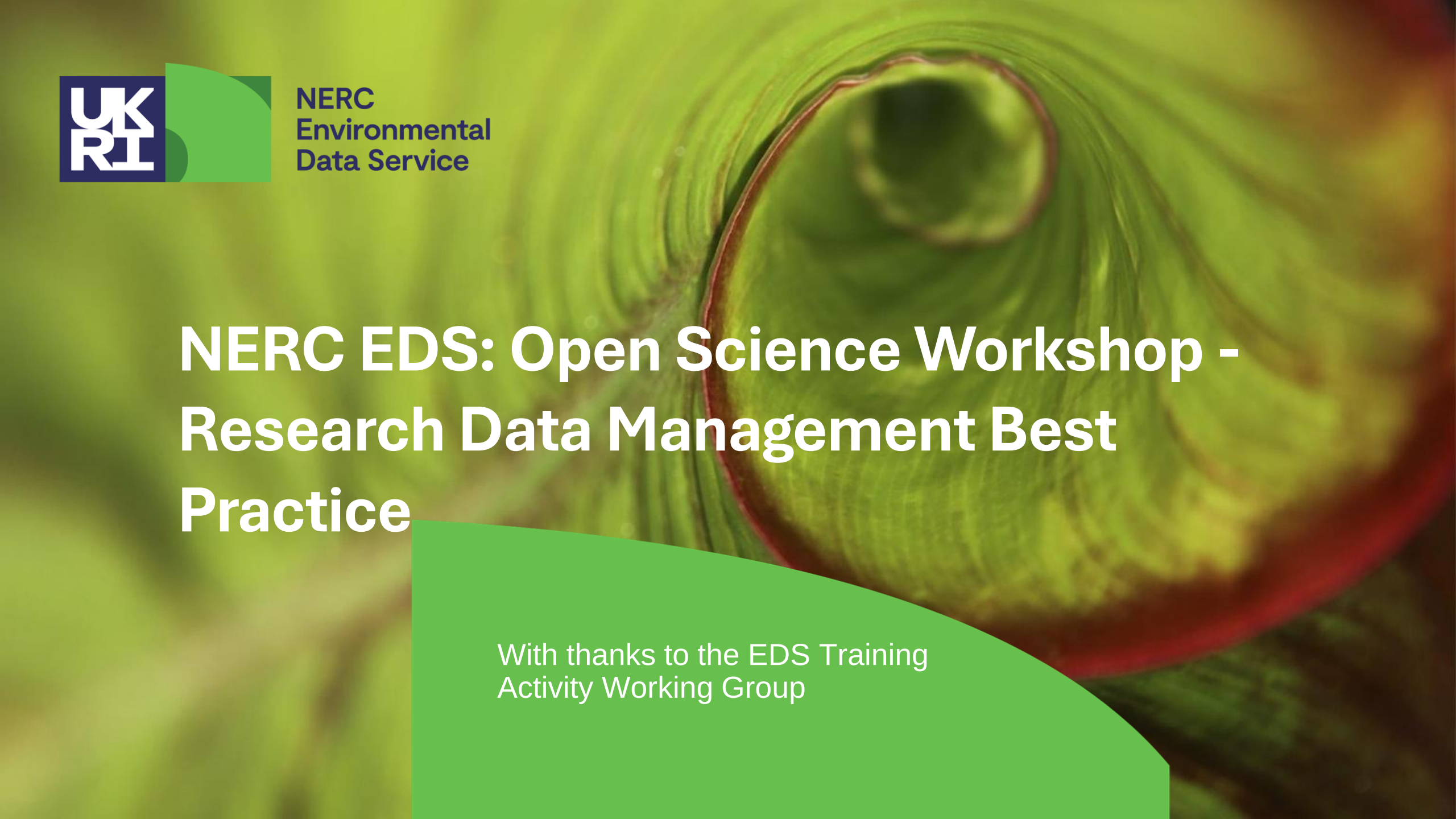


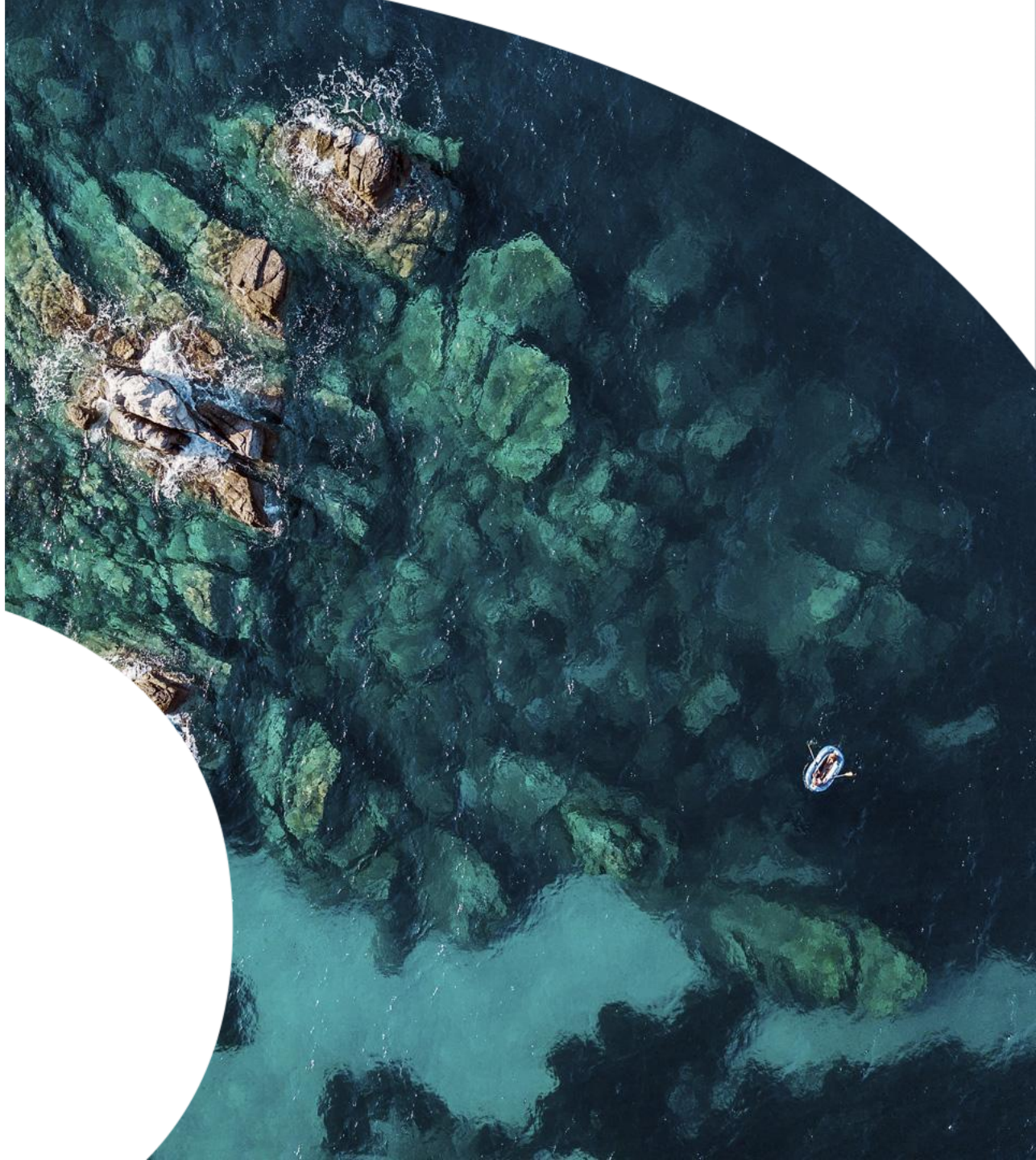
NERC Environmental Data Service

NERC EDS: Open Science Workshop - Research Data Management Best Practice

With thanks to the EDS Training Activity Working Group

Content

- Data Life Cycle
- Data Storage and Backup

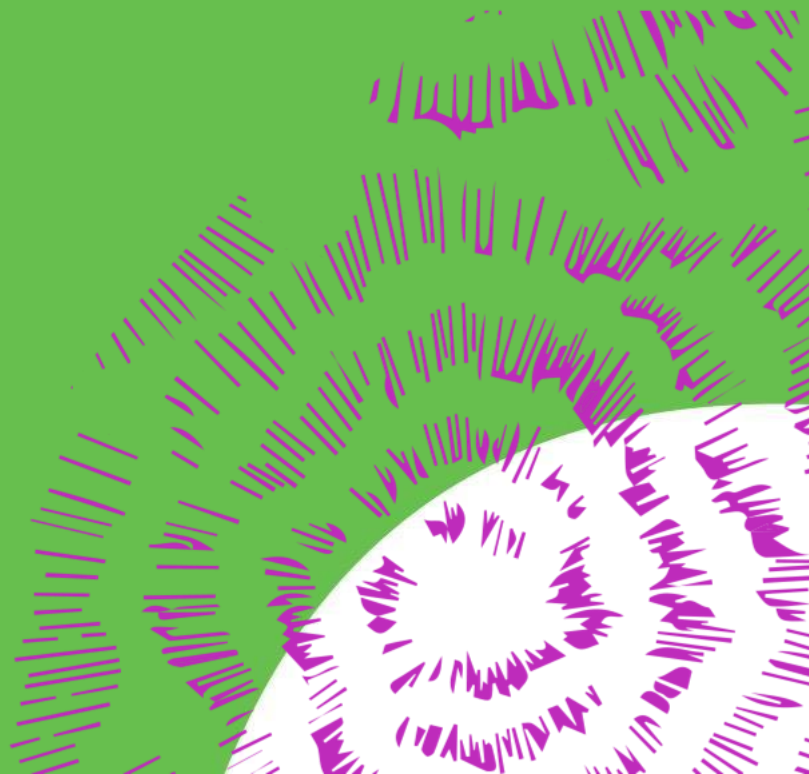


NERC Environmental Data Service

NERC Environmental Data Service

The Data Life Cycle

In this section, you will learn how to organise your data through planning, collection, analysis, publication and beyond

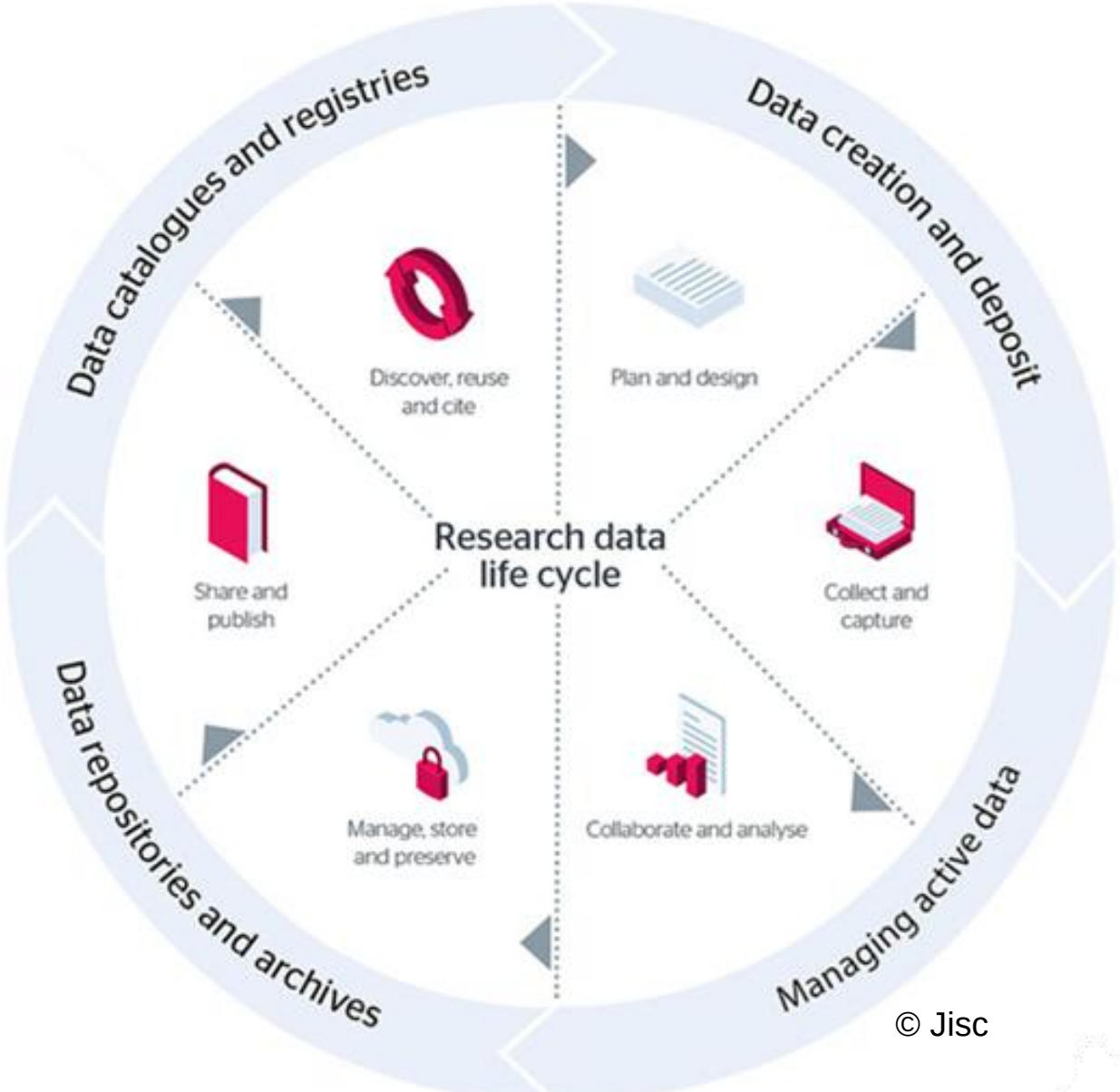


NERC Environmental Data Service

Research Data Management Lifecycle

The data life cycle is the sequence of stages that a particular unit of data goes through from its initial generation or capture to its eventual archival and/or deletion at the end of its useful life.

1) Data Management planning
2) Data collection and capture, collection and analysis
3) Data storage and archiving, sharing and publishing
4) Data cataloguing, discovery and reuse



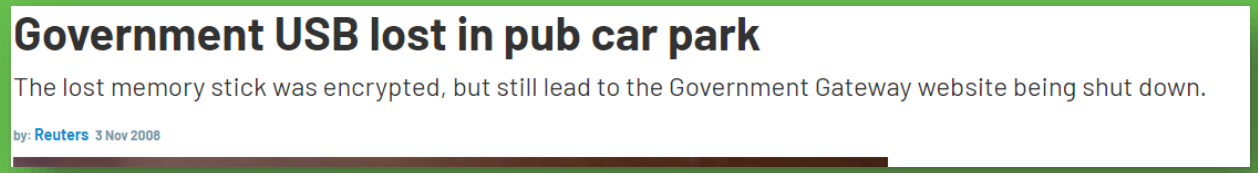
NERC Environmental Data Service

NERC Environmental Data Service

Storage and backup

Reason for data loss

- Failing hardware
- Software or media faults
- Virus and damaging malware
- Fires and natural disasters
- Computer theft
- Human error – spilling coffee



Day-to-day: Backing up data

Making regular backups is an integral part of data management. You can backup data to your personal computer, external hard drives, or departmental or university servers. Software that makes backups for you automatically can simplify this process considerably.

BACK UP OPTIONS

Back up options are context specific:

- **Cloud storage should be the first choice.** Institutional options exist. Do not hesitate to ask your colleagues.
- If without access to the cloud, the 3-2-1 rule is useful for remote data collection.
- Not always possible with huge datasets: ask your institute about other server options for large datasets.
- Conversely – for third-party data, avoid redownloading multiple times. Consider accessing via an API.
- Consider using Git for code versioning.

3-2-1 rule:

- Keep at least 3 copies of each important file
- On at least 2 different devices or storage media
- 1 of which should be off-site

Discuss options with your colleagues or your data centre

Data back up

Back-ups are not preservation!

Back-ups:

- Used to take periodic snapshots of data in case the current version is destroyed or lost
- Copies of files stored for the short-term
- Often performed on a somewhat frequent schedule
- Subject to change, no guarantee of their longevity

Data Archive and Preservation

Archiving and preservation = long-term storage

Archive:

- Used to preserve data for historical reference or potentially during disasters
- Archives are the final version, stored for the long-term and generally not copied over
- Often performed at the end of a project or during a major milestone
- Safeguarded and preserved

NERC Environmental Data Service

Tips for digital preservation

- Open vs proprietary formats
- Choose open-source where possible
- Digital obsolescence
- Refresh and migrate file formats often

NERC Environmental Data Service