

# Research Data Facility

## Research Data Facility

To accommodate our research community's growing need to store large datasets and facilitate collaboration among research groups, the Center for Research Computing has made available a variety of services collectively known as the Research Data Facility (RDF). The RDF consists of a combination of cloud-based and on-premises storage services that are robust and secure, flexible enough to meet a wide range of use cases, and scalable to meet future data storage needs.

## Security

The RDF is designed to securely accommodate non-regulated data. If your data requires [additional security precautions](#) to ensure regulatory compliance, please contact the CRC to discuss your requirements.

## Rice Box

Many researchers require a secure data archive or collaborative space where datasets can be shared easily with colleagues both inside and outside of Rice, but there is not a need for direct, real-time access to the data from software applications. The Rice Box service is designed to meet those requirements. Rice Box is:

- A cloud storage solution based inside the US
- Accessible via your Rice NetID and password
- Allows files/folders to be easily shared to Rice colleagues
- Allows links to files to be sent to external collaborators' email addresses
- Supports FTP bulk data transfer  
Currently, limited to commercial internet speeds – the CRC recommends consulting with us about integrating cloud storage solutions into workflows or archiving practices that involve large-scale datasets
- Can be synchronized to local folders
- Unlimited, free storage

To get started, visit [Getting Started with Rice Box](#) for more details

Start using Rice Box today - Go to <http://rice.app.box.com>

## Isilon

Some research workflows may require more interactive, larger scale network storage that can be accessed in real-time by users and applications. For these use cases, the CRC can allocate network shares from our Isilon storage appliance to faculty researchers, with a 500GB subsidized allocation and cost recovery for additional utilization beyond the 500GB limit. Dell EMC/Isilon is:

- Dell/EMC clustered storage appliance
- Highly redundant, can tolerate multiple disk failures/head failures without data loss  
(See our note on [disaster recovery and backup services](#) below)
- On-premises, managed by CRC
- Authenticated using your Rice NetID
- Mapped to your client computers as a shared network drive
- Real-time, shared access to data
- **Please note:** RDF Isilon is suitable for workstations or laptops which need access to an interactive, shared storage area for research data. It is *not intended* for high performance workloads, use on multiuser systems or servers, or in support of server infrastructure.

### RDF Isilon Subsidy Eligibility

Research storage allocations will be granted to all tenured faculty, tenure-track faculty and research faculty as defined by [Rice Policy 201](#). The first 500GB of storage will be subsidized, and utilization above the subsidized level will be charged back to the researcher. Research storage allocations for other research groups not defined above will be handled on a case-by-case basis.

### RDF Isilon Charge Back Rates

The current rate for RDF-Isilon is \$70/TB/year (7 cents/GB/year). The storage amount billed in a month is based on the average storage space used throughout the month. Your storage usage is measured in "GB-Month," which are added up at the end of the month to generate your monthly charges. For example:

- A user requests a new allocation in May 2018, and their May average utilization is 290GB. Since their utilization is below 500GB, there would be no charge for May
- The user's average utilization in June 2018 rises to 950GB during a major project. The monthly charge for June would be:  $(950-500) \times .07/12$  or \$2.56
- The same user finishes their project and removes several large datasets, so their average utilization for July goes down to 600GB. The monthly charge for June would be:  $(600-500) \times .07/12$  or \$.57

## Quotas and Scaling

When a researcher is granted an allocation, a warning (soft) quota will be set at 450GB (90% utilization), and a hard quota will be set at 500GB. When the soft quota is reached, the system will send a courtesy email to the researcher warning them that they are getting close to the hard quota. The hard quota will ensure that the user does not exceed their subsidy and receive unexpected charges. When the hard 500GB limit is reached, an additional email notification will be sent to the researcher, and a help desk ticket will be opened to alert the CRC. The researcher may then authorize that the quota be removed, with the understanding that they will be billed for any monthly average use above 500GB. Once the quota is removed, RDF-Isilon storage allocations will scale automatically to meet the demands of users, without additional intervention. However, to help researchers manage the costs of their storage, they may optionally request the CRC to set soft or hard quotas at a higher limit.

## Access to Shares (SMB/NFS)

The RDF-Isilon shares fully support authenticated SMB (CIFS) storage allocations

- Authentication using NetID/NetID password and Active Directory credentials
- OIT retains full administrative rights, to manage permissions on behalf of researchers
- OIT will grant read/write/modify permissions to researchers' shares based on Active Directory group membership
- SMB/CIFS is supported on Mac and PC platforms, with limited support available for Linux distributions
- CRC **strongly recommends** against using SMB for the following applications:

Supporting Unix/Linux network infrastructure (i.e., shared home directories, shared applications)

Multi-user Unix/Linux systems

For detailed instructions on how to access RDF shares via SMB, please see the [RDF User's Guide](#).

RDF-Isilon shares can be provisioned as NFS v3 shares on a case-by-case basis when SMB is not feasible due to client operating system limitations or application limitations.

- NFS v3 is unauthenticated, and carries a higher security risk since share permissions can be overridden by a local user with administrative rights
- NFS v3 shares can be compromised by an NFS client spoofing an IP address which has been granted access to the share
- Users should be aware that NFS v3 shares cannot be accessed by any systems which are not included in the access control lists
- Mixed Mode (i.e., SMB plus NFS v3 on the same share) is not supported
- Shares can be provisioned as NFS v3 once the researcher and systems administrator of the client machine acknowledges and accepts full responsibility for the security risk
- Client machines requiring access to NFS v3 shares must be added to access control lists maintained by CRC/OIT
- To request an NFS v3 share, a Request Tracker ticket should be filed with the help desk to document approval and acknowledgement of the security risks by the research PI and the systems administrator of the client system

## Backup and Disaster Recovery

Snapshots

- Each RDF-Isilon share will have nightly snapshots enabled by default
- Users can locate files in the snapshot directory and restore them within the 24 hour snapshot window
- Snapshots are *not* a backup or disaster recovery tool, they are intended to allow users to recover accidentally removed files from the last snapshot

- Data preserved by snapshots are included in utilization averages for cost recovery
- Snapshot policies can be customized to meet researchers' individual requirements, through consultation with CRC

#### Fee-Based Backup/Disaster Recover Services

- Backup to Amazon S3 Glacier cloud service
- Off-site, disk-to-disk disaster recovery
- Researcher is responsible for associated AWS Glacier charges. For current AWS Glacier charges, see: <https://aws.amazon.com/glacier/pricing>

Note that Rice University's Amazon S3 Glacier vaults are located in the US East (Ohio) region

***To request an RDF-Isilon Network share, please use our web form to [Request Help with the Center for Research Computing Resources](#)***

#### Data Transfer

*Coming Soon:* Globus Connect/Science DMZ

Globus Connect offers researchers a method of transferring large data sets between participating research institutions. CRC plans to integrate the RDF/Isilon appliance as a Globus Connect endpoint, with access to the facilities of the high-speed Science DMZ-Internet2 on-ramp