

Storage

Managing Data on the CoSTAR Cluster

Learning Goals:

- Storage Areas
- Data Transfer
- Best Practises



Storage Overview

The CoSTAR Cluster provides two main storage areas:

Type	Filesystem Path	Quota	Retention / Notes
Home directory	/users/<username>	233 GB	Personal working space
Scratch space	/scratch4weeks/<username>	500 TB (shared total)	Active job data, high I/O, temporary files, fast data processing. Deleted after 4 weeks of inactivity

Key Takeaways:

- Data in Scratch4weeks will be **removed after 4 weeks** after last use (read operation or creation).
- **No user data is backed up.**



Data Transfer: GUI methods - 1 of 2

For users who prefer a **graphical user interface** (GUI), the following are the easiest ways to move files between your local computer and the CoSTAR cluster.

Open OnDemand (OOD):

1. **How:** Log in to the CoSTAR Open OnDemand.
2. **Features:** Use the "Files" dropdown in the top menu to easily upload and download files directly in your web browser.
3. **Best For:** Quick edits, managing scripts, and transferring small to medium-sized files.



Data Transfer: GUI methods - 2 of 2

SFTP Clients (MobaXterm, WinSCP, FileZilla, Dolphin):

1. **How:** Connect to `costar-login01.surrey.ac.uk` (Port 22) with your University of Surrey credentials.
2. **Features:** Provides a familiar drag-and-drop interface to move files and folders.
3. **Best For:** Users comfortable with FTP clients, especially on Windows.



Data Transfer: Command line methods - scp

For more advanced control and efficiency, you can use **scp** (secure copy) command-line tools in your local terminal:

Use Case: Ideal for transferring single files or small directories.

Example:

```
# Copies a local file to your CoSTAR home directory
```

```
scp my_file.csv <username>@costar-login01.surrey.ac.uk:/users/<username>/
```



Data Transfer: Command line methods - rsync

A versatile alternative to scp is the command-line tool

rsync (remote sync):

Use Case: The recommended tool for large directories, synchronising changes, or resuming interrupted transfers. It is quite fast as it only copies the differences.

Example:

```
# Syncs a local directory to your CoSTAR scratch directory
rsync -avz /path/to/folder \
<username>@costar-login01.surrey.ac.uk:/scratch4weeks/<username>/
```



Best Practices for Storage – 1 of 3

Follow these rules to ensure your work runs smoothly and efficiently.

- **Always work in scratch:** Run your computational jobs so that they write data to your `/scratch4weeks/<username>` directory. Reading and writing files here avoids filling up your home directory.
- **Copy results back:** Scratch is temporary and **not backed up**. Once your job is finished, copy your important results and log files from scratch to your local machine for permanent storage.
- **Avoid Sensitive Data:** Do not store personally identifiable or other sensitive research data on the cluster.



Best Practices for Storage – 2 of 3

- **Clean As You Go:** Scratch is a shared resource. Regularly remove large intermediate files, old datasets, and failed job outputs that you no longer need. Files in scratch may be automatically removed after a period of inactivity.
- **Monitor Your Usage:** Keep an eye on your storage quotas, especially in your home directory. Use these commands to check your usage:
 - `du -sh ~` (Check home directory usage)
 - `ncdu` (Provides an interactive view of directory sizes)



Best Practices for Storage – 3 of 3

- **Fix Windows Line Endings:** If you edit a script on Windows and it does not run on CoSTAR, it may have incorrect line endings. To fix it, use the following command:

```
dos2unix <your_script_name>
```

- **Organise your work:** We recommend you create a new directory in scratch for each project or job. This makes it easy to manage your files and clean up later.

