



What is it?

Preserving is more than a backup. It is a set of actions, which ensure long-term retention of the integrity of data, by maintaining its (a) **accessibility**, (b) **authenticity** and (c) **longevity**. Digital objects are fragile, being susceptible to “**data rot**” which might influence their accessibility and authenticity. Preservation aims to ensure that datasets are in the best shape to be stored, discovered, accessed and reused. It implies recurring activities.

(a) **Accessibility**: Data can be retrieved, displayed and used.

(b) **Authenticity**: Data have not been manipulated, substituted or faked.

(c) **Longevity**: Data are reusable for long-term, independently of software and hardware decay.

There are different digital preservation methods, such as: migration, emulation, digital archaeology, and technology preservation.

How to do it

1. Be aware of the requirements of your data center/archive to preserve and store data.
2. If the data set is not already in an open access file format, then it should be migrated (transformed). The original data bit stream should be kept before **migration** to other formats is done.
3. **Preservation metadata** should be added and all migration procedures documented. A detailed description of the preservation actions should be included.
4. Quality procedures, such as checking the original and the migrated bit stream or checking the sums should be performed to assure the authenticity of the data and to prevent information loss during the migration process or after a certain period of time.
5. **Backup** is just one component of preservation and both steps should not be mixed up.
6. **Refresh** should be done every certain of time to prevent data rot.

Who does it?

Preservation is usually carried out by **archives, data centers or data managers** after data have been submitted by the data producer. However, **data producer** can facilitate the process performing good data management practices in the previous data life cycle steps. GFBio’s archives do data preservation and can be accessed via the [GFBio-Portal](http://www.gfbio.org). The portal acts as a bridge between researchers and infrastructures, providing support in data management to facilitate preservation actions later on.

Key elements

- Migrate data to the best format (open) and suitable medium.
- Document preservation actions (workflow).
- Create preservation metadata.
- Backup and store data.

Useful links

<http://www.dcc.ac.uk/resources/how-guides/cite-datasets> (DCC How-To-Guide)

<http://databib.org> (useful data repository registry)

<http://www.re3data.org/> (useful data repository registry)

http://en.wikipedia.org/wiki/Data_degradation (Information about data degradation)