# Documents/code (not for research data)

## Personal

C drive

* Only use for temporary files (dispensable) unless they are backed-up (see “backup” section below)

M drive (provided by UoL)

* 20Gb
* Backed up by UoL
* Can sometimes be slow, or can (rarely) lose access. Maybe best to use for backing up docs/code from specific folder on C drive (e.g. automatically using software such as AOMEI Backupper) rather than as main store.
* Docs accessible from elsewhere via AppsAnywhere

Cloud (e.g. personal Dropbox, Google Drive, Github)

* Backed up by 3rd party
* Not secure enough for research data
* Can be more convenient than using AppsAnywhere when accessing from elsewhere

## Project-specific

Sharepoint site

* 5Gb per site
* For study documents
* Shared with other investigators
* Backed up on cloud, can recover old/deleted versions
* Sync to local PC via OneDrive
* Secure

# Research Data

C drive

* Normally 400Gb free
* For Secondary data (analyses in progress). No need to store Primary (raw) data here (can use NAS – see below).
* Not directly shareable (until copies are placed on NAS).
* Best to have daily auto-backup of “Data” folder on C drive to an external hard drive, e.g. using Windows backup utility, while analyses specifically for sharing can go on the NAS.

Network-attached storage (NAS)

* 5.5TB (automatically backed up to secondary drive within NAS)
* For:
  + primary (raw) research data
  + analyses (for sharing and backup)
  + code (for sharing) – although better to use Github instead.
* Shared with other investigators (who have permission)
* Secure (encrypted) folders for each project
* Use:
  + Each project has a unique folder containing the following sub-folders: “Primary”, “Secondary” and “Completed” folders, for raw data, analysed data (for sharing and backup) and completed analyses respectively.
  + “Primary” data should be placed here immediately after data collection. Record of data stored here should be kept. Data is further backed up (Active DataStore – see below)
  + “Secondary” folder should contain a comprehensive copy of all analyses to date that are complete enough to be shared. Likely these will be updated prior to project meetings (e.g. monthly). Secondary data is normally a copy of data already on PC, either for sharing or to temporarily store if PC capacity is limited.
  + “Completed” analyses are those that are ready to be written up for publication. These will be backed up to Active DataStore (below) during the project, and later backed-up to online repositories.
  + Deletion of data should not be permanent until the administrator empties the recycle bin.

Active DataStore (UoL facility)

* 1TB (expandable) per project
* For medium-term backup during project:
  + primary research data
  + completed analyses
  + NOT for intermediate analyses (due to lack of capacity)

# Regular backup and long-term storage

1. C drive:
   1. Secondary (anonymised) research data (analyses in progress: Daily automatic backup to external HD (attached to PC)
   2. Other docs/code (not research data): on a day-to-day basis, backup to M drive and/or sync to cloud storage (for code, recommend Github).
   3. At the publication stage, final versions of code (i.e. contributing to the paper) can be organised into a separate Github folder/repository for sharing purposes.
2. NAS:
   1. Primary data and Completed analyses: auto-backup daily to Active DataStore
   2. Secondary data: Backed up (instantly) within NAS only (second NAS drive). Normally this data will already be on PC (and therefore has another copy anyway) but may sometimes contain data not in PC due to C drive capacity limitations. This is acceptable as long as the secondary analysis is easily recoverable from a combination from Primary data and analysis code (both backed up).
   3. When the project ends, secondary analyses (if not part of the final “completed” analysis – see below) either be deleted or, if there is a possibility of future need, stored elsewhere (e.g. on an external HD) without further backup.
3. Active Datastore:
   1. Active DataStore copies (of primary and completed analyses) will eventually be deleted as this storage only lasts for as long as the project (current end date is March 2022).
   2. Before project ends, primary data, completed analyses and code will be shared publically via online repositories to ensure their longevity.