Open Storage Network Retrospective & the Future of Distributed Storage for eInfrastructure

Open Storage Network Concept Paper

August 9, 2021

Christine R. Kirkpatrick[1], Kevin Coakley[1], Melissa Cragin[1], Catherine Cramer[1], Kenton McHenry[2], Luigi Marini[2], Rob Kooper[2], Jim Glasgow[2], Ian Foster[3], Alex Szalay[4], Derek Simmel[5]

Affiliations: 1. San Diego Supercomputer Center, University of California San Diego, 2. National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 3. University of Chicago, 4. Johns Hopkins University, 5. Pittsburgh Supercomputing Center

The Open Storage Network is funded by the National Science Foundation under grants #1747552, 1747493, 1747507, 1747490, 1747483 and by Schmidt Futures

http://www.openstoragenetwork.org/

Executive Summary

The Open Storage Network (OSN) launched as a pilot designed to provide a national storage substrate for access and use of active scientific and scholarly data. This distributed service leverages existing investments in network infrastructure to support high speed transfer among research institutions and computational facilities. Over the last five years the U.S. National Science Foundation (NSF) has funded more than 200 high-speed connections to the Internet-2 backbone operating at 10-100Gbps speeds. One goal of the OSN was to develop standardized storage module *pods*, for a high performance distributed system that extends the usability of the existing high-speed interconnects. OSN pods each provide 1+ PB of active storage at low cost with leading edge automation and requiring little maintenance. These are linked together to form a larger storage pool that supports access, data movement, replication, and local caching near computational resources. At full scale, the OSN could serve Petabyte-scale research, which is generally beyond the capacity of individual academic institutions and independent research organizations. Scientific and scholarly communities are facing a major challenge dealing with the rapidly increasing amount of research data emerging from projects produced at all scales--from large facilities to small research labs.

The OSN is a boon to computing on the network, particularly for the growing needs of mid-scale researchers, which while focusing on collaboration and open data, also need to acquire cyberinfrastructure for research, without necessarily having the budget for data-intensive infrastructure. OSN pods provide a solution that helps to fill a gap in high performance and big data applications, building in performance, flexibility, and cost effectiveness that is competitive with, or exceeds current solutions. The OSN is a viable data storage and sharing solution for research groups that span multiple institutions and cannot span institutional firewalls and security domains. Research projects will increasingly need an option like the OSN to support transdisciplinary and at scale science.

Introduction

As researchers have become increasingly dependent on data, the open data movement has emerged, concurrent with policy trends such as open government, open publishing and open science. Scientific and scholarly communities face a major challenge working with the burgeoning amount of research data emerging from projects produced at all scales -- from large facilities to mid-scale, to small research labs. This concept paper describes how the Open Storage Network (OSN) addresses many of these challenges facing researchers, as well what motivates them to use the OSN.

A four-part webinar series was held to explore these challenges and solutions. In the first, research teams described the research drivers that motivated them to adopt the OSN to help share their research, and how the OSN helped address these issues. These motivating factors included the cost of sharing multi-petabyte data sets (particularly, data egress costs) with the public (example use case: Terra Fusion); providing domain experts and data scientists access to large and diverse sets of IoT sensor data to improve AI models (example use case: Sage); and using the OSN with IoT and AI applications to help provide pertinent data to scientists, engineers and public health officials to answer questions of public safety (example use case: HydroShare).

During the second OSN seminar, researchers presented examples of the state of the art in national and international trends in research storage at scale, including delivering data storage services that meet the needs of researchers, federating IT services and simplifying access to storage services for researchers, and using OSN Pods. (Pod performance is comparable and in some cases exceeds the performance of object storage services from commercial cloud providers.)

A third webinar reported on the OSN to date, including the OSN attributes and capabilities in detail, and the software stack that allows the OSN to support a large number of clients and tools including interoperability with commercial clouds such as AWS. The webinar also provided updates on the favorable OSN Pod dollar-per-terabyte price, the availability of twice the storage capacity concurrent with 99.9999% durability[1], and future design plans to extend capacity and add durability, while gaining economies of scale through centralized management via the Command Center. The only requirement for a site to participate in the OSN is a high-speed connection to Internet2, preferably via 100 GbE, and access to staff to replace hardware under warranty as needed. The OSN is a very viable alternative to commercial cloud storage for many use cases and as one use case found (Pangeo), 1/10th of the cost. Monitoring and remote administration enabled through the OSN Command Center alleviates Pod operators from significant investment in local IT personnel resources.

[1] Based on Backblaze's Erasure Coding Durability Calculator (https://github.com/Backblaze/erasure-coding-durability)

The user experience for the OSN includes client access controls, with three basic levels of access - read and write, read only and open (unauthenticated) access, client compatibility through data management frameworks, and integration with XSEDE for storage allocations.

The OSN provides the storage substrate using open standards so that individual domains or projects can choose the most appropriate middleware for access, often software already in use. In the sections that follow, several such use cases are described: The deployment of the Clowder instance over data stored within the OSN, as an alternative to the budgeted storage currently being used, is particularly suited to use cases with large files, such as the output of the LIDAR survey. The Clowder framework emphasizes flexibility to customize data management models and interfaces. The Globus research data management platform allows ease of data transfer to and from the OSN and between OSN nodes, as well as data sharing and creation of data portals and integration of OSN storage into science workflows, with the development of a portal to enable discovery and retrieval of TerraFusion data hosted on OSN as an exemplar, and the use of Dataverse open source software for researchers to easily deposit, curate, and share FAIR data. iRods brings together OSN storage with other data stores to provide metadata search, visualization and workflows in a single namespace. And finally, we describe advantages in flexibility and cost effectiveness for OSN in mid-scale computing.

Research Drivers and Capabilities

In the first OSN seminar, multiple researchers presented the research drivers that led them to the OSN as well as why data sharing was needed.

The first presentation was from Donald Petravick of the National Center for Supercomputing Applications (NCSA) at the University of Illinois on how the Terra Fusion project uses the OSN to share multi-petabyte data sets. The Terra Fusion project has 2.4 PB of data and is growing. Due to the size of the Terra dataset, NASA doesn't make the complete dataset available to the public -- instead NASA works with the Terra Fusion project to fuse data collected over the same time period from multiple instruments into a fused dataset. These fused datasets are then divided into smaller datasets called “Samplers” and shared to the public using the OSN. The cost of sharing the data to the public was a primary research driver for Terra Fusion’s adoption of the OSN. Due to the data egress costs of the commercial cloud, hardware for three OSN pods could be purchased for the cost to download the whole Terra Fusion dataset 3.3 times.

In the second presentation, Wolfgang Gerlach from the University of Chicago and Argonne National Laboratory presented how the Sage project uses the OSN with IoT and AI applications. The Sage project uses AI to streamline edge computing applications by doing as much of the analysis as possible on the IoT device, and then pushing the derived data to the OSN. The Sage project uses the OSN to store images, videos, sound files, LIDAR data and multispectral images from edge sensors across the United States. The research driver for the Sage project is to provide domain experts and data scientists access to a large and diverse set of IoT sensor data in order to improve the efficiency and accuracy of their AI models.

The last presentation was from Christina Bandaragoda of University of Washington and Chris Lenhardt of Renaissance Computing Institute (RENCI) at UNC Chapel Hill on how they use the OSN to make Integrated Hurricane Data easily accessible to researchers and public health officials in order to make critical decisions before, during and after a hurricane. The research driver for the Integrated Hurricane Data is to provide pertinent data to scientists, engineers and public health officials to answer questions of public safety, like identifying areas where flooding could occur and the potential impacts on quality of drinking water, during and immediately after a hurricane. With collaboration with HydroShare and NOAA, 30 TB of data from Hurricane Matthew and Hurricane Harvey are available for download via the OSN.

A webinar on OSN Research Drivers and Capabilities is available[2] and the full concept paper can be read at https://doi.org/10.6075/J0HM56Z7.

National and International Trends in Research Storage at Scale

In the second OSN seminar, researchers from three major projects presented the state of the art in national and international trends in research storage at scale.

In the first presentation, Lene Krøl Andersen, the EOSC-Nordic project manager from the Danish eInfrastructure Cooperation (DeiC), presented how EOSC-Nordic used four themes to bring together dozens of institutions across the Nordic and Baltic regions to utilize the capabilities of the European Open Science Cloud (EOSC). The four themes she presented were: Discovery and Re-Use of Research Data, Analysis and Post-Processing, Data Management Sharing and Archiving, and Sensitive Data and Orchestration. EOSC-Nordic uses these four thematic demonstrators to meet the challenges of delivering data storage services that meet the needs of their researchers.

Uwe Jandt from the Deutsches Elektronen-Synchrotron discussed in the second presentation how the HIFIS initiative federated the IT services among the 18 Helmholtz Association research centers and simplified access to storage services for their researchers. The Helmholtz Cloud integrated the dCache project into their storage fabric, allowing access to storage by users from all Helmholtz centres and their collaboration partners, either directly or via connected services. dCache seamlessly integrates into the authentication and authorization infrastructure of the Helmholtz Cloud and provides a user-transparent workflow to enable buffering and pre-fetching of data transfers to simplify the process of getting data on to the Helmholtz compute services.

The last presentation was from Anita Nikolich of University of Illinois Urbana-Champaign and Ilya Baldin of Renaissance Computing Institute (RENCI) at UNC Chapel Hill on the challenges of managing multiple tiers of data storage at scale across geographic and network borders. Most projects are not properly funded to provide adequate storage for collecting data from an

[2] https://www.openstoragenetwork.org/seminar-series/oct-22-2020/

instrument and doing the analysis on that data. To solve that problem FABRIC[3] has developed a tiered storage and backup model to determine what data to keep and for what period. By considering the required retention times, the data characteristics and the data function, FABRIC developed a network of campus, regional, and commercial cloud partners who work with them on long-term data storage challenges.

The recording of the seminar is available[4] and the full concept paper can be read at https://doi.org/10.6075/J00G3HQ1.

OSN Pod Hardware and Software Stack

OSN Pod Hardware

OSN hardware is deployed in scalable units known as Pods. OSN Pods deliver approximately 1.2 Petabytes of object storage, accessible via S3, for a target cost of $150,000. S3 is the de facto standard interface for cloud storage, developed by AWS. Pods are sold as complete units with a server rack, two power distribution units (PDU), a high-speed data network switch, a management network switch, three monitor/proxy servers, and five storage servers. The Pods are delivered directly from the value-added reseller to sites with all hardware mounted and cabled within the rack. Sites are only required to connect the PDUs to the datacenter's power and configure the network uplinks from the high-speed switch when the Pod is delivered. The only requirement for a site to participate in the OSN is a high speed connection to Internet2, preferably 100 GbE, but sites with slower connections can still participate. Sites that cannot accommodate the OSN provided rack and PDUs can rerack for their environment. Wiring diagrams are available on the OSN website.

Since the original award, subsequent collaborators have been awarded or have proposed additional OSN pods. The furthest along currently is Alabama A&M, who will deploy a pod on their campus that is connected to the OSN network via Internet2. Extending the OSN model, the University of Maine will deploy a virtual OSN pod, following a similar project within MGHPCC.

OSN Software Stack

The OSN uses the free, open source community supported version of Ceph to provide the S3 object storage service. Ceph supports a RESTful API that is compatible with the basic data access model of the S3 API[5]. S3 API compatibility allows the OSN to support the large number of clients and tools that have been built for the AWS ecosystem.

[3] https://fabric-testbed.net/
[4] https://www.openstoragenetwork.org/seminar-series/nov-12-2020-national-and-international-trends-in-research-storage-at-scale/
[5] https://docs.ceph.com/en/nautilus/radosgw/s3/

The OSN Pod employs Erasure Coding (EC) for data resiliency in order to provide the best dollar to terabyte ($/TB) price. The OSN currently uses a 3 data chunk with 1 coding chunk (k=3 m=1) EC profile, due to limitations of the first generation hardware and Ceph Nautilus' EC configuration support. This EC profile provides 2 times the storage of traditional 3 copy replication while still providing 99.9999% durability (the loss of 1 object per 10,000 objects every 10 years). Future Pod designs will be architected to support a wider range of EC profiles to achieve greater $/TB price and durability.

Pod Performance

The OSN pod performance is comparable to the performance of object storage services from Commercial Cloud providers. The IceCube Neutrino Observatory recorded 5.5 GBps read speed from a single OSN Pod using 10 nodes running concurrent downloads. IceCube was only able to get 4 GBps from Azure Storage and AWS S3 in similar tests. Pangeo, a community platform for Big Data geoscience, was extremely impressed with the results of their benchmarks. The Pangeo project also achieved 5.5 GBps read speeds from threaded tests and achieved higher throughput from the OSN than from Google's Cloud Storage. Pangeo reported that the OSN is a very viable alternative to cloud storage for many of their use cases.

Using the OSN

Client Access Controls

The OSN offers three basic levels of client access via the S3 API: read and write, read only and unauthenticated access. Read and write and read only access require an S3 compatible client together with a key and secret to authenticate. A portal has been created by MGHPCC for the secure generation and retrieval of the key and secret. Read and write access is intended to be used only by the PI and their data managers to upload data. Read only access can be distributed to other collaborators who need access to download the data while keeping the data private from the public. Unauthenticated access allows anyone with the URL to download the data via HTTPS without any special clients nor keys. A hybrid access model can be achieved using S3 bucket policies[6] to allow unauthenticated access to a portion of the data while requiring a read only key and secret to access all of the data.

Client Compatibility

Any S3 client in which the end point can be specified is compatible with the OSN. AWS's official SDK for Python and official command line client are both compatible with the OSN. The OSN is

[6] https://docs.ceph.com/en/nautilus/radosgw/bucketpolicy/

also compatible with most popular command line clients, including rclone[7] and s3cmd[8]. The popular CyberDuck[9] GUI S3 client for Windows and macOS is also compatible with the OSN.

Data Management Frameworks

Data management frameworks that support accessing data via the S3 API are also compatible with the OSN. The OSN is partnering with the developers of data management frameworks including Dataverse, Clowder and Globus to ensure they are compatible with the OSN and provide the best user experience possible. More information appears under *Software Spotlights*.

Integration with XSEDE

The OSN is an XSEDE level 2 service provider. Requests for storage may be made through XSEDE's XRAS proposal system. This gives researchers an opportunity to understand and leverage OSN resources in a consistent and comprehensive fashion as part of a project data management plan that includes data sharing and distribution. Allocation requests between 10TB and 300TB can be requested through XSEDE. Startup allocations of up to 10 TB enable prospective projects to test interoperability with OSN, and are allocated on a rolling basis directly by the OSN. Projects needing more than 300TB should contact the OSN team directly to discuss their request. OSN allocation requests submitted via XSEDE may be stand-alone storage requests or coupled with requests for other XSEDE resources. Current OSN allocations via XSEDE include several projects in Physics, Biology and Chemistry, ranging in size from 10 to 200TB. New OSN allocation requests can be submitted by going to https://portal.xsede.org/submit-request#/ .

OSN Policy

Policy Development for Shared Infrastructures

Organizations building and managing shared infrastructure require policies that guide coordination, stakeholder and customer engagement, implementation planning, and end-user requirements. The term "guide" is key, as policies should be durable and require less frequent changes, emphasizing consistency. Policies provide a "frame" for procedures, which should be generated separately and document actions that support the practical functions of complex and distributed systems.

Cyberinfrastructure and other resources that will be provisioned require some initial policy "framing," and there must be consideration for governance processes to monitor and update the policies as a resource and its environments evolve. Policies take time to develop, and all policy work is iterative. This is necessary as social arrangements evolve, technical decisions are

7 https://rclone.org/
8 https://s3tools.org/s3cmd
9 https://cyberduck.io/

made, implementations are tested, and challenges solved. It is notable that in a distributed, low barrier infrastructure service like OSN, it is necessary to accommodate local variation while ensuring uniform global behavior.

Designing the Frame

When designing the policies, input from multiple stakeholder groups was taken into consideration: internal organizational groups (e.g. infrastructure and oversight teams); customers (i.e. data managers, data users); and external service partners and collaborators.

All policies are meant to embody these purposes:

- Simplify/accelerate access to data (in active use) by researchers
- Facilitate smooth flow of large data sets between sites
- Make it easier to expose large datasets with "long tail" characteristics to the scientific and scholarly community
- Compliance with applicable laws and regulations is required by all

Status

To ensure uniform global behavior, the OSN organization has developed "rules of engagement" for the team, OSN users, and 3rd party or partner services. As indicated above, the policies focus on the "what;" "how" is left to implementation decisions and regularized procedures. It is important to note that as an object store, the OSN does not support data that changes in real time (as in "live" databases), although snapshots from a live data stream would constitute an active data set. Further, this service is intended to support "Active Use" data (as defined earlier in the report), and does not provide preservation services nor support archival storage of datasets.

The OSN continues to build out and implement governance structures and functions that will be flexible to allow the OSN to scale. The OSN has completed the following policies:

- Overview and Definitions
- General policy (ver. 1)
- Guidelines for Hosting an OSN Pod
- Cybersecurity

OSN Cybersecurity policies were established and reviewed in consultation with the NSF TrustedCI CyberSecurity Center of Excellence (https://www.trustedci.org/)[10]. The allocation and use policies are still in progress at the time of this report and will be posted to the OSN website upon completion.

A recording of the webinar titled OSN Outcomes Update is available.[11]

---

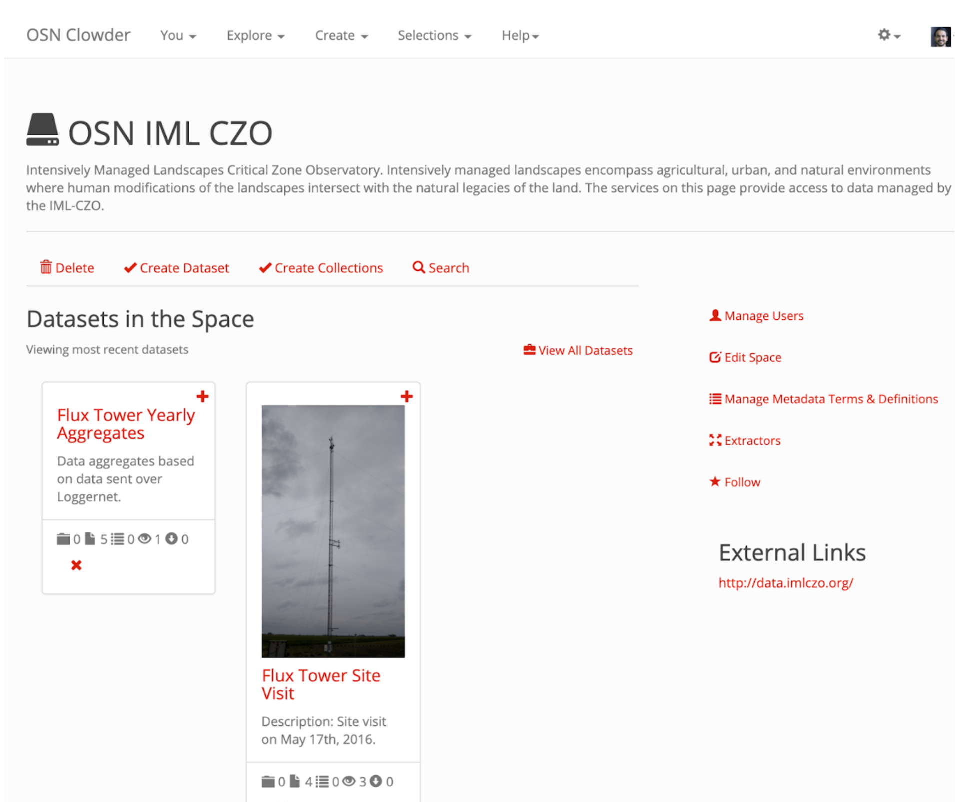
[10] https://blog.trustedci.org/search?q=OSN
[11] https://www.youtube.com/watch?v=xN2sH7m17lw&list=PLyckeMIsNe-_Lxa8jjgvmbcTWoa7519jT&index=1

Software Spotlights

The OSN infrastructure is meant to work in concert with a lab or community's platform already in use. The Pod design and S3 interface are useful in combination with a variety of existing platforms, many of which were developed with prior NSF investments. Clowder, Dataverse, and Globus are highlighted.

Clowder

With the increasing dependency on shared data as part of scientific research, there is a growing need for tools, in particular user friendly tools, to enable data discovery, organization, and analysis of large, heterogenous, and potentially interdisciplinary datasets. A number of scientific domains have developed such tools specific to their domains. Doing so, however, in this isolated manner has led to a great deal of duplication in effort, as well as challenges in terms sustainability of developed tools due to focus on a limited user base. The Clowder framework (https://clowderframework.org) attempts to address this. Having been built up over time in support of dozens of scientific efforts across every domain, Clowder supports the capabilities needed overall by the scientific community, from data still actively being used to published data. Clowder does so with an emphasis on customization, allowing every aspect of it to be tuned for a particular application from the metadata used, to the data formats supported, data visualizations, storage leveraged, and its user interfaces. Further, released under the NCSA (a BSD-3 like) open source license, Clowder, unlike many other domain specific solutions, can be taken and modified by anyone and deployed anywhere as desired, with support provided by a growing international community utilizing the software. As part of the Intensively Managed Landscapes Critical Zone Observatory (IML-CZO) use case described below, the deployment of Clowder on top of the OSN addresses an identified need within the CZOs



for bringing together community data collections and having a place to host them in an actively usable manner.

The IML-CZO (https://imlczo.org/) is an NSF funded activity to understand how land-use changes affect the long-term resilience of the *critical zone*, which is defined as the zone between treetops and bedrock of the Earth. It brings together scientists and researchers in a variety of earth science disciplines and from institutions across the Midwest. All data collected by the project is managed in a central repository https://data.imlczo.org/ developed and maintained by NCSA on project funded resources. Data produced by the project includes results of lab analysis, output of LIDAR surveys, and data streaming directly from sensors in the field, including a flux tower for atmospheric data, ISCO samples for water sampling, pressure transducers for logging water levels, soil sensors and many more. The data systems include an instance of Clowder and of the Geostreaming Data Framework (https://github.com/geostreams) to visualize data on the web in near real time. As of now, the Clowder instance contains 2.7 TB of data, 80K files, and 140 registered users. This work will continue in the new NSF CINet project (https://cinetcluster.org/).

As an Open Storage Network (OSN) use case, the deployment of the IML-CZO Clowder instance over data stored within the OSN was explored as an alternative to the project budgeted storage it currently uses. An instance of Clowder was established, running in Kubernetes on nearby NCSA resources that communicates with the OSN pod over the S3 API. Clowder is capable of using S3 buckets natively as a backend for storing data. The figure above shows the deployed instance with a couple of the CZO datasets migrated over to the OSN storage. Overall, the migration of the backend storage for the IML-CZO Clowder instance was greatly simplified due to the adoption of the commonly used S3 API by the OSN. This emphasizes that existing NSF technologies like Clowder can readily be employed as user-friendly front-ends to OSN hosted data. Fully deployed on the OSN, the IML-CZO could take advantage of the distributed nature of the OSN and its access to fast networks to aid in collecting and moving these community datasets regionally, in particular when analysis is desired, to a location in close proximity to a suitable computational resource. This would be particularly beneficial for large files, such as the output of the LIDAR survey.

Globus

Globus (www.globus.org) is a research data management platform that enables secure, reliable, and highly performant file transfer, sharing, and data management automation throughout the research lifecycle.

Globus is used by over 1,600 non-profit and commercial research organizations, national laboratories, and government facilities worldwide, including many supercomputer centers, scientific facilities, university campuses, and national e-research infrastructure providers. In 2020, the service had over 30,000 active, connected storage systems (“endpoints”) and over 170,000 registered users. Since its launch in November 2010, Globus has helped investigators move over one exabyte and 100 billion files.

Globus is developed and operated by a committed and experienced team at the University of Chicago. Globus is provided under a freemium model, with core authentication and transfer services freely available to all non-profit research institutions, and the vital issue of sustainability provided for by more than 150 paying subscribers, as well as by grant funding from the Department of Energy, National Science Foundation, National Institutes of Health, National Institute of Science and Technology, and Sloan Foundation.

Globus capabilities are accessible via an intuitive web interface (for casual users), from the command line (power users), and via REST and Python APIs (developers).

Globus capabilities are frequently integrated into data portals, science gateways, and other web applications that manage data distribution from instruments and provide access to reference datasets. For example: automating data egress (and analysis) from instruments such as cryo-electron microscopes and advanced light sources (e.g., TBs of Advanced Photon Source beam line data processed per day in COVID-19 research); making large corpora of data available to research communities (e.g., >3PB of data served by the NCAR Research Data archive to climate researchers worldwide); ad hoc data sharing in life sciences research (e.g. >2,000 shared endpoints created by researchers at the NIH); and reliable data migration for archival at massive scale (e.g., 2PB of irreplaceable data moved from the failed Arecibo observatory).

Globus broadens usability of OSN by making it simple for people to transfer data rapidly and reliably to and from OSN, and between OSN nodes, share data with collaborators, create data portals enabling search and retrieval of large data, and integrate OSN storage into science workflows. As an exemplar use case, we have developed a portal to enable discovery and retrieval of TerraFusion data hosted on OSN.

The success of the Globus platform teaches three lessons with broad applicability to research infrastructure, namely that (1) a high-quality research platform can transform science practice by automating previously intractable tasks, (2) modern cloud services allow a small professional team to build and operate such a platform, and (3) a subscription-based freemium model can allow such a platform to be sustained for the long term.

OSN believes that, given the opportunity, these lessons can be applied to a sustainability model that would enable OSN to grow into a robust, cost-effective, and performant national data-sharing resource that would substantially improve access to data for US researchers.

Dataverse

Dataverse is open source software used to build research data repositories at institutions and around the world. Once the software is installed, Dataverse makes it easy for researchers to deposit, curate, and share FAIR data. There are currently 70 Dataverse installations across 6 continents. These installations are associated with universities, research institutes, and some serve as national data repositories. In addition to the individual installations, the Harvard

Dataverse repository[12] is open for deposit by any researcher, and has 45,000 registered users. Dataverse publishes its user metrics at http://www.dataverse.org/metrics. Dataverse is an example of software already in use by many domains that can be layered onto the OSN for increased capabilities.

The capabilities of the Dataverse Software includes:

- Easy interface for researchers to deposit and share their data.
- The ability to upload any kind of file with special ingest available for some formats (e.g tabular data) and infrastructure to support larger data.
- Flexible metadata and permission systems to support many kinds of research domains and publishing workflows.
- The ability to connect to External Tools for previewing, exploring, and curating datasets (with upload and other functionalities on their way).
- A complete list at of features is at https://dataverse.org/software-features

Dataverse sees opportunities as they begin work on the Data Commons project at Harvard, as described in the presentation from Merce Crosas at the European Dataverse Workshop 2020[13], and they continue to work with the TRSA project (https://cyberimpact.us/). Both projects are focused on use cases that support large data throughout the research lifecycle.

Visions for the Future: The Role of Data Sharing and Distributed Storage in Research

The Trend Towards Mid-Scale Computing

Research in the US has traditionally been bimodal with funding going to lots of small projects and a few big projects. The big research projects are risk averse, use technology that was chosen a decade before the research starts and is built with its own vertically integrated software and hardware stacks. Today the trend is to fund mid-scale computing research in order to maximize science. Mid-scale computing focuses on creating a new unique instrument and taking risks. Mid-scale research projects are typically awarded to young researchers who are using cutting edge technology with a focus on collaboration and open data. Shoestring budgets are common for Mid-scale research projects and researchers are often required to search for cheap compute cycles wherever they can find them. Research data is copied and stored near these disparate compute resources, whether they are found on university HPC clusters, national supercomputer centers, science grids or commercial clouds. Each of these compute resources have different methods of access and different requirements for usage. Mid-scale computing needs access to reliable cyberinfrastructure for their research but doesn't have the budgets nor the expertise to operate it.

[12] http://dataverse.harvard.edu

[13] https://scholar.harvard.edu/files/mercecrosas/files/datacommons-tromso.pdf

The Commercial Cloud

One of the biggest developments in research computing cyberinfrastructure over the last 10 years has been the rise of the commercial cloud. Commercial clouds have become a popular source of cheap compute cycles due to their availability to sell their excess computing cycles provisioned in large facilities worldwide at low cost, and support for on-demand and burst computing. Proprietary compute hardware that is faster and cheaper than what can be purchased by universities is being developed by the commercial cloud providers so they can escape the Intel and NVIDIA monopolies in order to compete with each other on performance and price. The API driven approach to acquiring commercial cloud computing cyberinfrastructure allows researchers to easily move between commercial clouds when compute cycles become cheaper elsewhere.

While the commercial cloud has provided researchers with an excellent source of cheap compute cycles, they haven't been able to leverage the same techniques to provide cheap storage for researchers. Commercial cloud providers do not sell their excess storage at a discount and the on-demand and burst storage models can only be used when there is a second persistent copy of the data stored elsewhere. Proprietary storage hardware is not being produced by commercial cloud providers; they are forced to purchase hard drives from the same three manufacturers as everyone else. A large disadvantage of using the commercial cloud for data storage is data egress fees. Universities can typically transfer data between each other for free due to investments that have already been made by the US to create high-speed research networks. These free to use high-speed networks are critical to researchers who need to move their data next to cheap compute cycles. The commercial cloud providers use commercial networks, which can be expensive and the costs unpredictable due to complex charging models. Unless commercial cloud providers can peer with the private research networks to provide free data egress, the costs will likely remain a barrier to research that is dedicated to data sharing.

The Role of the OSN in Mid-Scale Computing

The federally funded cyberinfrastructure in the U.S. has been built on three pillars: compute, high-speed networking and data-intensive infrastructure. Large sums of money have been spent to provide compte resources and high-speed networking, but the funding for data-intensive infrastructure, specifically data storage, has not met the needs of researchers of mid-scale computing.

The instruments funded in mid-scale research projects often generate huge amounts of data. The amount of data generated causes a “First Meter” problem, in which an instrument needs to be coupled with network throughput and storage that can sustain the data flow from the instrument. In order to capture the data off of the instrument without slowing the instrument down, the storage has to be close to the instrument, have low latency and scale horizontally, which makes it very expensive. The OSN helps solve the first meter problem because it is geographically distributed in national supercomputer centers and advanced computing centers near compute resources, is on high-speed and low latency networks and can support horizontal scaling by allowing simultaneous reads and writes to multiple OSN pods.

Data access patterns for mid-scale projects have changed due to the investments in high-speed research networking. Before there were high-speed research networks, large datasets were distributed in total on physical media. With the proliferation of research networks, researchers were able to download subsets of the data to their labs for their own research. Now the current research networks contain low latency, 100-gigabit-per-second Ethernet hardware which enables researchers to stream data directly to their compute resources from the source, saving time staging data locally and requiring less directly attached storage on the compute resources. Mid-scale researchers are already using the OSN to take advantage of the high-speed research network on which the OSN pods are strategically distributed to stream OSN data to their compute resources.

Finally, the OSN uses cloud storage APIs that make it simple for mid-scale researchers to integrate the OSN into their cloud native workflows. The OSN is compatible with S3 APIs which allows researchers to easily transition from the commercial cloud to the OSN and from the OSN to the commercial cloud. Supporting cloud computing APIs allows mid-scale researchers to adopt the OSN once their commercial cloud bill exceeds their budgets or move from the OSN to the commercial cloud should they find a compelling research use case to have their data on the commercial cloud.

A recording of the webinar titled The Role of Data Sharing and Distributed Storage in Research - Visions for the Future is available[14].

Conclusion

This paper describes the advantages of an open storage network for large scale data storage, and results from a flexible, scalable Open Storage Network pilot (OSN pods), demonstrating its advantages over extant storage solutions and its ability to accommodate the growing needs of the research community through several cases. For future growth of OSN, continued engagement with the scientific computing community has been demonstrated to be essential for developing policies to fulfill their needs. Using input from stakeholders to ensure uniform global behavior, the OSN continues to build out and implement governance structures and functions that will be flexible to allow the OSN to scale and remain a distributed, readily accessible infrastructure. To date the OSN has completed policies addressing:

- Accommodation of local variation while ensuring uniform global behavior
- Coordination, stakeholder and customer engagement, implementation planning, and end-user requirements
- Rules of engagement for the team, OSN users, and 3rd party or partner services
- Cybersecurity

[14] https://www.openstoragenetwork.org/seminar-series/apr-29-2021-data-sharing-and-distributed-storages-role-in-research-next-steps/

As developed, the OSN provides a solution to several barriers, one of which is as a response to the lack of cost-effective, high-performance storage for researchers. While the commercial cloud can provide researchers with an excellent source of cheap compute cycles, the disadvantages of using the commercial cloud for data storage are data egress fees and potential vendor lock in. The commercial cloud providers use commercial networks, which can be expensive and the costs unpredictable due to complex charging models. Unless commercial cloud providers can work with private research networks to provide free data egress, the costs will likely remain a barrier to research that is dedicated to data sharing. The OSN uses common cloud computing APIs that make it simple for researchers to integrate the OSN into their cloud native workflows and allows researchers to easily transition from the commercial cloud to the OSN and from the OSN to the commercial cloud as their needs and budgets demand.

Distributed storage is increasingly used in response to research trends, particularly mid-scale research. The OSN is uniquely positioned to address data management challenges encountered in mid-scale research. Additionally, the OSN is geographically distributed in national supercomputer centers near compute resources, sits on high-speed and low latency networks, and supports horizontal scaling by allowing simultaneous reads and writes to multiple OSN pods. Mid-scale researchers are already using the OSN to take advantage of the high-speed research network on which the OSN pods are strategically deployed to stream OSN data to their compute resources.

The OSN is a boon to computing on the network, particularly for the growing needs of mid-scale research. With growing focus on collaboration and open data, institutions need access to cyberinfrastructure for their research community without necessarily having the budget to acquire and manage this data-intensive infrastructure. OSN pods provide a solution that can help fill a gap in high performance and big data applications, building in performance and cost effectiveness that is competitive with, or exceeds current solutions. Significantly, the OSN is a viable solution for research groups that span multiple institutions that cannot host shared data locally across institutional firewalls and security policies for data sharing. Research projects will increasingly need an option like the OSN to support transdisciplinary and at-scale science. Our experience in working with a broad range of researchers across a large variety of disciplines has demonstrated that the need for an easy to use, cost-effective, distributed data storage platform, such as the OSN, is here today and will continue to grow as small- to mid-scale projects continue to proliferate.