Data Lifecycle Management

- Home directories (NFS storage)
  - This file system is accessible from systems at the facility including compute nodes and data transfer nodes (DTNs).
  - Home directories are designed for source code and documents and are optimized for small to medium size files. Home directories are not designed for high bandwidth or high IOPS access to data.
  - In addition to being backed up on a daily basis, daily snapshots of directories are also made. Directory backups are kept for 90 days, while snapshots are kept for 7 days.
  - Quota per user is small (~5GB/user) , varies for experiments.

- Project file systems (GPFS and Lustre)
  - Project file systems at the SDCC are targeted primarily at collaborations for high bandwidth, large block access to large files. These file systems are accessible from systems at the facility including compute nodes and data transfer nodes.
  - Neither snapshots nor backups are made of project file systems.
  - Some project file systems, particularly but not limited to those designated as “scratch”, may have policy driven auto-deletion of files.

- Software distribution file systems (OpenAFS and CVMFS)
  - Software distribution file systems are read only, world accessible and readable file systems. These file systems are designed for worldwide, read only distribution of applications and libraries.

- Object storage (BNLBox)
  - Default quota is 50GB/user.
  - Write access to BNL Box is limited to authorized users, file read access controlled by owner of the data.
  - BNL Box has built-in file versioning and trash bin (30 day retention policy, up to quota limit)
  - Data in BNL Box is backed up daily and retained for 30 days.
  - Data in BNL Box Archive folder does not count against user quota and is transparently migrated to HPSS by policy

- Cold storage services. (HPSS - Tape)
  - This service is tailored for large files (files > 10GB in size).
  - Access to cold storage must be negotiated with the SDCC as the characteristics of the service provided can vary dramatically depending on access and quality of service requirements.