the amount of data transferred and the number of active users. Since April 2014, there have been at least 200 active users per month, with over 300 users in the last few months. The service is starting to replace use of other transfer tools such as scp and rsync, due to its higher performance and more robust and user-friendly transfer and sharing solution.

Globus transfer and sharing is also used as platform to integrate with various workflow systems: Swift and Galaxy both provide mechanism to use Globus transfer to stage data in and out for workflows. Galaxy has also been enhanced to use Globus identity and group mechanism to provide federated login and credential management capabilities to its users. Swift also integrates with the Globus metadata management capability, allowing users to associate metadata and tags with files that are part of a data set, search and discover data sets, and then stage them for further processing. Under ACME, work is being done to add Globus transfer to Pegasus workflow system.

Globus publication and discovery services [3] enable data publication and discovery with user-customizable collections. The hosted publication service provides the workflow needed for the publication process, while the data is stored in the institutional storage with references to the data added to the metadata. Using the Globus publication service, users can define collections and configure various aspects of the collection: policies on who can submit to the collection, curation requirements, permissions on who can view the published collection; metadata needed to publish to the collection, which is presented as forms to the user to provide input on; and lastly

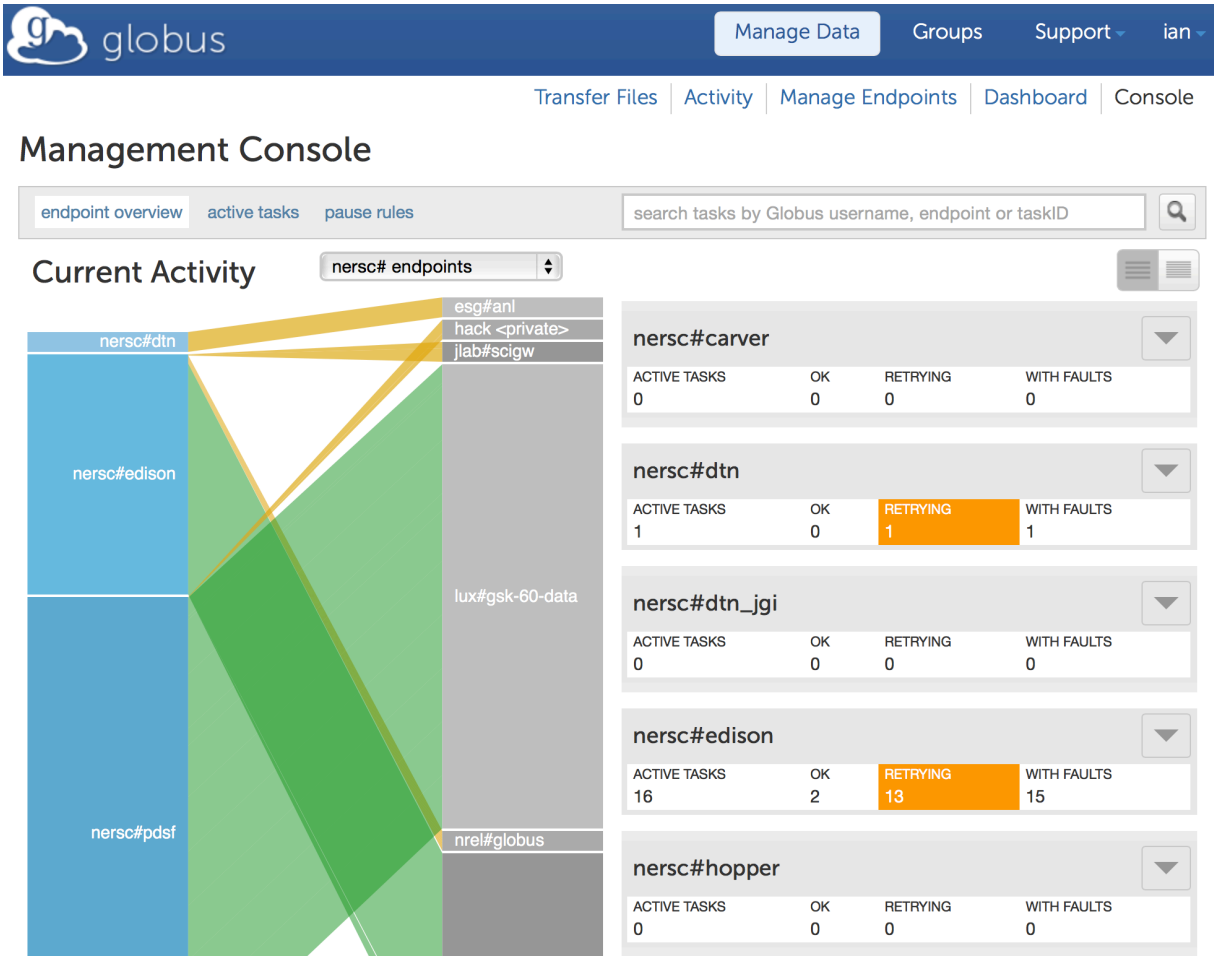


Figure 12.5: The Globus management console, showing here a view relating to selected NERSC endpoints.

persistent identifier to associate with artifacts published to this collections (DOIs or handles). The Globus publication service indexes the metadata and provides a rich search interface for discovery of the published artifacts. Users can search and find data sets, and depending on the access permissions set on the collection, download the data.

For the BER ACME project, the Globus publication service is being further enhanced to support automated extraction of metadata from files and publishing to the Thematic Real-time Environmental Distributed Data Services (THREDDs) service for Open-source Project for a Network Data Access Protocol (OpenDAP) access. The ACME project collection has been modified to have an additional step for extraction of metadata from NetCDF files, generation of THREDDs catalog and publishing to the ESGF search service for discovery.

12.7.2 Next 2–5 years

Current ESnet and Globus services enable high-speed data movement; current ASCR facilities enable high-speed computation. BER facilities provide some ability to discover and access important scientific data, but are far from adequate in terms of their ability to enable the most advanced data-driven science. We attribute these deficiencies to lack of investment and a lack of suitable underlying data architectures.

The next 2–5 years will surely require a new class of science data facilities that can enable large-scale aggregation, organization, and management of data within both individual discovery engines and across different facilities and disciplines. A key to success in establishing such facilities will be to leverage SaaS methods to establish and deliver common services, and to reduce demands on individual projects, facilities, and researchers.

12.7.3 Beyond 5 Years

We hope that DOE, BER, and ASCR will embrace the opportunities inherent in creating the world's most capable data-driven science infrastructure. With all data created within BER research accessible, discoverable, linked, and computable, the pace of discovery in Biological, Environmental, and other research will be greatly enhanced.

12.8 Cloud Services

Globus itself runs on cloud computers provided by Amazon Web Services. We find the high availability, specialized services, and support for state replication across availability zones highly useful for implementing a SaaS capability such as Globus. We believe that DOE should be investigating other opportunities to thus provide "science as a service," i.e., to outsource useful capabilities to cloud computers.

Independently of this use of cloud for SaaS, we see increasing demand for leveraging commercial cloud computing and storage services. For example, Amazon Elastic Compute is gaining traction as a platform for on demand compute resource, leading to increased need for efficient staging of data to and from Amazon storage. Globus introduced support for moving data to Amazon S3 storage, and has seem good adoption from users at researchers at various campuses and some from national labs. Internet2 Net+ offers Amazon Web Services to campuses and has provided direct AWS/Internet2 connectivity to enhance data transfers from campuses to cloud storage. With such seamless access to cloud resources, researchers will continue to leverage that and use that in tandem with other facilities and local resources.

On-demand cloud storage also provides an attractive offering for researchers. For example, LBNL recently adopted Google Drive as a research storage tier with unlimited storage for the campus, and encouraging all campus users to use Google Drive to store their data. There remains a gap in tools that provide reliable secure managed transfers to some of the cloud storage such as Google Drive that needs to be addressed.

12.9 Outstanding Issues

We observe that while high-speed networks are widely deployed across DOE laboratories, the tools that can allow users to make truly effective of those networks are less universally disseminated. ESnet has done an excellent job promulgating the Science DMZ concept; an important next step will be to promote use of tools such as Globus that can drive increased use of Science DMZs and ESnet capabilities.

The important role that Globus services play in BER and other DOE science makes it important to support both their continued operation and the development required to keep up with technological change and new requirements be supported. DOE currently has no coordinated mechanism for providing such support, leaving DOE facilities, projects, and researchers at risk of service interruption.

The Globus team is working towards sustainability via a subscription model, in which individual laboratories and projects pay the Globus group for access to advanced capabilities. At present, several DOE laboratories and facilities subscribe to this program (see Figure 12.1), but many major users do not.

Case Study 13

The DOE-UCAR Cooperative Agreement for Climate Change Prediction Program

13.1 Background

The Cooperative Agreement between DOE and the University Corporation for Atmospheric Research (UCAR) sponsors the development, enhancement, utilization, and analysis of the NCAR, DOE, and NSF Community Earth System Model (CESM)—one of the world's most complete and advanced climate models. CESM has participation from a very large community of scientists, and peer-acceptance, which is important to ensure excellence and relevance. Major modeling programs are no longer single PI research projects, because of the complexity of the problem and the technical sophistication of the models and computer codes. They are major technology development efforts, and are both shared-research tools and major code projects. The CESM community enables access to contributions from multiple sources in an open development process that allows incorporation and testing of a wide range of ideas in a broad spectrum of disciplines. The CESM program also has a mission to foster the creative involvement of university researchers and students in the subject area, and thus contributes to the development of highly trained people for the future. The CESM program is a complement to the other major modeling programs in CCSP that are specifically oriented towards a government mission to provide decision-support information.

Development of a comprehensive CESM that accurately represents the principal components of the climate system and their couplings requires both wide intellectual participation and computing capabilities beyond those available to most U.S. institutions. The CESM, therefore, must include an improved framework for coupling existing and future component models developed at multiple institutions, to permit rapid exploration of alternate formulations. This framework must be amenable to components of varying complexity and at varying resolutions, in accordance with a balance of scientific needs and resource demands. In particular, the CESM must accommodate an active program of simulations and evaluations, using an evolving model to address scientific issues and problems of national and international policy interest.

The CESM project will address important areas of climate system research. In particular, it is aimed at understanding and predicting the climate system. The long-term goals of the CESM project are simple but ambitious. They are:

- to develop and to work continually to improve a comprehensive CESM that is at the forefront of international efforts in modeling the climate system, including the best possible component models coupled together in a balanced, harmonious modeling framework;
- to make the model readily available to, and usable by, the climate research community, and to actively engage the community in the ongoing process of model development;
- to use the CESM to address important scientific questions about the climate system, including global change and interdecadal and interannual variability; and

- to use appropriate versions of the CESM for calculations in support of national and international policy decisions.

Complementary efforts using simplified models are also important and will be undertaken by many individuals, including some CESM participants. However, the CESM project will remain focused on comprehensive climate modeling.

We anticipate many important changes in the climate modeling enterprise over the next five years, including:

- increasing computer power, both in the United States and abroad, that can support more elaborate and more sophisticated models and modeling studies, using increased spatial resolution and covering longer intervals of simulated time;
- improved understanding of many of the component processes represented in the CESM, including cloud physics; radiative transfer; atmospheric chemistry, including aerosol chemistry, boundary-layer processes, polar processes, and biogeochemical processes; and the interactions of gravity waves with the large-scale circulation of the atmosphere;
- improved understanding of how these component processes interact;
- improved numerical methods for the simulation of geophysical fluid dynamics; and
- improved observations of the atmosphere, including major advances in satellite observations.

Under the auspices of the DOE-UCAR CA, CESM simulations are carried out on a number of supercomputers, including the NCAR/University of Wyoming Yellowstone system, and NERSC's Hopper and Edison systems, ANL's Mira system, and others. CESM is utilized for large international MIPs, including the upcoming CMIP6, and similar projects, including large-scale community efforts like the current Large Ensemble (LE) and Last Millennium Ensemble (LME) projects. The total data volume available from the LE and LME together is about 3.5 times (610TB vs. about 170TB) that of what was provided for CMIP5.

Results from these simulations are often transferred between the various computing sites for analysis, depending on specific aspects of the simulations involved. The preferred tool for data transfer is Globus, but if an endpoint does not exist at one site, then other means, scp or bbcp, can be used.

The volume of data transferred from one site to others can vary considerably—from a few hundred megabytes to tens or hundreds of terabytes. Efforts are made to keep model results local to the system upon which they were generated, but that is not always possible, especially in regards to the MIP-related simulations. For example, versions CCSM4 and CESM1.0 were used for the 2010–2012 CMIP5 simulations, and all the data (in total, about 170 TB) were transferred from NERSC and ORNL to NCAR for hosting on the ESGF, the infrastructure for CMIP5 and the other MIPs (TAMIP, GeoMIP, PMIP3 and others) to which the CESM project submitted data.

Data from hundreds of non-MIP simulations and community projects like the LE and LME are hosted via the NCAR Earth System Grid (ESG) portal.

13.2 Local Science Drivers

13.2.1 Instruments and Facilities

Table 13.1 shows the current instruments, facilities, and resources for the DOE-UCAR CA.

Over the next 2-5 years, it is expected that the Yellowstone supercomputer at the NCAR-Wyoming Supercomputing Center (NWSC) will be upgraded in terms of processor cores, memory, disk storage, and the other resources.

Table 13.1: This table describes the local computing and other resources the DOE-UCAR CA uses for carrying out simulations with the CESM, as of mid-2015.

Site name	Name and type	Processors	Memory	Disk storage	Archival storage capacity
NCAR-Wyoming Supercomputing Center (NWSC)	Yellowstone IBM iDataPlex	72,576 Intel Sandy Bridge processors	144.6 TB	15 PB	> 160 PB

13.3 Remote Science Drivers

This table describes the remote computing and other resources the DOE-UCAR CA uses for carrying out simulations with the CESM, as of mid-2015.

Site name	Name and type	Processors	Memory	Disk storage	Archival storage capacity
National Energy Research Scientific Computing Center (NERSC)	Hopper Cray XE6	153,216 Opteron	212 TB	2 PB	240 PB
National Energy Research Scientific Computing Center (NERSC)	Edison Cray XC30	133,824 Intel Ivy Bridge	357 TB	8 PB	240 PB
Argonne Leadership Computing Facility (ALCF)	Mira IBM Blue Gene/Q	786,432	768 TB	27 PB	16 PB

Over the next 2–5 years, the supercomputers at each of the remote computing sites (NERSC and ANL) will be upgraded in terms of processor cores, memory, disk storage, and the other resources. The new NERSC HPC resource, Cori, the first phase, is expected to become available in late fall 2015. The new ANL HPC resource, Aurora, should begin production-level service in Q2 CY2019.

Just as with the NWSC resource, it is anticipated that the computing resources at the DOE sites will be utilized to carry out the simulations with CESM in accordance with the DOE-UCAR CA plans, as well as the CMIP6 project.

13.4 Process of Science

The typical process for the use by the DOE-UCAR CA of the CESM for knowledge discovery involves an experimental design created by either an individual scientist, small NCAR group of scientists, or one of the CESM Working Groups (a collection of scientists and others with a common interest). Once the design is finalized and the necessary resources (computing, storage, and so on) are determined, the project applies for those resources at the computing center. Once those resources are allocated, then the model is executed at the center, the output is analyzed and archived, made available via the ESG or ESGF as appropriate, and papers are written and submitted to various science journals detailing what was learned from the experiments.

The same basic process for CESM simulations executed at NCAR is accomplished at DOE computing centers located elsewhere. Experiments are designed, resources allocated, the simulations run, post-processed, analyzed and made available via the ESGF and ESG. Some of the original model output and post-processed data from these simulations is transferred back to NCAR, but because all of the DOE-UCAR CA computing resources are associated with nodes in the ESGF, it is not necessary to transfer all of the data just for the purpose of making them publicly available. To get an idea of the total data volume generated by thousands of CESM simulations, the graph below

uses the archival volumes at NCAR and the DOE sites to extrapolate CESM data holdings for period 2016–2025, using the 2005–2014 period for extrapolation:

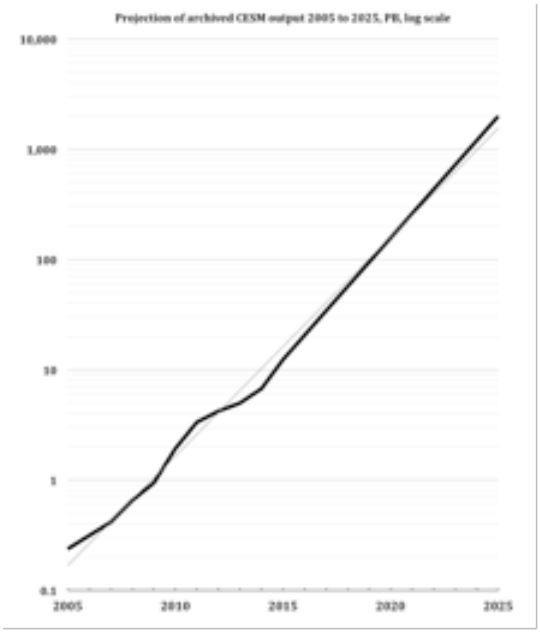


Figure 13.1: Archival volume of CESM output at NCAR and DOE HPC sites.

The science processes regarding the DOE-UCAR CA's use of the CESM are anticipated to be very similar over the 2017–2020 time range as they are currently, with the possible exception that the model output will be written in a transposed format compared to the current history format. This shift, from putting all fields from a single time period in a single file, to writing all time periods for each individual model output field into a single file, will reduce the requirement to post-process the model output to make it more usable for the community. Work is nearly complete on this aspect of workflow re-engineering. This enhancement of CESM will be part of the release of CESM2, expected in mid-2016.

One key project that is will begin and continue during this period is the anticipated CMIP6. The specifics of the scope of this project, number of simulations, experiment types, and output requirements—by the WCRP's WGCM—should be available by mid-2016.

It may be the case that the current ESGF architecture will be enhanced and expanded over this period, so that any CESM CMIP6 simulations will remain resident at their host sites, without the need to transfer large volumes to model data to NCAR or between the sites.

13.5 Beyond 5 years

See Table 13.2.

13.6 Network and Data Architecture

The CESM project as a whole may participate in future Big Data initiatives, but has not until this time.[1]

[1]The current CESM Data Management and Data Distribution Plan is available at `http://www.cesm.ucar.edu/management/docs/data.mgt.plan.2011.pdf`.

13.7 Collaboration Tools

The weekly meeting of the DOE-UCAR CA team uses Readytalk's services for remote collaborator call-ins, as well as sharing the desktop of the meeting convener. Skype is used on occasion to collaborate with colleagues located at remote locations. It is not anticipated that these practices will change.

13.8 Data, Workflow, Middleware Tools and Services

The most significant change to current DOE-UCAR CA practices will be the nearly complete re-engineering of the CESM workflow, to enable the creation of single-field timeseries format data as the simulation is ongoing. This will enable the global user community to have easier and more efficient access to CESM results. The DOE-UCAR CA and CESM will continue to rely on Globus, the ESGF and ESG and their follow-on projects to publish and deliver model output to the user community. Other projects may be incorporated into the ESGF to enable data format changes (to other binary formats, from netCDF to GIS-compatible formats, for example) and the ability to extract, subset, and additionally process the model results. Whatever tools ESG and/or ESGF make available will be exploited by the DOE-UCAR CA.

Table 13.2: The following table summarizes data needs and networking requirements for the DOE-UCAR CA.

Key Science Drivers			Anticipated Network Needs	
Instruments, Software, and Facilities	**Process of Science**	**Data Set Size**	**Local-Area Transfer Time**	**Wide-Area Transfer Time**
0-2 years				
NCAR-Wyoming Supercomputing Center (NWSC), hosting yellowstone National Energy Research Supercomputing Center hosting hopper and edison Argonne National Labs hosting mira	Experimental design created Necessary resources applied for and allocated Simulation completed	File size varies from 10 MB to 200 GB Total volume for simulation ranges from 10s GB to 10s TB	Better than 1 Gbit/s	Sustainable rate via Globus Online of about 1-10 Gbit/s
CESM version 1	Postprocessing and diagnostics completed		Highly "bursty" depending on scientific need	Highly "bursty" depending on scientific need
Globus Online for data transfer	Papers written	netCDF files of size 100s MB to 100s GB, number 100-10000 per simulation		Globus Online endpoint in CISL
ESG and ESGF for data distribution	Data published into ESG or ESGF			
2-5 years				
Upgrade to yellowstone Upgrade at NERSC Upgrade at ANL	Workflow re-engineering completed and model writing timeseries format natively Autopublishing of model output to ESG/ESGF	File size varies from 100 MB to 200 GB Total volume for simulation ranges from 10s GB to 100s TB	Better than 10 Gbit/s	Sustainable rate via Globus Online of about 10-20 Gbit/s
CESM version 2		netCDF files of size 1s GB to 1s TB	Highly "bursty" depending on scientific need	Highly "bursty" depending on scientific need
CMIP6		Number 100-10000 per simulation		Globus Online endpoint
5+ years				
Upgrade to yellowstone Upgrade at NERSC ? Upgrade at ANL ?	· What is the strategic direction for data flow, science process, etc.?	File size varies from 10s GB to 10s TB Total volume for simulation ranges from 100s GB to 100s TB	Better than 50 Gbit/s	Sustainable rate via Globus Online of about 50-100 Gbit/s
CESM version 2+		netCDF files of size 10s GB to 10s TB	Highly "bursty" depending on scientific need	Highly "bursty" depending on scientific need
CMIP7?		Number 100-10000 per simulation		Globus Online endpoint

References

[1] Bryce Allen et al. “Software as a Service for Data Scientists”. In: *Communications of the ACM* 55.2 (2012), pp. 81–88.

[2] K. Chard, S. Tuecke, and I. Foster. “Efficient and Secure Transfer, Synchronization, and Sharing of Big Data”. In: *Cloud Computing, IEEE* 1.3 (Sept. 2014), pp. 46–55. ISSN: 2325-6095. DOI: 10.1109/MCC.2014.52.

[3] Kyle Chard et al. “Globus Data Publication as a Service: Lowering Barriers to Reproducible Science”. In: *11th IEEE International Conference on eScience*. Munich, Germany, 2015.

[4] Eli Dart et al. *An Assessment of Data Transfer Performance for Large-Scale Climate Data Analysis and Implications for the Design of CMIP6*. 2015.

[5] Ian Foster et al. “Networking materials data: Accelerating discovery at an experimental facility”. In: *Big Data and High Performance Computing*. Ed. by Gerhard Joubert and Lucio Grandinetti. In press, 2015.

[6] Prabhat et al. *TECA: Petascale Pattern Recognition for Climate Science*. 2015.

[7] D N Williams et al. “The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data”. In: *Bulletin of the American Meteorological Society* 90.2 (2009), pp. 195–205.