

OURRstore (OU & Regional Research Store)

- Overview
 - Costs to use OURRstore
 - Restrictions on data that can be archived on OURRstore
 - How to request storage on OURRstore
 - How data is stored on OURRstore
- Using OURRstore
 - Archiving data to OURRstore
 - Simple option for data on Grace (possibly the research NAS)
 - Simple option for data on ROSS
 - Slightly less simple option for data on Grace (possibly the research NAS)
 - Complicated Option
 - Retrieving Data from OURRstore
 - Simple option for data archived from Grace (possibly the research NAS)
 - Simple option for data archived from ROSS
 - Slightly less simple option for data archived from Grace (possibly the research NAS)
 - Complicated Option

Overview

The OURRstore system, an NSF-funded project, is a very cost-effective way to archive data that needs to be reliably kept. OURRstore uses LTO media, which has a bit error rate 10 times lower than spinning disk, with an expected lifetime of at least 8 years. (The tapes theoretically can last longer if properly stored, but the hardware that reads the tapes might not.) OURRstore increases the reliability and safety of storage by creating redundant copies of the data. OURRstore procedures creates at least 2 copies of archived data, though we strongly recommend using the 3 copies option. One copy stays in the OURRstore robot, and is the 'online' copy, retrievable at any time. A second copy gets shipped back to UAMS, where we store it in an environmentally controlled, physically locked location, giving an offsite backup of our data stored in OURRstore. In a pinch, we can recover data from the copy sent to us. The optional (but highly recommended) third copy is taken out of the robot and stored in an environmentally controlled storage facility at the University of Oklahoma Health Sciences Campus. This third copy allows them to generate replacements for the primary copy should that be necessary without the risk of shipping the backup copy kept at UAMS to Oklahoma.

OURRstore is the least expensive, most reliable archival storage option available to UAMS basic science researchers.

Costs to use OURRstore

The cost to use OURRstore is only the media costs (extremely low compared to other options) plus a small additional amount for shipping cartridges to and from Oklahoma. All the equipment and management costs are covered by NSF and OU. All a researcher has to do is make certain that their research project, their lab, their department, or someone has provided cartridges and return shipping materials to the OURRstore team. The HPC Admins will assist researchers, labs, and departments in ordering the cartridges, to make certain that they get credited appropriately.

The last batch of 100 cartridges that we sent to OURRstore cost \$67 each, and cost approximately \$135 to ship 100 cartridges via Fedex Ground, fully insured, to Oklahoma. USPS Priority Prepaid Forever small boxes to return individual cartridges (the second offsite copy) cost \$8.45 currently. For the recommended 3 copy option this worked out to a one-time charge of \$213.50 for approximately 7.6 TB usable triple redundant storage that should keep data safe for at least 8 years (\$28.09 per TB). The less resilient (higher risk) 2 copy option would be a \$145.15 one time charge for the same 7.6 TB, minimum 8 year storage (\$19.10 per TB). These prices are expected to drop over time as we qualify less expensive vendors. We are also exploring other shipping options. Due to UAMS procurement policies the cartridges first have to be received at UAMS, then shipped to OURRstore. Then after filling them with data redundantly, OURRstore ships 1 out of every 2-3 cartridges back to UAMS. Hence we pay 3 to 4 increments of shipping for every 2 to 3 cartridges, depending on the storage option chosen. The above numbers reflect all these estimated shipping charges.

For comparison to other storage options, this works out to \$3.51 per TB per year for triple redundant storage currently, with pricing continually dropping, compared to, for example, Amazon Glacier Deep Archive (their least expensive option) at about \$12 per TB per year plus any data egress and networking charges. (Often the networking and data egress charges are more expensive than the storage costs for Amazon.) And often cloud storage providers charge extra for keeping data redundantly in multiple storage locations, increasing the cost. For OURRstore, geographically separated redundancy is included in the costs. And as media prices drop, and capacities rise, the costs per TB for OURRstore will only go down.

More importantly, with OURRstore you purchase tape cartridges as a one-time, up-front payment good for OURRstore's lifetime, instead of paying monthly recurring charges as you would with most cloud storage providers. So when a particular grant expires and the money ends, your files are still safe and sound. With a monthly storage service charge, cloud storage services will delete your files if you run out of funds to pay for the storage.



If you decide to use OURRstore for archiving your data, please be prepared to provide for the media and shipping costs.

Here are a few suggested options for covering the media and shipping costs for using OURRstore:

- Researchers should include the archiving costs in their grant budgets. Since most grants now require archiving and data sharing plans, this is a reasonable expense to include in a grant budget. Of course, ongoing grants might not have budgeted for archiving costs, so would need to look for other options. Pilot research projects also might not have a budget for archiving costs. **Including archiving costs in grant budgets is the preferred funding mechanism, as it does not require UAMS funds.**
- Research labs could include costs for archiving in their annual lab budgets.
- Departments might be willing to cover the costs for archiving data as part of their departmental budgets, for example as infrastructure costs (i.e. part of indirects).

Due to the nature of how OURRstore works, the data you archive could end up on any cartridge, or even be spread across multiple cartridges. Therefore, when you purchase cartridges, you are actually just purchasing an allotment of archival storage that is **added** to your quota, commensurate with the usable capacity of the cartridges, divided by 2 or 3, depending on which storage option that you choose (double or triple redundant).

Please coordinate with us on the shipping address and the specific model of LTO cartridges allowed. In fact, we are willing to put together the order for you, which might simplify things for you. OURRstore will not accept unapproved cartridges, and may not accept cartridges that are not shipped with an appropriately addressed shipping label. In addition, one has to provide to OURRStore materials for return shipping, including pre-paid shipping labels and arranging for pickup. Everything has to be exactly right in order to add cartridges to the OURRStore system. Currently only the following specific models of cartridges are allowed:

- IBM 38L7302 [LTO-7, formatted as Type M, 9 TB raw, ~7.6 TB usable]
- IBM 01PL041 [LTO-8, 12 TB raw, ~10.2 TB usable]



Please do not ship cartridges to OURRstore without consulting with the HPC Admins. The OURRstore team may reject and discard or return anything that does not meet their criteria for cartridge type, labelling, addressing, etc. or that is not associated with a valid OURRstore account.

We cannot guarantee that your data is stored on the cartridges that you purchase. Instead, the usable storage capacity of the cartridges would be added to your quota for you to draw from, and your cartridges would be put into the pool of cartridges that we use for everybody. If you need tighter controls (i.e. your data and nothing else goes on the cartridges that you purchase), then you would need to use the "complicated" option for archiving data to OURRstore mentioned below (i.e., you would need to get your own OURRstore account and manage your own archiving, shipping and storage instead of using the pooled services that the HPC admins provide).

Restrictions on data that can be archived on OURRstore

Being an NSF project, there are certain stipulations on the kind of data that can be placed in OURRstore. OURRstore is intended for **NON-CLINICAL STEM RESEARCH DATA** that is **NOT LEGALLY REGULATED**. Non-STEM data is currently **FORBIDDEN** on OURRstore, because OURRstore was funded by the NSF, and non-STEM data is outside the NSF's mandate.

1. The data should be relatively static (i.e. does not change), as OURRstore is only intended as a robust, resilient archive, not a backup solution where one is making daily or weekly copies of changing data. (You may use ROSS, the Research NAS, or a cloud option if you need backup.)
2. The data must be STEM related data (Science, Technology, Engineering, Math). NSF's definition of STEM includes physical sciences, biosciences, geosciences, engineering, mathematics, technology (for example, computer and information sciences), and social sciences.
3. While the data may include deidentified human subject data, it may not be clinical research data (i.e. data directly related to patient care or clinical studies of human disease). If the human research is basic science research, that is acceptable.
4. Legally regulated data (for example, HIPAA, Controlled Unclassified Information, FDA clinical trial, ITAR/EAR, FERPA) is currently **FORBIDDEN** on OURRstore, per their agreement with NSF.
5. If your files are subject to one or more Institutional Review Board (IRB) agreement(s) governing human subjects research, then it's **YOUR RESPONSIBILITY** to ensure full compliance with your IRB agreement(s).



If you decide to use OURRstore for archiving your data, you must insure that your data complies with the above rules.

How to request storage on OURRstore

If you have a means for covering the costs and agree to the data restrictions of OURRstore, please send a request via e-mail to hpcadmin@uams.edu confirming that you agree to the terms, and requesting access. We will then work with you in archiving your data.

How data is stored on OURRstore

Data stored on OURRstore should be collected into compressed archive files, preferably between 20 and 200 GB in length, for the best storage efficiency without excessive access times. Currently, the absolute minimum size of an archived file is 1 GB. The absolute maximum size is 1 TB. These archive files need to be created at UAMS prior to electronic transfer to the OURRstore system. The initial transfer is disk to disk, hence goes pretty quickly. Once the data is in the OURRstore disk cache, the OURRstore archive management software will start copying the data onto a media cartridge for safekeeping. When a cartridge is full, the system makes a copy of the cartridge, ejects it from the system, and the OURRstore team, using the prepaid label that we provide them, ships that copy back to us. We then store that copy offline in a locked, environmentally controlled location in Arkansas. If the optional third copy is requested, the OURRstore system makes that third copy, which is ejected from the system and stored in an environmentally controlled location in Oklahoma.

Using OURRstore

Archiving data to OURRstore

Because of the need to collect data into bundled, compressed archive files, we offer several options to assist users in archiving data. In the simplest option, you just create a directory tree or bucket where you place the data to be archived, and the HPC admins take care of bundling up the data for OURRstore. Or you can create the compressed archives yourself, and then ask the HPC Admins to transmit to OURRstore using the pooled account. Power users who want complete control of the process are welcome to work directly with the OURRstore team to get trained in using OURRstore, get a private account, and manage the archiving process themselves (not recommended, but possible).

Simple option for data on Grace (possibly the research NAS)

For the simple option, all you need to do is collect your data to be archived into a sub-directory tree with just the files to be archived. Please move the sub-directory under a parent directory named "ToBeArchived" in your home directory. Please name (or rename) the sub-directory tree to be archived with the current date in "yyyy-mm-dd" format, for example "/home/john/ToBeArchived/2021-08-07/". The sub-sub-directory tree under the dated subdirectory can be organized any way you see fit. Please use the "mv" command, not "cp" or "rsync" to collect data, since you eventually want the data to disappear from Grace once it is safely in OURRstore, and you don't want to run into a space crunch while organizing your archive directory. (Remember, Grace's storage is only supposed to be a temporary holding place for running jobs.)

Then send a request via e-mail to hpcadmin@uams.edu, giving us the name of the archive subdirectory you want archived. The HPC admin team will then convert that subdirectory tree into a set of compressed, encrypted, multi-volume tar files, broken into blocks of appropriate sizes for transmission to OURRstore. If the data that you are archiving is smaller than a 20 GB archive file, the HPC admin team may combine your data with other data to maintain archiving efficiencies. Before transmitting the tar files to OURRstore, the HPC admin team will create manifests of the tar files, complete with listings of the directory tree being archived. They will place a copy of these manifests in your /home/<username>/archived directory as a record of what you archived when, to aid in retrieval. The names of these manifests will include the name of the subdirectory archived from your "ToBeArchived" folder, for example, "/home/john/archived/2021-80-07-<id>.manifest", where <id> is the block number (i.e. when splitting up what would be a large tar file into smaller pieces). You will be notified when the HPC Admins have confirmed that the archive files are safely tucked away on OURRstore. At this point the HPC admins (or you) may remove the archived files from the "ToBeArchived" directory.

This method can also be used for data on the research NAS, assuming that the HPC admins can get access to that storage area on the NAS. For data on the NAS, just create the "ToBeArchive" directory somewhere in your storage area, and let the HPC admin team know the storage area and path to the directory when you submit the archive request.

Simple option for data on ROSS

For data on ROSS, the simplest way to archive that data is to collect the data to be archived into a bucket. When ready for archiving, you could optionally alter the permissions on the bucket to read-only to minimize the chance of accidental changes, if you wish. Then send a request via e-mail to hpcadmin@uams.edu giving them the name of the bucket to archived and the namespace that the bucket lives in.

The HPC Admins will then pull data from the bucket into compressed, encrypted tar files broken into blocks of appropriate sizes for transmission to OURRstore. If the data that you are archiving is smaller than a 20 GB archive file, the HPC admin team may combine your data with other data to maintain archiving efficiencies. Before transmitting the tar files to OURRstore, the HPC admin team will create a manifest listing the objects in the tar files. They will then notify you, sending you a copy of that manifest when they have confirmed that the tar files have been safely archived on OURRstore.

For this option, the HPC Admin team will not delete the bucket just archived. It is acceptable to keep data both in ROSS and in OURRstore, though we suggest in this case to not use replication on ROSS, which doubles your storage charges. If you wish to free up space on ROSS (e.g. to avoid hitting your quota, or to avoid further charges), you are free to delete the bucket yourself once you get confirmation that the data is safely on OURRstore.

If you prefer to create your own compressed archive files ready for OURRstore you could simply create them and place them in a bucket. In this case, when you e-mail hpcadmin@uams.edu the name of the bucket and namespace, let them know that you have already generated the compressed archive files in the bucket. Or alternatively, you could use the "Slightly less simple option for data on Grace", temporarily placing copies of your archive files on Grace.

Slightly less simple option for data on Grace (possibly the research NAS)

If you would rather create the compressed tar files yourself, feel free to do so, and then collect the archive files into the top level of your "ToBeArchived" subdirectory in your home directory. In this option, feel free to use any method of your choosing (e.g. tar, zip, or some custom format) that can collect the data you want archived into files. The ideal choice should keep the archive files between 20 to 200 GB in length, though OURRstore will accept anything between 1 GB to 1 TB in length. ***We encourage you to use compression and encryption for efficiency and safety***, but that is your choice. If you *do* encrypt, please safeguard your encryption key, since no one but you likely knows it. (The HPC admins would not know your encryption keys, for example.) In this option you are responsible for creating your own manifests of the content of your archive files, if desired. The names of the archive files must be globally unique. In other words, do not name any 2 archive files with the same name. They are all going into a single directory in OURRstore, so none of the names of any of the archive files that you create can clash with the name of any archive file that you previously created. Otherwise you run the risk of losing the previous archive file (i.e. it could get overwritten).

Once you have your archives ready, send a request via e-mail to hpcadmin@uams.edu, giving the HPC Admins the names of the archive files that you want transmitted to OURRstore. They will take care of transmission to OURRstore and will notify you when they have confirmed that the archive files are safely tucked away on OURRstore. At this point the HPC admins (or you) may remove the archived files from the "ToBeArchived" directory.

This method can also be used for data on the research NAS, assuming that the HPC admins can get access to that storage area on the NAS. For data on the NAS, just create the "ToBeArchive" directory somewhere in your storage area, and let the HPC admin team know the storage area and path to the directory when you send the archive request e-mail.

Complicated Option

To exercise the complicated option, you would need to approach the OURRstore team directly, sign agreements, and go through the mandatory training to get your own account on OURRstore. You would then be responsible for following all of their rules, for purchasing your own media, for shipping to and from Oklahoma, for creating and tracking your own archive files, etc. This option really is for the power user who wants full control over the process of archiving and retrieving data with only minimal or no assistance from the the HPC admin team. This option is also appropriate for users whose data is on systems that the HPC Admins do not have access to. While we do not encourage people to use this option due to the complications and responsibility of getting your own account on OURRstore, it is a possibility for those who prefer. For more information, see [OURRstore: OU & Regional Research Store](#).

Retrieving Data from OURRstore

How to retrieve archived data depends on which of the above options you used to archive it.

Simple option for data archived from Grace (possibly the research NAS)

When the HPC Admins archived data for you, they left manifests of which files or objects are in what compressed tar file in your in your /home/archived directory. You can search through those manifests (e.g. using grep) to find which archive file or files the data you are interested in is located. Send a request via e-mail to hpcadmin@uams.edu with the name of the subdirectory tree that you would like retrieved and the name of the archive file or files it is in.. If you lost the manifests, don't fret. The HPC Admins kept a backup copy and can help. In the case of a lost manifest, still send a request via e-mail to hpcadmin@uams.edu with the name of the subdirectory tree that you would like retrieved and the approximate date that it was archived. The HPC admins will do their best to find the archive file names in their manifest backup copies. However, since yours is the primary copy, please do not lose it, as there is always a chance that the backup copy gets lost as well.

Once the HPC admins receive your e-mail, they will pull the pertinent archive files from OURRstore and restore the archived data to your /home /<username>/RestoredArchives directory. In general, your data should be restored within 1 business/work day (i.e. things might not get restored on weekends and holidays). The HPC admins will then notify you by e-mail that your data are restored.

Simple option for data archived from ROSS

Simply send a request via e-mail to hpcadmin@uams.edu with the name of the bucket that you want restored, and a name prefix of the objects that you want restored. Leave the name prefix blank if you want the entire bucket restored. You should also include the namespace where the bucket should be located.

Once the HPC admins receive your request, they will restore the objects into the bucket in the namespace, and notify you when it is ready.

Slightly less simple option for data archived from Grace (possibly the research NAS)

In this option, since you created the archive files yourself, the HPC Admins did not create manifest files. It is up to you to keep track of what data is in which archive file. When you want to retrieve one of your archive files, send a request via e-mail to hpcadmin@uams.edu with the names of the archive files that you would like retrieved.

Once the HPC admins receive your e-mail, they will pull the pertinent archive files from OURRstore and restore the archive files to the top level of your /home/<username>/RestoredArchives directory. In general, your archive files should be restored within 1 business/work day (i.e. things might not get restored on weekends and holidays). The HPC admins will then notify you that your archive files are restored.

Once you get notification that your archive files have been restored, you may then use whatever means you chose to pull data from those files. Don't forget that you could have compressed and encrypted the archive files before you archived them. Remember that the HPC Admins would not know the encryption key if you encrypted the files before archiving, and cannot help if you have lost it. So do make certain when you created encrypted archive files that you safely store your encryption key.

Complicated Option

You are in complete control of your retrieval of archived data, since the archived data is on your OURRstore account. The HPC admins are not involved.