ORNL/LTR-2019/1075

Background and Roadmap for a Distributed Computing and Data Ecosystem

Contributors:

Co-Chair: Eric Lancon, BNL
Co-Chair: Arjun Shankar, ORNL

Members:
Ray Bair, ANL
Amber Boehnlein, JLab
David Cawley, PNNL
Eli Dart, ESnet
Michael Hofmockel, PNNL
Amedeo Perrazzo, SLAC
Rob Roser, FNAL
Lauren Rotman ESnet
Panagiotis Spentzouris, FNAL
Adam Stone, LBNL
Craig Tull, LBNL
Kerstin Kleese Van Dam, BNL
Theresa Windus, AMES

April 2019

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE US DEPARTMENT OF ENERGY

ORNL/LTR-2019/1075

National Center for Computational Sciences

BACKGROUND AND ROADMAP FOR A DISTRIBUTED COMPUTING AND DATA ECOSYSTEM

Author(s)

Co-Chair: Eric Lancon, BNL
Co-Chair: Arjun Shankar, ORNL

Members:
Ray Bair, ANL
Amber Boehnlein, JLab
David Cawley, PNNL
Eli Dart, ESnet
Michael Hofmockel, PNNL
Amedeo Perrazzo, SLAC
Rob Roser, FNAL
Lauren Rotman ESnet
Panagiotis Spentzouris, FNAL
Adam Stone, LBNL
Craig Tull, LBNL
Kerstin Kleese Van Dam, BNL
Theresa Windus, AMES

Date Published:
April 2019

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831-6283
managed by
UT-BATTELLE, LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

ACRONYMS

ANL	Argonne National Laboratory
ARM	Atmospheric Radiation Measurement
ASCR	Advanced Scientific Computing Research
BER	DOE Office of Biological and Environmental Research
BES	DOE Office of Basic Energy Sciences
BNL	Brookhaven National Laboratory
BSSD	Biological Systems Science Division
CADES	Compute and Data Environment for Science
CESD	Climate and Environmental Science Division
DCDE	Distributed Computing and Data Ecosystem
DMZ	De-Militarized Zone
DOE	US Department of Energy
DTN	Data Transfer Nodes
DUNE	Deep Underground Neutrino Experiment
EMSL	Environmental Molecular Sciences Laboratory
EOD	Experimental and Observational Data
ESGF	Earth System Grid Federation
ESnet	Energy Sciences Network
FLC-WG	Future Laboratory Computing Working Group
FNAL	Fermi National Laboratory
FTP	File Transfer Protocol
JLab	Jefferson Laboratory
HEP	DOE Office of High Energy Physics
HPC	High Performance Computing
HTTP	Hyper Text Transfer Protocol
IdM	Identity Management
iRODS	Integrated Rule-Oriented Data System
KBASE	Systems Biology Knowledgebase
LBNL	Lawrence Berkeley National Laboratory
LCF	Leadership Class Facilities
LCRC	Laboratory Computing Resource Center
LDRD	Laboratory Directed Research and Development
MRDP	Modern Research Data Portal
ERSC	National Energy Research Scientific Computing Center
NLRCG	National Laboratory Research Computing Group
NP	DOE Office of Nuclear Physics
OLCF	Oak Ridge Leadership Computing Facility
ORCID	A nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors
ORNL	Oak Ridge National Laboratory
OSG	Open Science Grid
PB	Petabytes
PHI	Personal Health Information
PKI	Public Key Infrastructure
RP	Resource Provider
SC	DOE Office of Science

SLAC	Stanford Linear Accelerator
SMEs	Subject Matter Experts
VO	Virtual Organization
WLCG	Worldwide Large Hadron Collider Computing Grid
XSEDE	Extreme Science and Engineering Discovery Environment

EXECUTIVE SUMMARY

The drivers for using distributed computing and data resources are emerging from many science domains and have been documented through several recent Department of Energy (DOE) workshop reports. The increased usage and availability of geographically distributed computing and storage resources also highlight the need for easy-to-use tools and procedures that can support the paradigm of distributed computing and data management transparently. Aiming to meet each laboratory's programmatic needs, a variety of computing resources are available throughout the DOE complex under various access policies, operating systems, and hardware configurations. However, today no straightforward mechanisms or policies exist for a given researcher or group of researchers to easily access resources at different locations across the DOE laboratory system. At the same time, DOE research programs have supported the research and development of technologies and tools that could enable researchers to access distributed data and computing resources globally. Academic and industry activities have also developed mechanisms for the distributed federation of resources. We observe that such existing capabilities can be harnessed to create mechanisms that enable researchers to work seamlessly across facilities.

We envision the creation of a DOE Office of Science (SC) wide federated Distributed Computing and Data Ecosystem (DCDE) which comprises tools, capabilities, services and governance policies to enable researchers to seamlessly use a large variety of resources (i.e., scientific instruments, local clusters, large facilities, storage, enabling systems software, and networks) end-to-end across laboratories within the DOE environment. A successful DCDE would present to the researchers a range of distributed resources through a coherent and simple set of interfaces, and allow them to establish and manage computational pipelines and the related data lifecycle. Envisioned as a cross-laboratory environment, a DCDE would be overseen by a governing body that includes the relevant stakeholders to create effective use and participation guidelines.

To validate the DCDE approach, the working group proposed the development of a prototype which implements in a coherent and progressive manner the main components of a DCDE. The prototype will help in defining a general set of recommendations, supported by implementation experiences, for expanding the DCDE to the broader laboratory complex and produce an applicable governance model.

1. INTRODUCTION

The National Laboratory Research Computing Group (NLRCG) and the Advanced Scientific Computing Research (ASCR) program office jointly established the Future Laboratory Computing Working Group (FLC-WG) with a charter to identify the benefits and obstacles in creating and operating a DOE/SC wide federated Distributed Computing and Data Ecosystem (DCDE). The FLC-WG organized a series of topical presentations from researchers, facility leaders, and scientists to understand the community needs and current practice. The objective is to take advantage of past research results in this area to create a DCDE with as little reinvention as possible.

We define a DCDE for DOE as the set of tools, capabilities, services, and governance policies that enable researchers to seamlessly use a variety of computing-related resources (i.e., scientific instruments, local clusters, large facilities, storage, enabling systems software, and networks) end-to-end across laboratories within the DOE environment.

The methodology adopted by the FLC-WG included:

1. Review past ASCR research activities and current laboratory practices.

2. Review scientific community and academic approaches, and commercial solutions.

3. Identify current and future challenges that need to be addressed.

4. Compose a report describing the current state of lab computing, a set of research challenges, and recommendations for achieving a federated DCDE.

This report includes the findings of the working group and reflects the summarized outcome of the FLC-WG charter activities stated above. We observed that DOE has supported several research activities over the years that are vital components of a DCDE. In addition, DOE supports world-leading facilities in a host of scientific areas. Although SC laboratories provide computing and storage resources to lab research staff and visiting scientists, demands on these resources are dramatically increasing, requiring researchers to reach across laboratories to accomplish their scientific objectives. We believe that the laboratories have the capability to leverage decades of research to create a modern DCDE to actively address the growing demands of DOE scientists.

Enabling a DCDE requires a set of functional services and capabilities to come together, which will enable a coherent, easy-to-use environment that makes it straightforward for individual investigators to make rapid progress in advancing their science. The main patterns of functional services and capabilities that emerged from the review of the capability and technology landscape and needs include: (i) seamless authorization and access control across facilities, (ii) coordinated resource allocation and cross-facility workflows, (iii) enhanced data storage, transfer, and dissemination capabilities, (iv) support for functional variety and portability across environments, and (v) well-understood governance structures and policies to ensure equitable and transparent operations.

The rest of the report is structured as follows: in Section 2, we summarize the science drivers which a DCDE will need to support. We begin by outlining representative science use cases that have been presented to the working group and refer to several prior documents that describe the requirements of the science community. In Section 3, we briefly list examples of institutional facilities, as well as the primary DOE ASCR facilities. Next, in Section 4, we distill the critical techniques from the state-of-the art in research and existing practices required to create the middleware and components of a DCDE. Section 5 addresses the important topic of organizational concerns and the governance of a DCDE. Concluding the report, Section 6 discusses next steps and proposes a pilot implementation for the DCDE.

2. SCIENCE DRIVERS

The emergence of unprecedented needs for computing and storage from current facilities and science programs has been described in several recent workshop reports [3,5,14]. Managing computing at larger scales, increasing complexity and variety, and processing exponentially growing volumes of data are becoming common concerns across the various DOE/SC program offices.

2.1 GROWTH OF DATA AND A MULTIFACETED SCIENTIFIC DISCOVERY LANDSCAPE

The 2015 DOE workshop on Management, Visualization, and Analysis of Experimental and Observational Data [18] recognized these challenges in their final report. The typical workflow identified in the report involves first collecting data at the instrument, performing some processing close to the

instrument before moving data to local or remote facilities for more lengthy calculations and preparation of data products, and then disseminating data products. The way each project implements this pattern varies according to their needs and available resources, but most use cases offer variants of data handling and processing activities that can be characterized as distributed computing models. The report acknowledges that:

> *"Meeting the challenges of the explosion of data from EOD projects requires computational platforms, networking, and storage of greater capacity and lower latency, along with software infrastructure suited to their needs. However, existing HPC platforms and their software tools are designed and provisioned for high-concurrency HPC workloads, single-project data products, and comparatively simpler data needs. The result is a significant gap between the needs of EOD projects and the current state of the art in computational and software capabilities and resources. [...]"*

Several notable science experiments exemplify these needs. For example, the Deep Underground Neutrino Experiment (DUNE) experiment is targeting the design and development of a project-wide software infrastructure that aims to maintain portable and accessible software that can be used at any institution and run transparently on modern grid and/or cloud resources as part of a distributed processing data-centric workflow. Another example is the major Linac Coherent Light Source (LCLS) upgrade planned in the 2020 timescale, which will require a new paradigm in how data is acquired, managed, and analyzed. A suggested approach involves a design that provides both access to local capabilities for storage and analysis of standard experiments, and the ability to surge to remote capabilities for the highest demand experiments, or whenever the local capabilities are fully utilized. Projections suggest, when integrated across the entire DOE/SC complex, that user facilities will soon collectively acquire exabytes of data per year [18].

The sociological landscape of scientific discovery is also evolving, with more diverse collaborations exploiting geographically distributed resources. In some cases, DOE science is carried out by teams of scientists with limited connections to computer scientists, or with expertise in computing limited in scope to their specific needs. While the growth of data is one dimension of the growing problem, there is also a greater need for multiple analysis modes (e.g., real-time, near real-time, twinning, steering, and in situ, etc.). In addition, certain domain communities build capabilities they depend on that are unavailable to other communities due to a lack of expertise or awareness; these domain communities also find it hard to adapt and accept new ways of accelerating scientific discovery. Ultimately, the problems of data growth, multiple analysis modes, heterogeneous resources available and needed, and the diversity of extant tools and skills must be pulled together into a framework that is not a hindrance but instead is an enabler for scientific productivity.

2.2 DOE/SC PROGRAM OFFICE NEEDS

Each of the program offices in the DOE/SC supports domain scientists and facilities that reflect the needs and requirements we discuss above.

The Office of Basic Energy Sciences (BES) within SC [4] supports multiple user facilities that house more than 240 different instrument types, ranging from X-ray light sources to neutron scattering facilities and nanoscale science research centers. A recent report to examine BES computational needs [5] identified the following as one of its key findings:

> *"BES and the ASCR facilities are experiencing a pressing need to mature their capabilities in data science. Improvements and new capabilities at BES facilities are creating challenges that the community is not prepared to address. These include unprecedented growth in data volume,*

complexity, and access requirements; the need for curation of the massive amounts of data that are retained; and integration of diverse datasets from different experiments to enable new scientific conclusions. Efficient and effective use of BES facilities requires real-time access to ASCR HPC facility-class resources to support streaming analysis and visualization to guide experimental decisions."

Multiple BES priority research directions were identified in quantum materials and chemistry, catalysis, photosynthesis, light harvesting, combustion, materials and chemical discovery, and soft matter. All these areas require new mathematics and computer science models to facilitate the transformative science of BES. Of particular relevance was the need for multiple levels of theoretical models to enable multiscale science and complex materials and chemistry phenomena. In addition to the need for very large computational resources, the need for midscale computing and the ability to manage the very large data sets from these simulations was a vital need.

DOE SC's Office of Biological and Environmental Research (BER) consists of two divisions: Biological Systems Science Division (BSSD), and Climate and Environmental Science Division (CESD). Each division has its own science use cases that drive their own needs for different kinds of computational infrastructure, yet both divisions have common needs that also overlap with other SC programs. BER partners with ASCR to use computational resources. For example, BER sponsors the DOE Systems Biology Knowledgebase (KBASE) to provide computational workflows and data management for systems biology. BER also provides a high performance computational and data capability of its own in the Atmospheric Radiation Measurement (ARM) and Environmental Molecular Sciences Laboratory (EMSL) user facilities. In order to advance the predictive understanding of earth's climate and environmental systems, CESD has expressed a need for computational infrastructure to integrate "*multiscale observations, experiments, theory, and process understanding into predictive models for knowledge discovery*" [20]. CESD describes the need to:

"advance the predictive understanding of Earth's climate and environmental systems… the innovation most needed is a framework that allows seamless integration of multiscale observations, experiments, theory, and process understanding into predictive models for knowledge discovery… processing should be *automated and robustly integrated, and they must support computational environments from small clusters to the leadership class facilities (LCFs) with one - time authentication (single sign - on). Further integration with model output archives and delivery systems, such as the Earth System Grid Federation (ESGF), is expected to be important for combining observations and model results to answer various research questions and support model benchmarking [20]."*

Similarly, the BSSD computational needs are diverse and rapidly evolving, driven both by the need to understand experimental results and to simulate complex, natural systems. The BSSD Strategic Plan states:

"A key counterpart to BSSD interdisciplinary research is the development of new approaches to efficiently combine high-performance computational techniques with experimental analyses to enable iterative hypothesis-based research on whole biological systems. [8]"

The Office of Nuclear Physics' (NP) scientific program, outlined in the Nuclear Physics Long Range Plan [32], expresses a requirement for a tight coupling and feedback between theory calculations, simulated data, and experimental data from multiple experiments. The process of developing the necessary infrastructure and scientific applications is getting underway and is likely to reach full maturity for the proposed Electron-Ion Collider. Like what is outlined for DUNE, this program will need to utilize a

heterogeneous set of resources, which opens the possibility for processing frameworks in which each sub-task is optimally matched to its compute hardware. The NP Long Range Plan states:

> *"Nuclear physics is poised for a period of new discoveries made possible by the upgrade and commissioning of state-of-the-art experimental and observational facilities, remarkable advances in high-performance computing, and the relentless pursuit of transformational ideas. A strong interplay between theoretical research, experiment, and advanced computing is essential for realizing the* full *potential of these discoveries. New measurements drive new theoretical and computational efforts which, in turn, uncover new puzzles that trigger new experiments."*

The Office of High-Energy Physics' (HEP) Exascale Requirements Review [3] draws several main conclusions that echo the needs listed above, including requiring:

> *"(1) Larger, more capable computing and data facilities are needed to support HEP science goals in all three frontiers: Energy, Intensity, and Cosmic. The expected scale of the demand at the 2025 timescale is at least two orders of magnitude -- and in some cases greater -- than that available currently. (2) The growth rate of data produced by simulations is overwhelming the current ability, of both facilities and researchers, to store and analyze it. Additional resources and new techniques for data analysis are urgently needed. (3) Data rates and volumes from HEP experimental facilities are also straining the ability to store and analyze large and complex data volumes. Appropriately configured leadership-class facilities can play a transformational role in enabling scientific discovery from these datasets. (4) A close integration of HPC simulation and data analysis will aid greatly in interpreting results from HEP experiments. Such an integration will minimize data movement and facilitate interdependent workflows."*

The Office of Fusion Energy makes similar observations:

> *"The technical implementations for practical and affordable exascale platforms will present a number of significant challenges to approaches and algorithms used in today's codes. Additional challenges are presented in the areas of fault tolerance, software engineering, workflows, data management, in-situ analytics, and visualization. Close collaboration among stakeholders in various communities will be crucial to overcoming these challenges and realizing the advantages afforded by the new platforms."*

The Exascale requirements review from across the program offices in the DOE Office of Science is available online [21] and includes a cross-cut review [14] that raises commonly growing concerns of computing at various scales, data analytics integrated with simulation, and workflows to support the applications across facilities. This need again suggests the vital requirement for a DCDE. As we note above, common themes and needs emerge from the different science domains: (1) an ability for users to seamlessly execute cross-facility workflows, (2) easy resource allocation and cross-facility coordination mechanisms, (3) high-speed data transfers (including streaming) and mechanisms to effectively disseminate data and hold data for the long term, and (4) capabilities to enable functional variety and portability across the laboratories. Additional computing resources will be required, but considering tightening budgetary environments, a DCDE that supports the above requirements and facilitates sharing of best-practices is a logical solution for communities and facilities to come together to take advantage of a broader and unified view across DOE facilities.

3. COMPUTING FACILITIES OVERVIEW

Computing facilities with associated computing and data management infrastructure are ubiquitous at the DOE national laboratories. This includes a wide range of capabilities, from department and project servers and databases, to world-leading supercomputers and data stores approaching exabytes. Among them, the capabilities of ASCR's national scientific user facilities are well-documented. Some of DOE's experimental user facilities also have dedicated computing resources available to approved facility users. In addition to these national facilities, each of the Office of Science and National Nuclear Security Administration labs also have institutional research computing capabilities geared to their particular missions and their programs.

Institutional research computing facilities collocate computing and data resources and centralized system administration to reduce support and labor cost, streamline aggregate facility requirements, and facilitate good cybersecurity. The site funding (business) models, facility governance, levels of system and user support, and modes of integration of institutional computing with lab programs and user facilities vary across laboratories. Although the missions and research programs of the DOE labs vary, the creation of services at the institutional and major facilities that could be exploited more broadly, serves as the basis for the DCDE. For completeness, we briefly discuss the significant computing and data facilities in the laboratory environment.

The next two sections review the major ASCR computing resources and examples of laboratory institutional computing and data resources.

3.1 DOE COMPUTATIONAL AND DATA FACILITY CAPABILITIES

ASCR's mission is to develop and deploy high performance and leadership computing resources and high-performance networking for scientific discovery. ASCR facilities include high end computers at ANL, ORNL, and LBNL, and a High-Performance Scientific Network (ESnet).

The Office of Science sponsors 27 national scientific user facilities[1] to provide researchers with advanced tools to conduct their science. Many of them can be divided into two broad categories: computing facilities, and facilities for conducting observational science (accelerators, colliders, light sources, neutron sources, and facilities for studying the environment and atmosphere). User facilities that provide significant HPC and data transfer resources are summarized below:

The **Argonne Leadership Computing Facility (ALCF)** provides supercomputing capabilities to the scientific and engineering community to advance fundamental discovery and understanding in a broad range of disciplines. Available to researchers from universities, industry, and government agencies, the ALCF is a DOE Office of Science User Facility that helps accelerate the pace of discovery and innovation by providing supercomputing resources that are 10 to 100 times more powerful than systems typically used for scientific research. Through substantial awards of supercomputing time and user support services, the ALCF enables large-scale modeling and simulation research aimed at solving some of the world's largest and most complex problems in science and engineering.

The **Oak Ridge Leadership Computing Facility (OLCF)** is an Office of Science user facility that — along with its partner organization, the ALCF — offers researchers leadership capability: computing and data analysis resources many times more powerful than they could access elsewhere. Researchers can apply for time at leadership computing facilities through several allocation programs that cater to a range

[1] https://science.energy.gov/user-facilities/

of scientific disciplines and HPC experience levels. Every year, the OLCF hosts and works with several hundred users across a broad range of scientific domains at industry events, on-site meetings, and regular hackathons. The OLCF is operationalizing the Summit computer in 2019 - a pre-exascale system capable of 5 to 10 times the performance of Titan, the previous leadership computer at OLCF. (Summit is the fastest supercomputer in the world at the time of the writing of this report.)

The **National Energy Research Scientific Computing Center (NERSC)** is the broad-based scientific computing facility for the Office of Science in the U.S. Department of Energy. As one of the largest facilities in the world devoted to providing computational resources and expertise for basic scientific research, NERSC is a world leader in accelerating scientific discovery through computation. NERSC is a division of Lawrence Berkeley National Laboratory (LBNL), located in Berkeley, California. It is also one of three divisions in the Berkeley Lab Computing Sciences area, housed in the Shyh Wang Hall computational research and theory facility.

ESnet is DOE's user facility network, engineered and optimized for big data science, connecting all major DOE research facilities in the U.S. and abroad, and peering with other networks to enable seamless, high performance data transfer. It is a major enabling capability for integration of experimental and computational facilities and data across the DOE complex and beyond. ESnet provides the high-bandwidth, reliable connections that link scientists at national laboratories, universities, and other research institutions, enabling them to collaborate on some of the world's most important scientific challenges, including energy, climate science, and the origins of the universe. Funded by the DOE Office of Science, ESnet is managed and operated by the Scientific Networking Division at LBNL. As a nationwide infrastructure and DOE User Facility, ESnet provides scientists with access to unique DOE research facilities and computing resources.

The **Environmental Molecular Sciences Laboratory (EMSL)** at PNNL began operations in 1997, and currently provides the scientific user community with a broad range of premier instruments for molecular to mesoscale research, as well as production HPC and optimized computational codes for molecular to continuum-scale modeling and simulation. With more than 50 premier instruments, individual non-proprietary and proprietary users and user teams from academia, national laboratories, other federal agencies, and industry can use multiple capabilities and iterate between experiment and simulation to obtain a mechanistic understanding of physical, chemical, and biological processes and interactions that underpin larger-scale biological, environmental, climate, and energy challenges.

The Atmospheric Radiation Measurement (ARM) Climate Research Facility began operations in 1989 and supports a global network of permanent and mobile long-term atmospheric observational facilities, as well as the development of scientific data products. ARM is a multi-laboratory effort and is a major contributor to national and international research efforts related to global climate change.

3.2 INSTITUTIONAL COMPUTING EXAMPLES

The following are typical examples of institutional computing at four DOE science laboratories. They show the breadth and depth of computing and data infrastructures that are available to laboratory staff.

- **Argonne National Laboratory's (ANL)** Laboratory Computing Resource Center [LCRC][2] provides mid-range institutional high-performance computing (HPC) and data resources for laboratory research projects, provides user assistance and training, making science and engineering projects more successful and productive, and provides a spectrum of scalable applications and tools while training people to use them, enabling larger and more complex

studies. Every ANL research division uses LCRC. LCRC also promotes datacenter efficiency and consolidation, e.g., via hosting programmatic additions dedicated to specific projects.

- **Brookhaven National Laboratory's (BNL)** BNL Institutional Computing [BIC] has historically been designed to support NP and HEP programs [7]. Today, about 60,000 cores, 45 petabytes (PB) of disk, and about 100 PB of tape archive are dedicated to these domains. Institutional computing has ramped up in the past years, mainly with two institutional HPC clusters (CPU-GPU 256 nodes, KNL 144 nodes). About 1,600 users from 20 different projects have an account on BNL scientific computing resources.

- **Oak Ridge National Laboratory's (ORNL)** ORNL's Compute and Data Environment for Science [CADES] provides institutional computing and a big-data facility for ORNL staff and collaborators, and their programs' scalable computing and data analytics needs[11]. CADES supports three protection zones (i.e., Open, Moderate, Protected Health Information (PHI) zones), heterogeneous computing environments including HPC clusters, self-service birthright cloud, and purpose built (e.g., graph, quantum, AI) analytics appliances, all backed by high-speed storage and high-bandwidth networks. Operating adjacent to the Oak Ridge Leadership Computing Facility (OLCF), CADES enables staff to establish cross-facility workflows (e.g., from the Spallation Neutron Source (SNS) to OLCF) and facilitates an on-ramp to leadership computing.

- **Pacific Northwest National Laboratory 's (PNNL)** PNNL Institutional Computing [PIC] is a program designed to advance scientific discovery by providing staff with the resources required to accelerate their research outcomes [39]. The three main objectives of Institutional Computing are to remove barriers to innovation via access to compute capabilities, and build computational capabilities and a culture of computing, while driving efficiencies and economies of scale. These are achieved by offering scientists and engineers access to a broad portfolio of no-cost and cost-effective computational capabilities. Except for Laboratory Directed Research and Development (LDRD), projects consume PIC resources through a pay-by-use model that benefits from institutional funding of the base capability to lower costs.

3.3 ENABLING COMPUTING AND DATA ACROSS THE LAB ECOSYSTEM

Laboratories across the DOE complex support sizable institutional, leadership, and programmatic computing and network facilities. However, the capability at one laboratory may still not be sufficient for the computing needs of its local researchers. Reaching across seamlessly as a distributed system to available capacity and uncommitted resources in other laboratories is a potential solution. Furthermore, the laboratories' computing and data facilities today are already connected by ESnet and will be increasingly tied to distributed and non-local observational science instruments. Linking these resources together into a DCDE could answer the diverse needs and increasing computing and data requirements of DOE/SC researchers.

Several examples of projects federating resources exist that have been supported, piloted, and offered by programs applicable to certain primary domains. Examples include ESGF [19] and Open Science Grid (OSG) [35]. These projects were developed primarily in response to a computing challenge from a given research community. In the same way, the expertise that has been acquired and developed in various areas of computing through ASCR research support can be brought to bear to the broader community in the form of the DCDE. There are also past R&D initiatives supported by the National Science Foundation (NSF) cyberinfrastructure programs such as TeraGrid [44] and the Extreme Science and Engineering Discovery Environment (XSEDE) [49] which have had similar goals and produced valuable lessons and advances from which the DCDE should borrow.

Significant differences exist between the missions of user and institutional computing facilities, which inform the respective roles of the two types of facilities within the DCDE.

User facilities are built around unique capabilities, examples being unique instruments or experimental capabilities, capacity or capability computing facilities, or a distinctive ensemble of both. In the national laboratory context, they are likely to be major producers and/or consumers of scientific data. Their mission is to serve their research communities, which are often distributed across institutions and borders.

Institutional computing capabilities, on the other hand, exist primarily to serve the computing needs of staff at that institution (i.e., one national laboratory). Depending on the individual institution, that mission may extend to collaborators at other institutions. Institutional computing capabilities often face the challenge of satisfying increasing, sometimes 'peaky' requests for diverse resources within constrained budgets. These requests may not be aligned with available local technical expertise nor with multiyear hardware refresh plans.

Viewed in this context, the DCDE is clearly beneficial to the user facilities. It improves their ability to execute their mission by giving their widely-distributed user base better tools to use data and compute resources, which may themselves be widely distributed. The DCDE facilitates collaboration and knowledge sharing. The DCDE also enables offloading temporary excess of local demands and provides scientific communities with specialized hardware and technical support not available at their home laboratory.

Several labs allow for usage of hosted resources to non-local users. This happens because labs are either part of a distributed organization (e.g., OSG) or following agreements. A few examples are listed below:

- **BNL**'s programmatic resources (RHIC, ATLAS, Belle II) are available to affiliated researchers (from NSF, DOE and foreign entities) either through local access (ssh) or through grid workflow managers (PanDA, Dirac) and data management services (Rucio, GridFTP) used by the various projects. Users need to be registered on site to get a local account. X509 certificates are used to grant remote access to grid enabled services. Opportunistic usage of computing resources is available for OSG members at a level of a few percent reaching tens of percent for some periods. Institutional resources, including storage and archival, are made available to both local and non-local projects or organizations via allocations, or with agreements documenting resources and associated cost and level of services.

- **Fermilab** resources are available to CERN's Compact Muon Solenoid (CMS) Experiment researchers include the Tier-1 and LPC processing and storage facilities for production and analysis, respectively. The Tier-1 compute farm accepts opportunistic work from the OSG. Members of Fermi lab-based and affiliated experiments other than CMS utilize the shared FermiGrid processing and storage resources, which are also open to opportunistic use by the OSG. The archival tape facility is also available to external scientific customers, through individual agreements and for a corresponding fee. The facility support and operations processes, personnel, and tools are integrated across all supported research programs. In addition, to support the needs of both the dominant CMS the additional large number of smaller experiments, Fermilab has adopted a strategy that relies on common services – both among the experiments and across the wider HEP realm [48]. These services utilize tools for authentication and authorization (for example with X509 certificates), workflow and job management (HTCondor and GlideinWMS), and data management and transfer (xrootd, globus toolkit for all, phedEx and CERN FTS for CMS, and the FIFE toolset - Fermilab FTS, IFDH, and SAM – for the other experiments). The strategy for the future of these services is to adopt, adapt, and evolve in

conjunction with HEP and wider scientific efforts. These directions include exploration of Rucio as the common data management tool and developing and deploying HEPCloud as the common portal for accessing a wide variety of heterogeneous computing resources. HEPCloud, which is expected to be commissioned at the end of calendar 2018, will provide elasticity in resource provisioning through the Fermilab facility.

- **ORNL** resources in CADES are made available to collaborators in the open research zone through the single-factor sign up zone (LDAP implemented by ORNL XCAMS) that provides authentication and access control. Collaborators and staff in projects can self-request and set up cloud virtual machines, use opportunistically compute time on the HPC cluster queues, or be included in a program sponsored resource. These open protection zones allow workflows to be set up to enable cross-facility automation. Excess capacity is shared on a first come, first served basis, and by requested discretionary allocations. Run-time bursting from one programmatically bought resource into another with excess capacity is allowed across similar computing platforms.

4. TOOLS/CAPABILITIES

We discuss the state of the practice in the broad area of enabling and offering a DCDE. This section includes an overview of how existing tools and research efforts have shown components of the DCDE concept - enabling scalable distributed computing and mechanisms to manage computing and move scientific data. A conceptual DCDE may be divided into five high-level areas of emphasis shown in Figure 1.

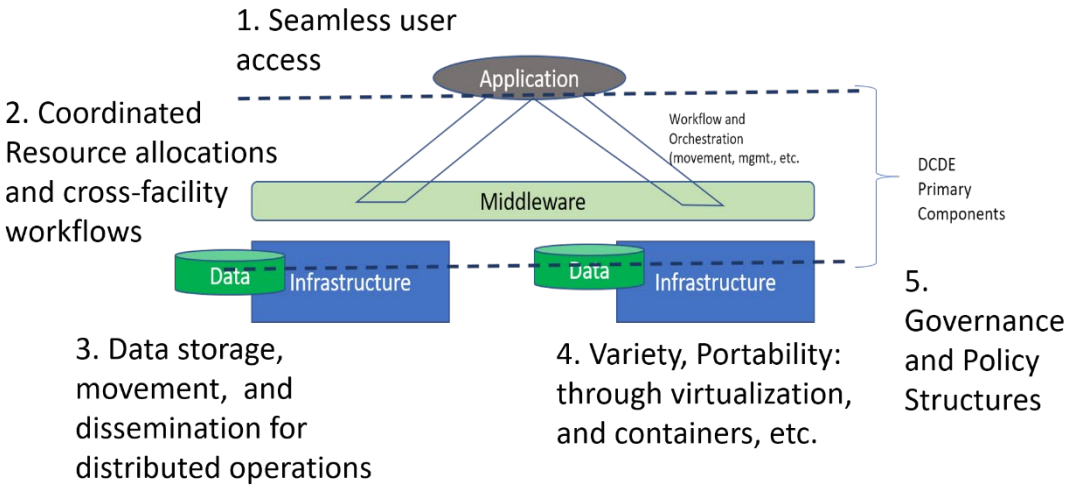

Figure 1. Conceptual block diagram to frame existing tools and capabilities discussion.

These high-level areas are (1) an ability for users to seamlessly access federated resources, (2) coordinate resource allocations and execute cross-facility workflows across the DCDE, (2) support the data management lifecycle, its movement at high-speed, and dissemination, (4) take advantage of functional variety and portability to co-opt system heterogeneity, and (5) operate within effective governance policies and mechanisms.

The FLC-WG and the broader NLRCG invited subject matter experts (SMEs) to speak to the state-of-the-art and practice in the topics identified above. Over the years, ASCR has funded research programs of several of these SMEs who had established collaborations with SC scientists and have been supported for

their widely recognized understanding of the needs and challenges in the scientific community. Our goals in inviting these SMEs for input were to build on existing research advances and identify and explore mechanisms to integrate them so that they could be operational, widely available, and span DOE facilities. We summarize the lessons learned and forward-looking guidance from the SMEs in the areas of emphasis shown in Figure 1. We also give the salient findings from the discussions for each of the areas in a corresponding subsection below. We discuss the concern of governance in the next section because, although there are exemplars of how this should be set up (such as in the formation of ESnet), it requires policy and procedural engagement of human stakeholders and is not simplified by an available technical solution.

4.1 SEAMLESS USER ACCESS

A first step in user access is an intuitive notion of user or application identity and identity management (IdM). Convenient IdM is vital for science for the ease of use for a diversity of facilities. As science has evolved in recent decades from scientists being collocated with facilities to a modern era of remote access to several facilities, being able to "carry" and use a single identity across institutions and facilities is a critical ingredient to seamless user access.

In recent work, Cowles et al. [12] have made the observation that scientific computing typically needs three types of data for effective IdM: (i) user identifier, (ii) contact information, and (iii) Virtual Organization (VO) membership and role. The basic relationship (classic trust model) between user and resource provider (RP) or a facility has evolved over the past two decades to accommodate new models of access, which include brokered trust, transitive trust, and organization-to-organization trust. To enable these access models, active developments of IdM delegation have come into vogue. From the long-standing Public Key Infrastructure (PKI)/X.509 certificates, we now have Security Assertion Markup Language (SAML)-based and social-id based mechanisms to delegate access control and authorization, and to simplify the ability for users to access diverse facilities. Recent recommendations to the DOE Science community [13] propose mechanisms to improve security and accelerate adoption through a risk-based approach. Ultimately, it should be the DCDE's goal to enable any domain user (who may even be a computing novice) to access and use our increasingly complex facility environments easily and effectively.

A memo from DOE on OneID on December 18, 2017 from the DOE Chief Information Officer recommends integration with OneID where feasible. The memo states: "All DOE elements are to consider the use of OneID as the preferred logical access management tool to simplify access to systems, particularly those that cross organizational lines." Although OneID makes it easier for DOE elements to share users, the DCDE should plan to address mechanisms to include potential partners in academia and industry as well. It should be noted that, even if the OneID implementation is a big step forward an integrated system, it will only directly address the first problematic item from above, i.e., user identity. Mechanisms to address membership in projects and other attributes need to be implemented.

Findings:

- DOE science is increasingly carried out by small teams or individual scientists, with few connections to computer scientists, and without expertise in DCDE based computing services.

- A common, standardized identity recognition mechanism across the various facilities of the complex is a prerequisite for a distributed system. Mechanisms should also exist to manage capabilities and rights across platforms.

- There are available options to provide single federated access to a DCDE, such as SAML-based solutions and Open Researcher and Contributor ID (ORCID) (uses OpenID and has persistence with increasing scope).

- InCommon [25] federated identity is widely used in academic research and recognized by some laboratories.

4.2 COORDINATED RESOURCE ACCESS AND CROSS-FACILITY WORKFLOWS

Several tools and middleware have been created to enable computational tasks to operate on distributed computing grids [42]. The FLC-WG reviewed leading examples involving the DOE labs in providing distributed workflow support and resource scheduling.

4.2.1 Workload management

For the past 15 years, the use of distributed computing has been one of the science-enabling successes of the Worldwide Large Hadron Collider Computing Grid (WLCG) [48], and its U.S. component, the OSG. Several workflow managers software have been developed by the HEP community to exploit a variety of distributed resources. PanDA [36], based on Condor [10] and developed in the U.S., is a pioneer in this field. Other workload managers are emerging, building on the experience acquired and aiming at addressing shortcomings and operational issues.

The science in HEP experiments utilize high-energy and high-intensity particle accelerators and sky surveys, coupled with massive, complex detectors (primarily at FermiLab in the U.S., and CERN (European Organization for Nuclear Research) in Europe) as the scientific domain's tools for exploring nature. The experiments' data analysis includes large, distributed, international collaborations (in some cases thousands of collaborators per experiment) working on large data-sets (O(100's PB) current). The deployment includes 100+ compute clusters (OSG and WLCG) using over 400,000 cores, ~450 PB Disk, and ~750 PB tape space. Strong networks connect the individual sites, leading to weekly transfer volumes between all sites between 4-6 PB, and a total Large-Hadron-Collider (LHC) Trans-Atlantic network capacity of 400 gigabits per second.

The ever-increasing computing needs of the HEP community have pushed the HEP Cloud/OSG infrastructure to investigate and use heterogeneous public clouds and DOE computing resource capabilities to solve their peak computing requirements. For example, HEPCloud [24], in partnership with several national laboratories and relying on ESnet and public cloud services, instantiates a data workflow, communicates with a facility interface for access control, and relies on a decision engine to provision resources available in the facility pool. Figure 2 illustrates this workflow.

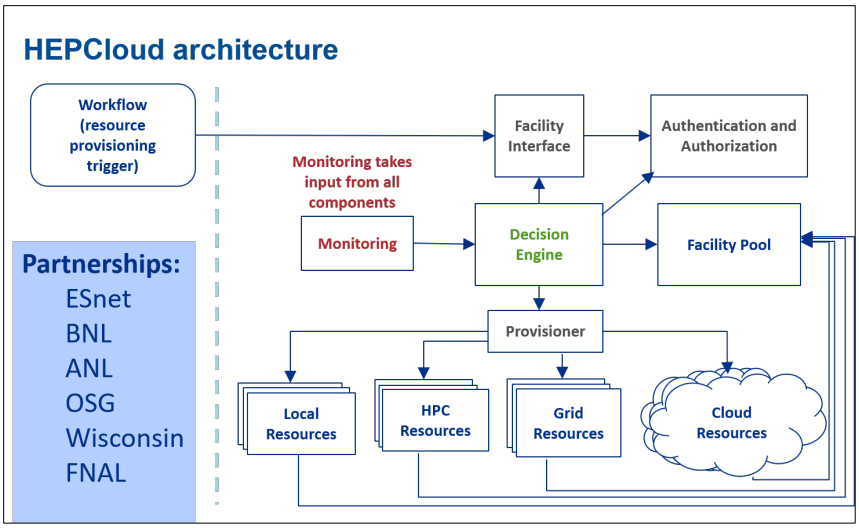

Figure 2. Layout of High Energy Physics (HEP) cloud.

While highly-sophisticated and incorporating years of operational expertise, workload managers developed by HEP have not yet emerged from the bounds of their original science domain for broader community use.

4.2.2 Domain agnostic workflow management

In the class of domain-agnostic workflow tools, we reviewed input from the creators of two tools in widespread use in the DOE and academic community: Pegasus [38] and Swift [47]. Both offer a programmable interface for users to orchestrate and execute tasks across multiple facilities. In the tradition of abstracting main requirements from a workflow tool, these tools support heterogeneous infrastructures and a separation of workflow description and execution. They interact with different facility resources, can be run on small to large-scale computing clusters, and interact with diverse storage systems.

Pegasus also supports data reuse, which is useful for collaboration and ensemble workflow runs, recovers from failures, and supports workflow restructuring for performance improvement. The Pegasus approach is to "Submit Locally, Run Globally," as illustrated in Figure 3 [38]. The architecture illustrates the manner in which a Laser Interferometer Gravitational-Wave Observatory (LIGO) [28] workflow description is submitted to a centralized Pegasus Workflow Management system, which submits jobs to a local or remote pool. Jobs and tasks pull data directly from the source.

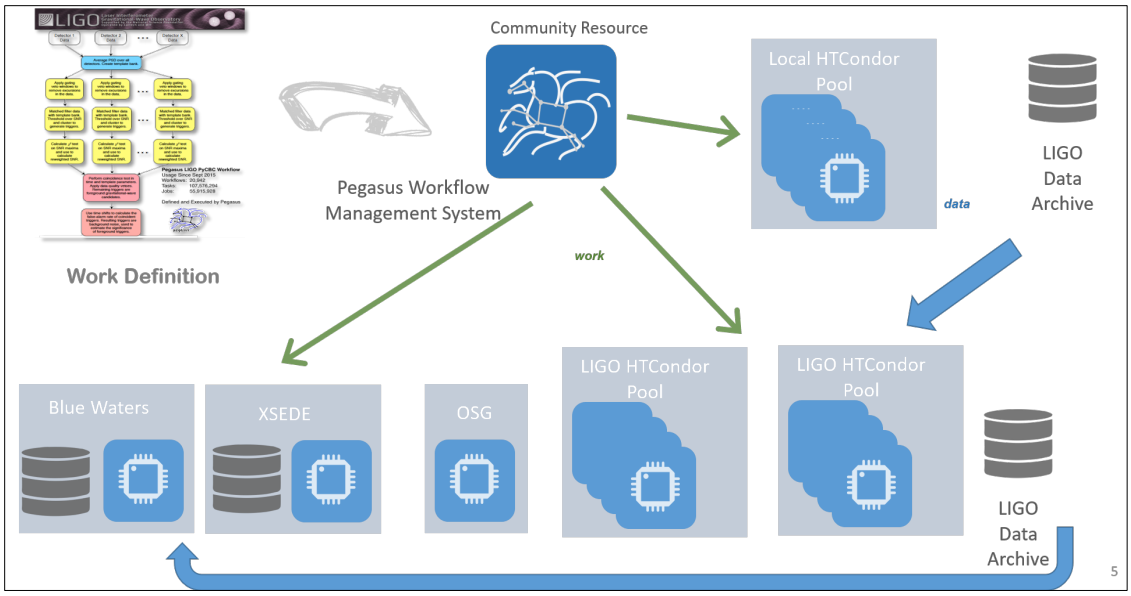

Figure 3. Pegasus example with Laser Interferometer Gravitational-Wave Observatory (LIGO) workflow.

Swift (and a recent follow on project, Parsl [37]) enables parallelism to be more transparent through dataflow programming, allows the computing location to be more transparent, supports basic failure recovery transparent (through retries/relocating failing tasks), and enables provenance capture (tasks have recordable inputs and outputs). In a manner similar to Pegasus, Swift enables a script to be submitted by a Swift host to the runtime system, which has the drivers to support a variety of runtime environments, as shown in Figure 4.

Finally, while there are existing tools for establishing workflows, an often-ignored component of these cross-facility middlewares is targeted collaboration tools within and across communities. The current solutions are a diverse set of commercial offerings with each facility or team using their collaborative environment of choice.

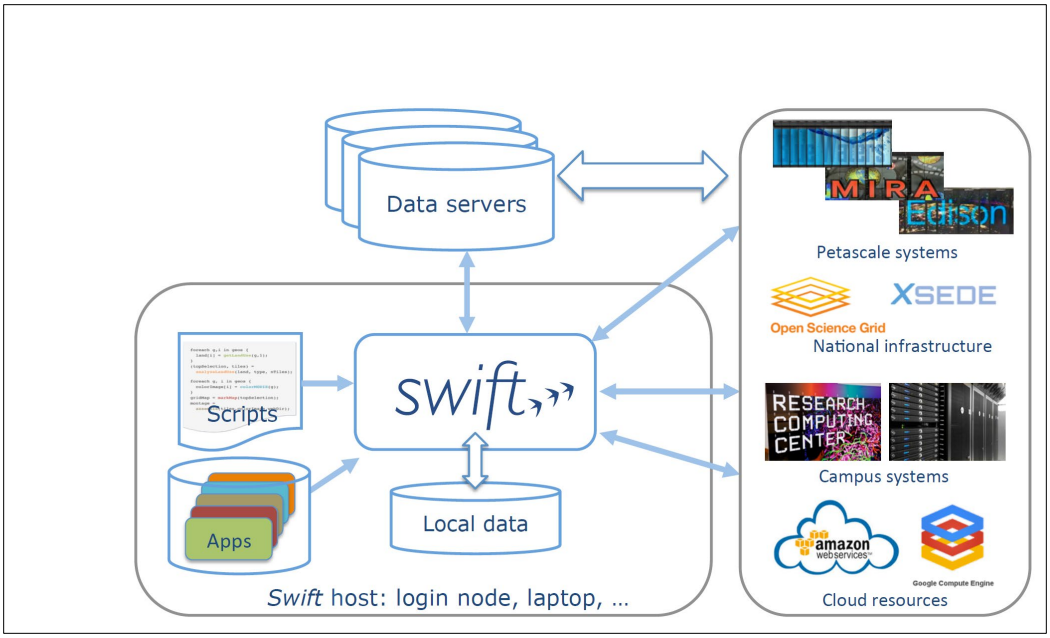

Figure 4. Swift Workflow Components

Findings:

- There already are existing large-scale middleware and workflow enabling capabilities that can satisfy the workflow and resource coordination need for a DCDE, but they have differences in interface and implementation, and offering all the tools may be difficult. The DCDE will need the right abstractions so that we do not perpetuate technology silos. The DCDE toolkits will require individual facilities to support common resource abstractions and well-defined interfaces. These will need to be easy to evolve and co-opt into available tools. (See also the section on Governance.)

- Researchers need guidance and simplification of tools so that workflows may be available as a service.

- Existing tools are evolving to meet the need, e.g., by moving to Python and supporting Jupyter notebook interfaces.

- We lack an existing generalized collaboration environment that connects all the tools one would need to actively collaborate with other researchers in real time.

4.3 SCIENTIFIC DATA MANAGEMENT: MOVEMENT, DISSEMINATION AND LONG-TERM STORAGE

Scientists have a generalized need for the capability to place their data where it will be scientifically useful. That capability must scale up as the data sets increase without a corresponding scaling up of the human effort required to move the data. DOE's ESnet provides the long-distance part of this capability: it interconnects DOE labs and facilities with state-of-the-art services and connects them to collaborating institutions and the wider internet. In collaboration with ESnet, the site and facility local networks provide the connections between facilities, data sources, HPC resources, and the ESnet network.

4.3.1 Scientific Data Management

The NLRCG has initiated a working group on scientific data management [51], which has proposed a systematic methodology to support a scientific data management lifecycle. Data creation sets off actions that include a basic storage step from which the data can be moved to an archival setting (Figure 5), publishable state for dissemination, or a state of being deleted. From the data store, users execute a cycle of using the data to create (zero or more new information) or return the data to storage. Policy decisions influence the actions associated with the various steps of the management of the scientific data sets. For example, DOE defines a digital research data management policy [16] focused on stewardship of the data artifacts associated with sponsored projects.

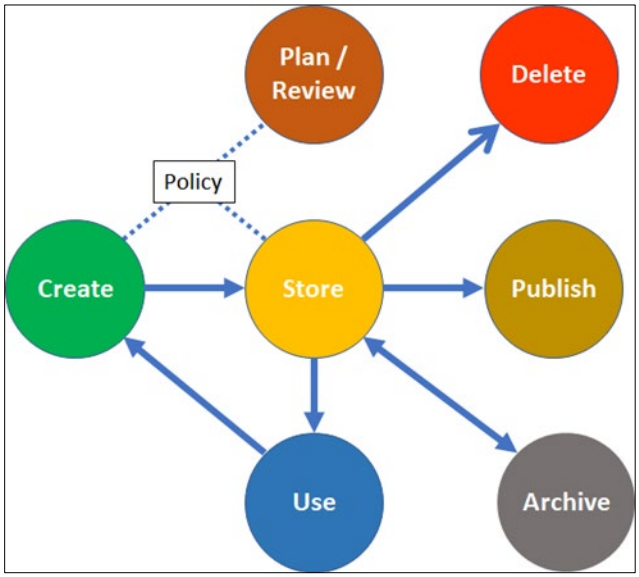

Figure 5. Example of actions performed on data.

The FLC-WG reviewed the integrated Rule-Oriented Data System (iRODS) [26] industry consortium's offerings in data management. iRODS offers an open source, distributed, metadata-driven capability, which combines various distributed storage technologies (existing file systems, cloud storage, on-premises object stores, and archival systems) into a Unified Namespace. Through a set of rules and policy enforcement points, iRODS enables workflows across the infrastructure to enable secure collaboration and federation (Figure 6).

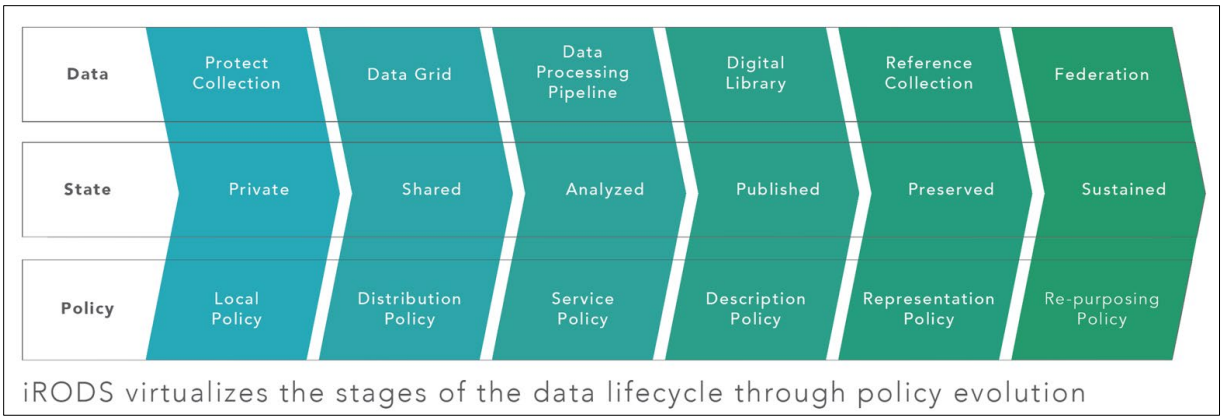

Figure 6. iRODs data lifecycle.

Another example of a distributed data management capability with established large-scale support is the Rucio tool [40], originally created for data sharing for collaborators on the ATLAS experiment. Rucio enables declarative management of data copies, supporting federated access and a namespace management facility across groups. Third party copy and information services (e.g., Globus) may be incorporated to support specific modes of data handling and movement.

Other science communities have developed unique tools to share and disseminate data, including visual tools to display the data. Notable examples include the environmental science community's ESGF and the astronomy community's SAGA data management software suites [50].

4.3.2 Data Movement, Streaming

Over the past several years, the Science DMZ (De-Militarized Zone) design pattern [15,22] has emerged as best practice for the design and configuration of site local network enclaves that support high-speed, at-scale data transfers for scientific applications. The Science DMZ is typically a separate security enclave, where the security policy and its enforcement mechanisms provide both the protection and defense of the Science DMZ resources, and the performant operation of those resources. Within the Science DMZ are Data Transfer Nodes (DTNs), which are high-speed data servers configured specifically and solely for data transfer.

The DTNs in a Science DMZ run data transfer tools and platforms to enable scientists to transfer data. A set of legacy tools, such as FTP (file transfer protocol), HTTP (hypertext transfer protocol), and rsync, are familiar to scientists but typically do not scale to the requirements of terabyte- and petabyte-scale data sets. To address these failings, high-performance tools such as GridFTP were developed, which perform better but are difficult for many scientists to use. The solution has been the emergence of high-performance data orchestration platforms that hide the complexity of the tools behind a software-as-a-service interface.

The most common of these platforms in the DOE labs is Globus [23], which is built on top of the legacy GridFTP tools but eliminates the need for users to understand such details as performance optimization, transfer management, integrity verification, and fault recovery. These features are key to scalability from both performance and productivity perspectives: if the software can gracefully recover from faults and manage large operations with a simple user interface, scientists need not expend additional effort to optimize data transfers as the scale of their data sets increases.

An area of active research which does not yet have easy-to-deploy cross-facility capabilities is data streaming. DOE, in collaboration with the NSF and the Air Force Office of Scientific Research, has conducted recent workshops on data streaming identifying the current state-of-the-art and future research directions [41]. As the needs of streaming become explicit, the DCDE will need to begin incorporating the emerging tools developed by this community.

4.3.3 Scientific Data Dissemination

Dissemination through legacy data portals used a web server with a database for user and metadata, computer generated imagery programming for server-side functionality, and access to a filesystem containing the data objects. The web browser provided a graphical user interface, which the portal developer did not have to build. The combination of these capabilities provided a significant improvement over command-line FTP, which was the state-of-the-art for data access before data portals.

However, the legacy data portal design comes with several significant limitations:

- It is inefficient for a large data sample, error-prone, and time consuming. If a scientist wants to retrieve several thousand data objects from the portal, clicking on individual files in a web browser is an error-prone and inefficient way of getting the data. The scientist can write a script using wget or another similar tool to fetch individual files, but this is also error-prone (e.g. what if the 547th file out of 984 files doesn't transfer correctly?). Also, this necessitates the scientist spend time creating and managing scripts to interface with the data portal.

- Another limitation is performance and scalability due to the coupling of storage read to web server performance.

The Modern Research Data Portal design pattern (MRDP) solves most of these issues. The web and data service functions of the data portal are separated from each other, and the data service portion is outsourced to a data transfer platform, which uses DTNs in a Science DMZ (Figure 7). When the modern data portal gives the user references to data objects, the references don't point back at the portal web server, rather they point to a DTN cluster that can be scaled up as needed without modifying the data portal web application at all.

Findings:

- A lack of an existing generalized cross-facility data management solution is a gap. One of the main challenges stems from the lack of a generalized, holistic data management solution. While there are tools that claim to resolve data management issues, many of them do not resolve the capturing of the many layers of metadata required to describe the complicated experimental designs of observation science.

- Provenance and Ownership. We need to be able to know the provenance of the data, who the current (active) data owner is, and when to delete data. In addition, system owners need to be able to engage directly with the data owner through the system and get approval to use, cite, and move the data.

- Effective mechanisms to protect and share sensitive data. We don't have a process or system to assess risk in sharing data. The data could have business sensitive data, personally identifiable information, or other risks embedded in the data. We need a method for identifying, quantifying, and ultimately managing the risk in sharing data.

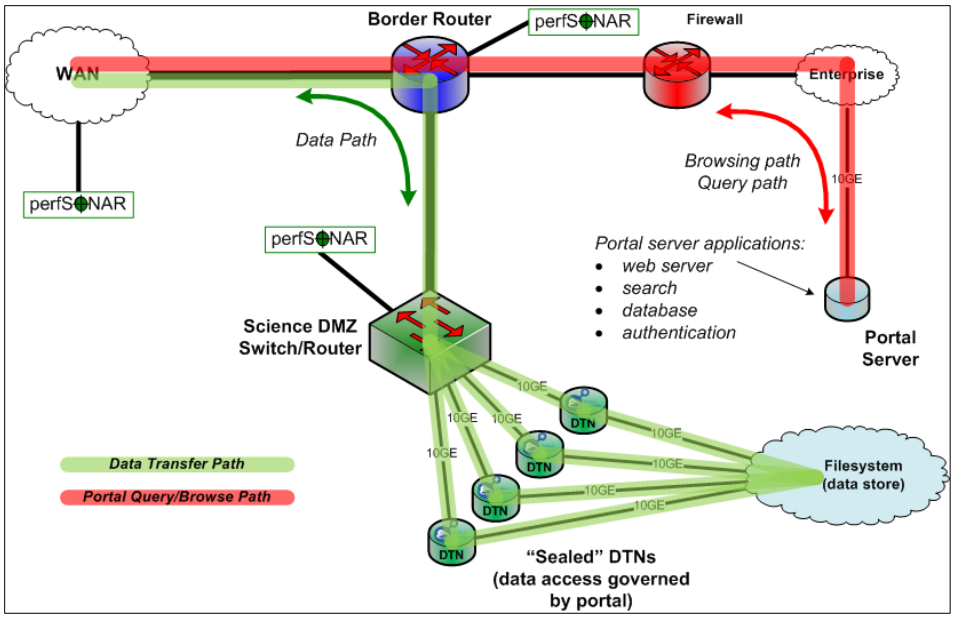

Figure 7. The Modern Research Data Portal design pattern, showing the portal web server separated from the Science DMZ, which serves the data objects and the separate paths from the web server and the Science DMZ.

- Data are produced at facilities that may not provide long-term storage or archival capabilities. Large-scale, long-term storage capabilities and knowledge are available at some laboratories that have already invested in costly archival infrastructures.

- Scientific data transfer and dissemination can be enabled by emerging patterns such as the modern data portals backed by science DMZs and high-speed data transfer infrastructures currently operational in the DOE environment.

- End-to-end data streaming solutions from observational experiment to computing resources remains an area of active research.

4.4 SUPPORTING FUNCTIONAL VARIETY AND PORTABILITY

For an effective DCDE, software written and executed in one environment must be portable and runnable in another environment. While programming languages have evolved to support tremendous portability (e.g., Java using its virtual machine approach), large gaps have remained in encapsulating growing numbers of libraries in an executable module. The relatively recent technology advancements in Containers promise to mitigate backward (and forward) portability challenges. The NLRCG and Large-Scale Network Interagency Working Group Middleware and Grid Interagency Team known as the MAGIC [29] subcommittee received presentations on this topic, which were shared with the FLC-WG.

Containers promise mechanisms for software to be "packaged" along with its associated libraries in a way that the reliance on the underlying operating system infrastructure is reduced to a bare minimum — no longer are individual disparate modules required to be built separately and made available to run an application; the delivery mechanism is the whole packaged application with only an operating system kernel dependency. Capturing a similar philosophy of a virtualized operating system environment, the dependencies are limited to key system calls — the rest of the libraries and filesystem dependencies are encapsulated into the container. Although dependent on kernel calls, the application-dependent library

isolation facilitates portability significantly. While virtual machines abstracted the operating system and only depended on the hardware, containers offer a lightweight, application-layer instantiation of this approach. The state-of-the-art is evolving in this space particularly in its relationship to underlying hardware (e.g., accelerator and network capabilities), but the technology remains a promising approach towards the end goal of "write/build once" and "run everywhere."

Findings:

- The needs for virtualization of resources and services are increasing. Virtualization and containerization technologies are improving and making run-time environments portable.

- A federated DCDE will need to span a wide variety of heterogeneous compute systems (i.e., desktops to supercomputers) and deeply interconnect them with other physical devices (i.e., storage systems, distributed sensors, and large/unique instruments) using a high-performance network.

5. ORGANIZATIONAL CONCERNS AND GOVERNANCE

A key element for the success of a DCDE is a functioning governance model. The details of the implemented model will depend on the details of computing resources and funding mechanisms, but at a minimum will include mechanisms to track and allocate resources by connecting with users and setting standards of performance and metrics for resource providers.

Distributed projects like a DCDE need distributed and coordinated information systems and accounting. If computing and storage resources are made available broadly, they should be publicized to the researchers to know their capabilities and their availability. Currently no straightforward mechanisms or policies exist for a given researcher or group of researchers to easily access resources at different locations across the DOE complex. Each site has its own policy for managing and attributing accounts. Accounting of resource usage and information about computing systems are also not easily available outside a given facility for a global overview. For example, there is currently no simple process for a research project to get multi-year allocations across facilities. Research projects extend over several years and need resource planning beyond a given year for proper project planning. It is also very difficult for non-computer scientists to compare the relative units used for allocations (e.g., CPU-hours are hardware dependent).

The governance model also needs to support all elements of planning, including staffing, funding and sustainability, procurement and requirement processes, resource planning, user support, and support for services. Organization and governance of the DCDE will evolve as the DCDE itself evolves with time. Operating a DCDE will require lab staff and management to develop new skills via training, and to collaborate with their peers across the lab complex.

A number of organizations that administer shared computing resources can serve as models. The details of their governance models differ—capacity computing user facilities, such as NERSC, have a different model than the OSG model, which is a federated set of resources that are owned and administered by local sites, but with OSG providing user support and the software stack. Compute Canada [9] provides another model where distributed resources are planned, administered, and allocated through a centrally-funded organization at the federal level, with local contribution from its provinces. XSEDE provides an additional model that might have applicable experience. All these projects are articulated around three main components: organization and technical support, user communities, and service providers. While

key technical components have been developed in the DOE, NSF, and commercial communities, the specifics of integration to bring the technical components together in a DCDE will need to be fleshed out in a prototyping exercise. The integration specifics would reveal the following aspects of organizational structure and governance:

- Economical model (resource sharing and publication) to determine levels of service and structure of the laboratory communities, service level agreements and enforcement of policies, and adoption and reuse promotion.

- Accounting model (resource usage and accounting) showing how users will gain access to, and be properly billed for, spare CPU resources across facilities while preserving local users' priorities. Creating resource-to-user matches and discriminating between temporary allocations (e.g., for compute cycles) and more static allocations (e.g., for storage).

- Operational model showing how user communities interact with service providers while specifying the minimum set of services to be supported by all resource providers, including a description of how services will be deployed, operated, and maintained uniformly across the DCDE.

6. MOVING FORWARD TO A PROTOTYPE

As noted earlier, the science use cases for using distributed resources are emerging from many science domains. The goal of a successful DCDE would be to present to the researchers a variety of distributed resources through a coherent and simple set of interfaces, which allow them to manage data and related computations throughout the scientific discovery lifecycle, from idea inception to archival after publication. The locations of these resources will not coincide with the places where data is produced, stored, analyzed or archived. The increased usage and availability of geographically distributed computing and storage resources highlight the need for easy-to-use tools that are capable of handling this distributed paradigm transparently. The success of a DCDE will be proven by its simplicity and ease of use. It should promote widely used standards and tools without imposing a single mandated solution on a large variety of science programs and resource providers. In addition, since it is envisioned as a cross-laboratory environment, a DCDE should establish a governance body that includes the relevant stakeholders.

To test the complexity and robustness of the DCDE concept and deployment approach built on existing technologies and little reinvention, a prototype should be established in the near term which implements in a coherent and progressive manner the five main components of a DCDE. The prototype would be limited in size and in duration, while addressing the key concepts of a DCDE. The prototype will help in defining a general set of recommendations, supported by implementation experiences, for expanding the DCDE to the whole lab complex and to produce an applicable governance model.

Such a pilot project will expose gaps and challenges clearly while testing available technical solutions that will enable a DOE-wide DCDE. To be successful the pilot will need strong support and commitment from the stakeholders involved.

7. SUMMARY

The science use cases for using distributed resources across the DOE ecosystem are emerging from many science domains. The FLC-WG's main findings are:

- A DCDE is required to address drivers and requirements for the science use case as expressed in past workshop reports from offices within the DOE.
- Five areas of activity support a DCDE: seamless user access, coordinated workflows across sites and facilities, data management and movement, portability and accommodating heterogeneity, and robust governance frameworks.
- Previous research programs have provided existing technologies and approaches that developed large parts of the solution set that would comprise a DCDE.
- There do exist technology, policy, and governance gaps and hurdles that need to be overcome, but several technical components are available to take steps towards deploying a DCDE. Remaining technical gaps, policy, and governance frameworks should be identified during a pilot DCDE project.

8. REFERENCES

[1] [Online] https://www.alcf.anl.gov/
[2] [Online] https://www.lcrc.anl.gov/
[3] [Online] https://arxiv.org/abs/1603.09303
[4] [Online] https://science.energy.gov/bes/
[5] [Online] https://science.energy.gov/~/media/bes/pdf/reports/2017/BES-EXA_rpt.pdf
[6] [Online] https://www.bnl.gov/compsci/
[7] [Online] https://www.sdcc.bnl.gov/#ic-cluster
[8] [Online] http://genomicscience.energy.gov/pubs/BSSDStrategicPlanOct2015.pdf
[9] [Online] https://www.computecanada.ca/
[10] [Online] https://research.cs.wisc.edu/htcondor/
[11] [Online] http://cades.ornl.gov
[12] Cowles, R., Jackson, C., and Welch, V. (2014) A Model for Identity Management in Future Scientific Collaboratories. International Symposium on Grids and Clouds. DOI= http://pos.sissa.it/archive/conferences/210/026/ISGC2014_026.pdf
[13] Facilitating Scientific Collaborations by Delegating Identity Management: Reducing Barriers & Roadmap for Incremental Implementation, http://hdl.handle.net/2022/20357, Presented at CLHS 15, June 2015
[14][Online] CrossCut Report, Exascale Requirements Reviews, March 9-10, 2017, https://science.energy.gov/~/media/ascr/pdf/programdocuments/docs/2018/DOE-ExascaleReport-CrossCut.pdf
[15] E. Dart, L. Rotman, B. Tierney, M. Hester, and J. Zurawski, "The science dmz: A network design pattern for data- intensive science," in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, ser. SC '13. New York, NY, USA: ACM, 2013, pp. 85:1–85:10. [Online]. Available: http://doi.acm.org/10.1145/2503210.2503245
[16]https://www.energy.gov/datamanagement/doe-policy-digital-research-data-management
[17] Deelman, et al., The Future of Scientific Workflows, International Journal of High-Performance Computing, April, 2017, https://doi.org/10.1177/1094342017704893

[18] [Online] https://science.energy.gov/~/media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report_160524.pdf
[19] [Online] https://esgf.llnl.gov/
[20] [Online] https://science.energy.gov/~/media/ber/pdf/workshop%20reports/Towards_a_Shared_ESS_Cyberinfrastructure.pdf
[21] [Online] http://exascaleage.org/
[22] [Online] http://fasterdata.es.net/science-dmz/
[23] [Online] https://www.globus.org/
[24] [Online] http://hepcloud.fnal.gov/
[25] [Online] https://www.incommon.org/federation/
[26] [Online] https://irods.org/
[27] [Online] lcrc.anl.gov
[28] [Online] https://www.ligo.org/
[29] [Online] https://www.nitrd.gov/nitrdgroups/index.php?title=Middleware_And_Grid_Interagency_Coordination_(MAGIC)
[30] Chard K, Dart E, Foster I, Shifflett D, Tuecke S, Williams J. (2018) The Modern Research Data Portal: a design pattern for networked, data-intensive science. PeerJ Computer Science 4:e144 https://doi.org/10.7717/peerj-cs.144
[31] [Online] http://www.nersc.gov/
[32] [Online] https://science.energy.gov/np/nsac/
[33] [Online] https://www.olcf.ornl.gov/
[34] [Online] https://cades.ornl.gov/
[35] [Online] https://www.opensciencegrid.org/
[36] T. Maeno, K. De, T. Wenaus, P. Nilsson, G. Stewart, R. Walker, A. Stradling, J. Caballero, M. Potekhin, D. Smith et al., "Overview of ATLAS PanDA workload management," in J. Phys.: Conf. Ser., vol. 331, 2011, p. 072024.
[37] [Online] http://parsl-project.org/
[38] [Online] https://pegasus.isi.edu/
[39] [Online] https://pic.pnnl.gov/
[40] [Online] https://rucio.cern.ch/
[41] [Online] http://streamingsystems.org/
[42] [Online] Xavier Grehant, Isabelle Demeure, and Sverre Jarp. 2013. A survey of task mapping on production grids. ACM Comput. Surv. 45, 3, Article 37 (July 2013), 25 pages. DOI=http://dx.doi.org/10.1145/2480741.2480754
[43] [Online] http://swift-lang.org/main/
[44] [Online] https://en.wikipedia.org/wiki/TeraGrid
[45] [Online] https://science.energy.gov/~/media/ber/pdf/community-resources/Technologies_for_Characterizing_Molecular_and_Cellular_Systems.pdf
[46] The Trusted CI Vision for an NSF Cybersecurity Ecosystem And Five-year Strategic Plan (2019-2023). V Welch, J Basney, C Jackson, J Marsteller, B Miller - 2018
[47] Michael Wilde, Mihael Hategan, Justin M. Wozniak, Ben Clifford, Daniel S. Katz, and Ian Foster. 2011. Swift: A language for distributed parallel scripting. Parallel Comput. 37, 9 (September 2011), 633-652. DOI=http://dx.doi.org/10.1016/j.parco.2011.05.005
[48] [Online] http://wlcg.web.cern.ch/
[49] [Online] https://www.xsede.org/
[50] [Online] http://sagasurvey.org/
[51] National Laboratory Research Computing Group – Scientific Data Management Discussions. Picture courtesy: Stuart Fuess, chair.

APPENDIX A. NOTES