
OSN

Release 0.1

csim

Nov 07, 2022

CONTENTS

1	Contents	3
1.1	System Overview	3
1.2	Allocations	5
1.3	Accessing Datasets	5
1.4	Landing Pages	9
1.5	Open Access & Protected Datasets	9
1.6	Managing Files and Data	12
1.7	Transfer Data to the OSN	12
1.8	Help	12

The Open Storage Network (OSN) is a distributed data sharing and transfer service intended to facilitate exchanges of active scientific data sets between research organizations, communities and projects, providing easy access and high bandwidth delivery of large data sets to researchers.

The OSN serves two principal purposes: (1) enable the smooth flow of large data sets between resources such as instruments, campus data centers, national supercomputing centers, and cloud providers; and (2) facilitate access to long tail data sets by the scientific community. Examples of data currently available on the OSN include synthetic data from ocean models; the widely used Extracted Features Set from the [Hathi Trust Digital Library](#); open access earth sciences data from [Pangeo](#); and Geophysical Data from [BCO-DMO](#). These data sets are being used by researchers to machine learning models, validate simulations, and perform statistical analysis of live data.

Note: This documentation is under active development.

CONTENTS

1.1 System Overview

OSN data is housed in storage pods interconnected by national, high-performance networks creating well-connected, cloud-like storage that is easily accessible at high data transfer rates comparable to or exceeding the public cloud storage providers, where users can temporarily park data, for retrieval by a collaborator or create a repository of active research data.

This user guide is designed for the following categories of OSN user:

- End Users who wish to view metadata and retrieve data.
- Data Curators who maintain data sets
- Data Managers who grant access to data sets for Curators and End Users

1.1.1 Configuration

Key characteristics of OSN storage are:

- Ability to access data from anywhere via a RESTful interface that follows S3 conventions
- Federated identity management, allowing access to protected information with existing identity via InCommon or commercial services
- High speed access and transfer via national research and education networks
- Security and data integrity

OSN storage pods are located in science DMZs at Big Data Hub sites, interconnected by national, high-performance networks. 5 petabytes of storage are currently available for allocation.

1.1.2 File Systems

OSN Storage is disk based and primarily intended to house active data sets. OSN storage is allocated from the pod(s) closest to the requestor with capacity to fulfill the request. Allocations of a minimum 10 terabytes and a maximum of 50 terabytes can be requested through the XRAS process. If your project needs more than 50 terabytes, please contact the OSN team directly to discuss before you submit your request.

The OSN supports two types of data sets:

1. Open Access Data Sets that are readable by anyone and writable by Curators and Data Managers.
2. Protected Access Data Sets that are readable by invitation from a data manager and writable by Curators and Data Managers.

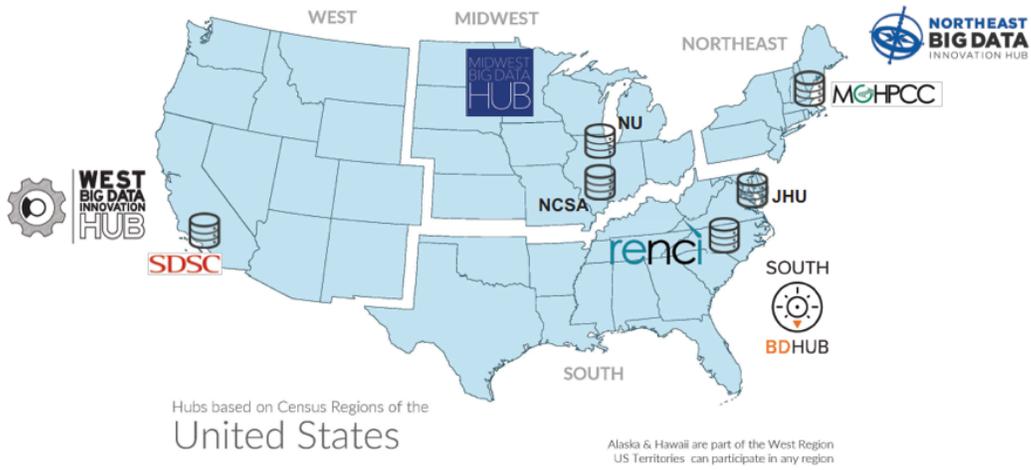


Fig. 1: OSN Pod Deployment at six sites as of January, 2021

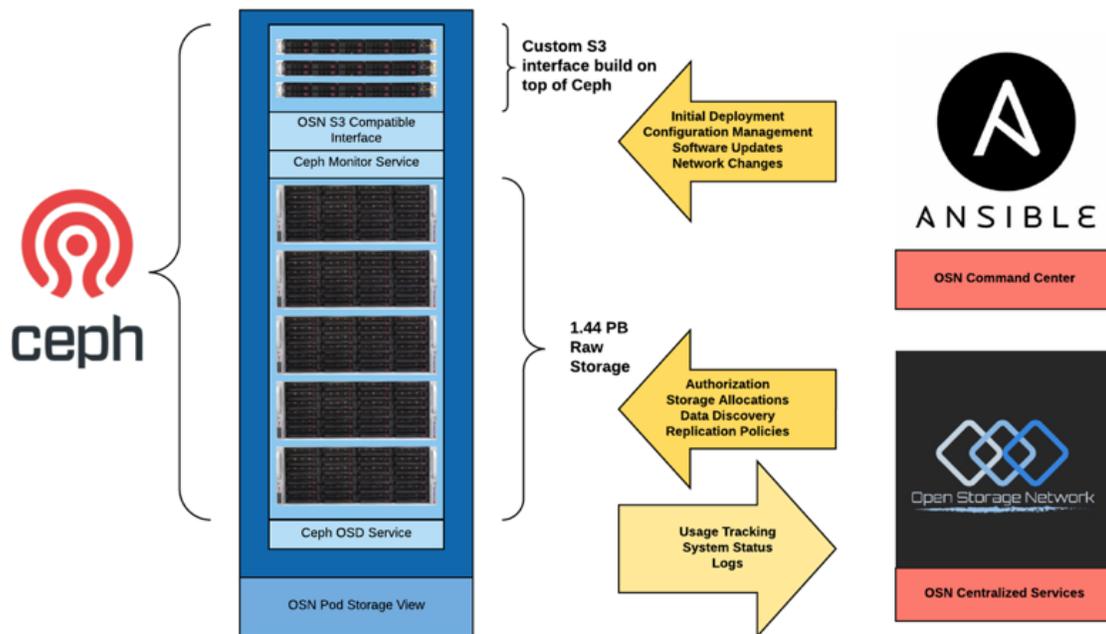


Fig. 2: OSN Storage Pod

Every data set is a collection of objects that are individually and uniquely accessible from anywhere. For Open Access data sets, an S3 RESTful interface allows users to manipulate storage objects simply by issuing commands in the form of Uniform Resource Identifiers.

For Protected Access Data Sets, the user first obtains an access key which is then embedded into the access command. Examples of each are provided below.

Coming soon: Consistent with FAIR principles, every OSN data set will have a landing page that makes it easy to “visit” a data set from a browser, search engine, or data catalog. The landing page contains metadata that describes the data set, along with the links to preconfigured, downloadable tools for accessing the data.

An active research data set can remain in OSN storage up to five years and usage must comply with the OSN Acceptable Use Policy.

1.2 Allocations

Storage on the OSN is allocated in standalone buckets independent of HPC allocations. There is a one-to-one mapping between buckets and allocations. This User Guide uses “Allocation” when referring to outward-facing operations such as Allocation requests, and “Bucket” when referring to inward-facing operations such as Bucket creation.

OSN storage is allocated from the resources at the location(s) closest to the requestor with capacity to fulfill the request. Allocations of a minimum 10 terabytes and max of 50 terabytes supporting up to 1.6 million files can be requested through the XRAS process. Larger allocations can be accommodated with additional review. If your project needs more than more than 50 terabytes or more than 1.6 million files, please contact the OSN team directly to discuss before you submit your request.

An active research dataset can remain in OSN storage up to five years.

1.3 Accessing Datasets

OSN supports a RESTful API that is compatible with the basic data access model of the [Amazon S3 API](#). Any software that complies with that API can access data stored on the OSN.

There are three common methods for connecting to and using OSN resources:

1. OSN portal built-in web tools
2. Third party desktop applications (e.g. Cyberduck, Rclone)
3. Third party data management server applications (e.g. Globus and iRods)

1.3.1 OSN Portal Built-in Web Tools

The OSN portal (portal.osn.xsede.org) supports a simple UI that allows end users to browse allocations and to upload and download objects via the browser. This mode of access is most appropriate for browsing a dataset and uploading/downloading smaller files (typically <100G).

To use the built-in browser, a user logs onto the OSN portal and clicks on one of the allocations that they have been granted access to. This brings the user to a searchable/sortable table listing of the allocation and its subdirectories. Clicking on any of the objects shown initiates a download of the object to the local disk.

To upload a file, the user locates the file on their local filesystem and drags the file to the browser window. This initiates an upload to the bucket location that the user is currently browsing.

OSN Basic Bucket Browser

Storage Portal Home Admin ▾ James Culbert (culbertj@mit.edu) ▾

Ceph Bucket Explorer

jimtes2    151

Show 25 entries Search:

Select	Object ↑↓	Folder ↑↓	Last Modified ↑↓	Timestamp ↑↓	Class ↑↓	Size ↑↓
	Chlorophyll/					
<input type="checkbox"/>	darwin_v0.2_cs510_BiologicalRates_Denit.nc	/	7 days ago	2020-09-25 14:57:59	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0a.nc	/	6 days ago	2020-09-26 02:11:46	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0b.nc	/	6 days ago	2020-09-26 02:11:13	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0c.nc	/	6 days ago	2020-09-26 02:11:03	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0d.nc	/	6 days ago	2020-09-26 02:11:07	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0e.nc	/	6 days ago	2020-09-26 02:11:41	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0f.nc	/	6 days ago	2020-09-26 02:11:05	Standard	1 TB
<input type="checkbox"/>	darwin_v0.2_cs510_Chlorophyll_TRAC0g.nc	/	6 days ago	2020-09-26 02:11:13	Standard	1 TB

Fig. 3: OSN Basic Bucket Explorer

1.3.2 Third Party Applications

There are numerous commercial and open source software tools for moving files to and from S3 buckets. These tools provide more sophisticated capabilities than the built-in browser tool including transfer management, multi-upload management, and provide configuration options that can help optimize data transfer for a given computer/network environment.

To use these tools, you will need to retrieve a pair of keys that are used to access the buckets stored on OSN. To retrieve these keys, you can contact your data manager and she will either give you keys or create an account for you on the OSN portal where you can retrieve these keys. If your data manager creates a portal account for you and gives you access to the keys you can visit [OSN Portal](#) to retrieve them; the allocations you have access to and their associated keys will be listed on your home page.

Note that the “Bucket” information displayed in the portal has two components (this will be important when you configure third party tools). The bucket information contains the OSN site/pod location and the specific allocation on that pod.

Storage Management Portal

Welcome to the storage management site.

You are a member of the projects shown below.

You may manage groups that you are a datamanager for.

- [AYSE \(manage\)](#)
- [JIMTEST \(manage\)](#)
- [OSNDEMO \(manage\)](#)
- [SCOTTPROJECT \(manage\)](#)
- [TWOSIGMA \(manage\)](#)
- STORAGEADMINS

You have access to the following keys.

Type	Bucket	Access	Secret
read/write	https://mghp.osn.xsede.org/osndemo1	716PCBCHM5LM1JT9BVAE	huPRC9yrDTTgY/Y1enhJcQg+Howy
read/write	https://minio.mghpcc.org:9000/jimio	jimio_datamanager	DHCRYnOhsZoseOxNa3NZ7olpGAQw+Y
read/write	https://mghp.osn.xsede.org/twosigma1	QZ6KC404RE0QS1NJIU8E	24PhTdWwHoqS6ulzRbPsXoqymlexeH
read/write	https://mghp.osn.xsede.org/mghpbucket	RZ0YD49FQFOMMOFNB78A	cq4gAD2EB3OFgNLzLWmRnV6Ce/yTCb
read/write	https://jhui.osn.xsede.org/jhubucket	7I0114S07FAYVN471MGF	aOCiww7LEzXzXQukKJd3MXI2gb1XIJ
read/write	https://ncsa.osn.xsede.org/ncsabucket	7Z1206I4R76TTIO4JOB2	DAhrxmVtt5jkkAOw8wM18mOLpHhheM
read/write	https://renc.osn.xsede.org/renbucket	L536KG5NL68YH1BXRJZA	NsyZfjRu8ZNOQWmmqZormtnnEYfsoU

Fig. 4: OSN Portal User Home page

Cyberduck

Cyberduck is a popular file transfer tool that supports the S3 API. The following describes how to configure Cyberduck to connect to an OSN resource. Cyberduck is a “cloud storage browser” for Mac and Windows that supports multiple storage providers/protocols. The software may be downloaded at: [The Cyberduck Download Page](#)

Using Cyberduck with OSN is straightforward.

1. Visit the OSN portal to retrieve your Bucket location and allocation names (see image below)
2. Visit the OSN portal and retrieve your allocation keys or retrieve them from the data manager for your project
3. Open Cyberduck and select the bookmarks icon (see image below)
4. Click the add icon at the bottom left of the screen to create the bookmark
5. Edit the new bookmark to point at the desired OSN pod using you allocation key pair



When specifying the server, use the hostname portion of the location (i.e. if the location is <https://mg hp.osn.xsede.org> the hostname is “mg hp.osn.xsede.org”).

When specifying “Port”, use 443 if the location starts with “https://”; use 80 if the location starts with “http://”.

Anonymous Access Data Sets

Some datasets provide anonymous read access; if you are accessing buckets anonymously, type “anonymous” into the Access ID portion and Cyberduck will then select the grayed out anonymous access box in the window.

Exit the window for the bookmark to save.

Browsing, Uploading, and Downloading

Once a bookmark is created, you can use it to access data by double-clicking the bookmark. This logs your user in and lists the contents of the dataset.

Note: If your buckets have large object counts, you will need to increase the Timeout settings for connections.

Go to Preference > Connection and change the box next to Timeout for opening connections (seconds) and change the setting to 90 seconds.

Cyberduck client is a full-fledged transfer client so desktop up/downloads can be easily performed for data sets.

The tool supports multiple upload/download streams, chunking, pausing and restarting.

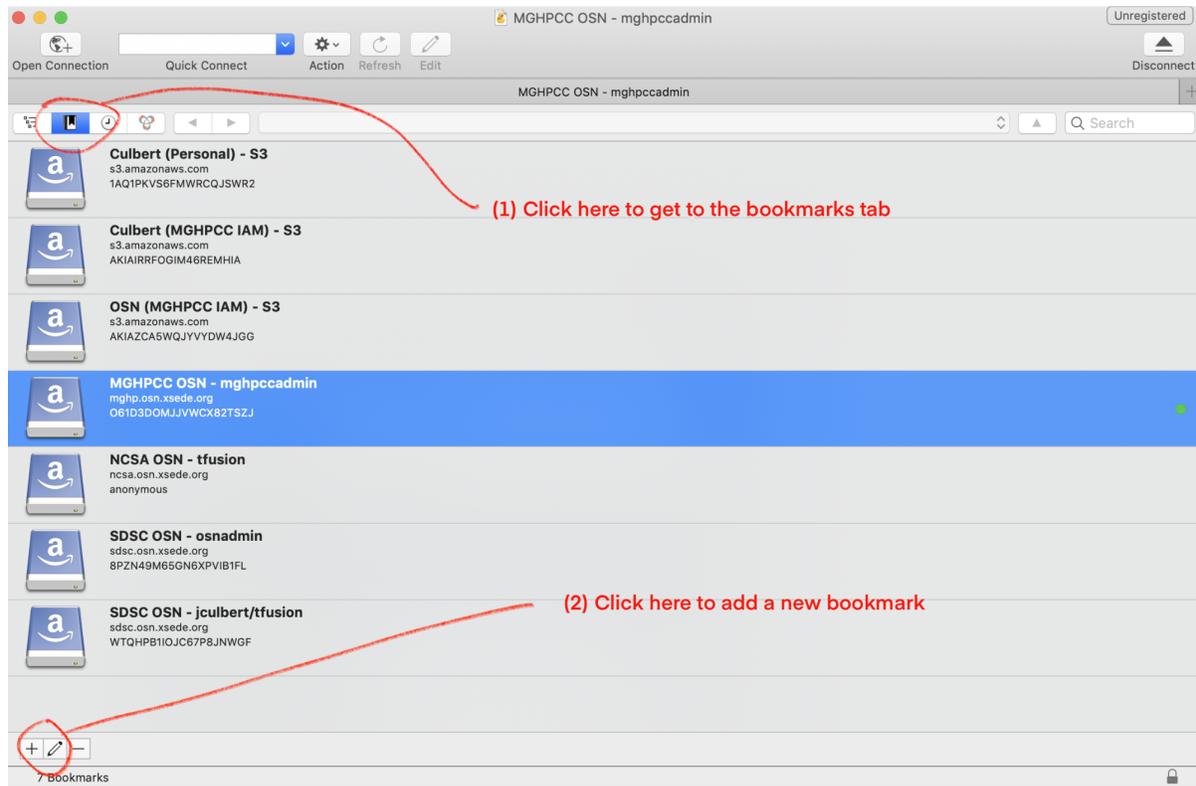


Fig. 5: Selecting the bookmarks page and adding new bookmark

Rclone

Rclone Configuration

Rclone Commands

1.3.3 Third Party Data Management

1.4 Landing Pages

blerg lorem ipsum

1.5 Open Access & Protected Datasets

blerg lorem ipsum

OSNDEMO Allocation

Amazon S3

Nickname: OSNDEMO Allocation

URL: <https://mghp.osn.xsede.org/osndemo1>

Server from location here → Server: mghp.osn.xsede.org Port: 443

Access and secret keys from datamanager or portal here → Access Key ID: KJHASD897asdkjh

Secret Access Key: Secret Access Key

SSH Private Key: None

Client Certificate: None

▼ More Options

Bucket (with / prepended) here → Path: /osndemo1

Web URL: <https://mghp.osn.xsede.org/>

Download Folder: Downloads

Transfer Files: Default

Timezone: UTC

Encoding: Default

Connect Mode: Default

Notes:

?

Fig. 6: Adding OSN pod and user information to bookmark

The screenshot shows a window titled "OSNDEMO Allocation" with a header bar containing three colored circles (red, yellow, green) on the left. Below the header is a search bar with the Amazon S3 logo and the text "Amazon S3". The main configuration area includes the following fields:

- Nickname: OSNDEMO Allocation
- URL: <https://mghp.osn.xsede.org/osndemo1>
- Server: mghp.osn.xsede.org
- Port: 443
- Access Key ID: anonymous
- Anonymous Login
- Secret Access Key: Secret Access Key
- SSH Private Key: None
- Client Certificate: None

A red annotation "Type 'anonymous' here" with an arrow points to the "Access Key ID" field.

Below these fields is a section titled "More Options" with a downward arrow. It contains the following fields:

- Path: /osndemo1
- Web URL: <https://mghp.osn.xsede.org/>
- Download Folder: Downloads
- Transfer Files: Default
- Timezone: UTC
- Encoding: Default
- Connect Mode: Default
- Notes: (empty text area)

A help icon (question mark in a circle) is located in the bottom left corner of the window.

Fig. 7: Using anonymous access as your user

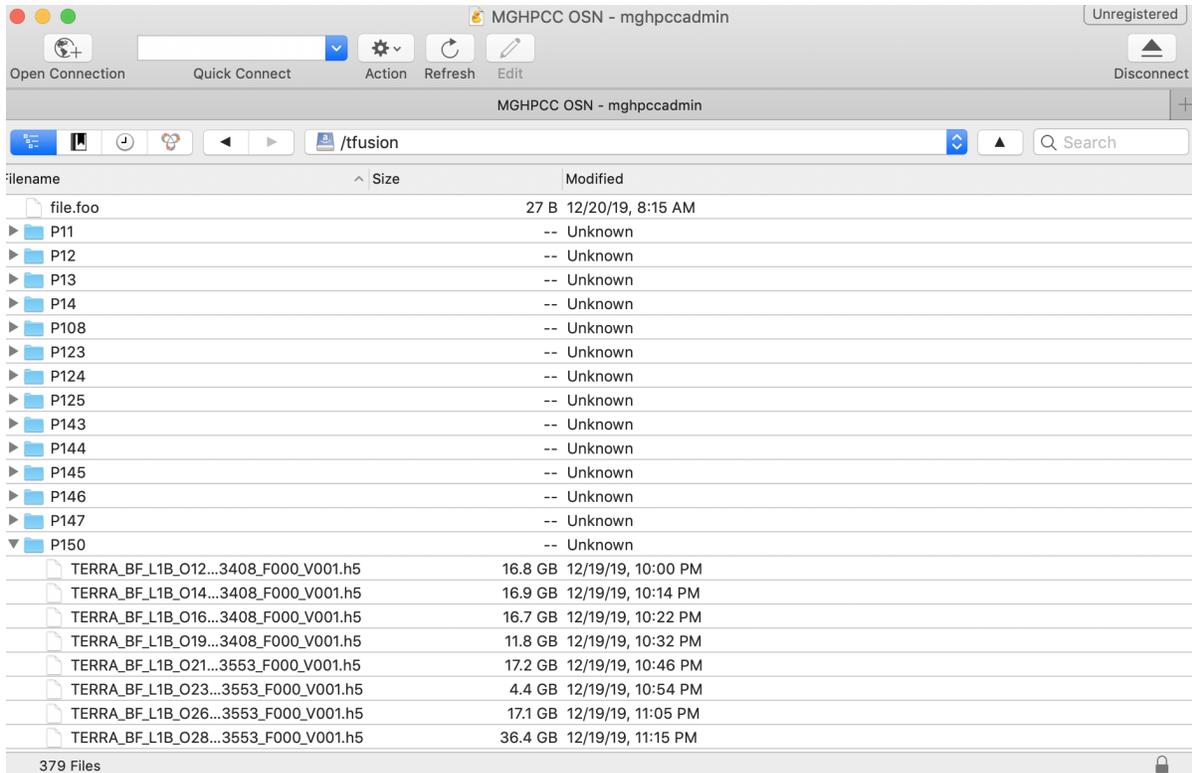


Fig. 8: Directory listing within bucket

1.6 Managing Files and Data

blerg lorem ipsum

1.7 Transfer Data to the OSN

blerg lorem ipsum

1.8 Help

blerg lorem ipsum