Data Life Cycle

Description

This page will help users to understand how to manage their data on the cluster. We provide a quick example here, but for more details, please consult the Data Management Plan (DMP) provided by Unige.

Each user is responsible for their data and must manage it from generation until deletion from the cluster.

From Terms of use → Storage:

```
The HPC clusters are not a long-term storage provider: users are requested
to manage their files
on a regular basis by deleting unneeded files and migrating results or
valuable data to a permanent location such as Tape NASAC or Yareta.
```

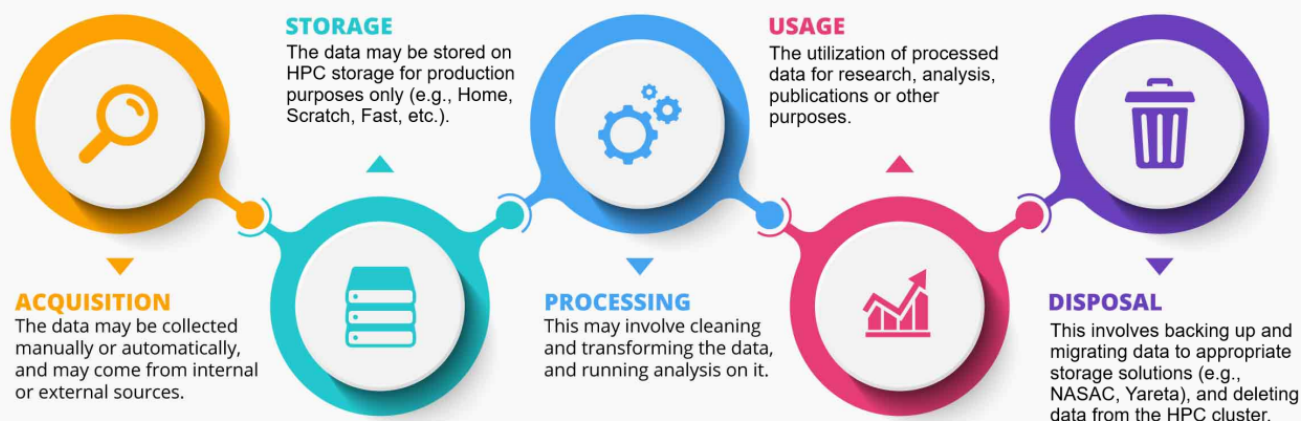This ensures enough space for everyone and guarantees optimal performance for computing.

Data Management

Below is a schema representing an example data life cycle, which includes the following stages:

- **Acquisition:** The process of collecting or generating data.
- **Storage:** The data may be stored on HPC storage for production purposes only (e.g., Home, Scratch, Fast, etc.).
- **Processing:** The manipulation or analysis of data to extract useful information.
- **Usage:** The utilization of processed data for research, analysis, or other purposes.
- **Disposal:** This involves backing up and migrating data to appropriate storage solutions (e.g., NASAC, Yareta, Hereda), and deleting data from the HPC cluster.

This example should be adapted to your needs; however, it must comply with the terms of use. Any unused or unnecessary data for computation must be removed from the cluster. Additionally, old data should be removed if it will not be used in the near future. Keeping a small amount of old data is tolerable, but several hundred gigabytes or terabytes can become problematic. If everyone stores too much data, there will be no space left for new projects, impacting the overall performance and availability of the HPC cluster. (cf. hpc-community: baobab-urgent-scratch-partition-nearly-full)

DATA LIFECYCLE MANAGEMENT

STORAGE
The data may be stored on HPC storage for production purposes only (e.g., Home, Scratch, Fast, etc.).

USAGE
The utilization of processed data for research, analysis, publications or other purposes.

ACQUISITION
The data may be collected manually or automatically, and may come from internal or external sources.

PROCESSING
This may involve cleaning and transforming the data, and running analysis on it.

DISPOSAL
This involves backing up and migrating data to appropriate storage solutions (e.g., NASAC, Yareta), and deleting data from the HPC cluster.

From:

https://doc.eresearch.unige.ch/ - **eResearch Doc**

Permanent link:

https://doc.eresearch.unige.ch/hpc/data_life_cycle

Last update: **2025/06/11 12:27**