

Ensure integrity and accessibility when making backups of data

 dataoneorg.github.io/Education/bestpractices/ensure-integrity-and

Data Life Cycle stage(s): Preserve

For successful data replication and backup:

- Users should ensure that backup copies have the same content as the original data file
 - Calculate a checksum for both the original and the backup copies and compare; if different back up the file again MD5: algorithm to determine check sum
<http://en.wikipedia.org/wiki/MD5>
 - Compare files to ensure that there are no differences
- Document all procedures (e.g., compression / decompression process) to ensure a successful recovery from a backup copy
- To check the integrity of the backup file, periodically retrieve your backup file, open it on a separate system, and compare to the original file
- A data backup is only valuable if it is accessible. When access to a data backup is required, the owner of the backup may not be available. It is important that others know how to access the backup, otherwise the data may not be accessible for recovery. It is important to know the “who, what, when, where, and how” of the backups:
 - Have contact information available for the person responsible for the data
 - Ensure that those who need access to backups have proper access
 - Communicate what data is being backed up
 - Note how often the data is backed up and where that particular backup is located including
 - physical location (machine, office, company)
 - file system location
 - Be aware that there may be different backup procedures for different data sets:
 - Not all backups may be located in the same location
 - Depending upon the backup schedule, each iteration of the backup may be located in different locations (for example, more recent backups may be located on-site and older backups may be located off-site)
 - Have instructions and training available so that others know how to pull the backup and access the necessary data in case you are unavailable

Description Rationale

For successful preservation a backup data file should contain the same information as the original.

Additional Information:

Related Best Practices

Tags

Cite this best practice:

DataONE Best Practices Working Group, DataONE (July 01, 2010) "Best Practice: Ensure integrity and accessibility when making backups of data". Accessed through the Data Management Skillbuilding Hub at <https://dataoneorg.github.io/Education/bestpractices/ensure-integrity-and> on Aug 31, 2020