

# WLCG Workshop S.F. 2016 – Data Management Session

<https://indico.cern.ch/event/555063/sessions/207257/>

Sat 8 Oct, 09:45.

Session concentrating on reducing the cost of storage.

List of possible topics for contributors. Topics relevant to infrastructures, experiments and storage providers have been combined - choose the ones relevant to you, ignore the rest, and add any that you see are missing.

- Reduce cost/volume
  - cost of storage management
    - integrating standard (non HEP) solutions e.g. ceph
    - protocol zoo, SRM-less operation
    - T2 storage as cache
    - multi-site storage
    - regional federations
    - cloud storage
  - storage overheads; replication, erasure, RAID levels etc
    - reduce reliability requirements?
    - reduce cost/impact of data loss
- Reduce volume used
  - reduced number of global replicas
    - remote access
    - latency hiding
      - applications, overcommitting
    - global federations
    - CPU-only resources (inc cloud)
  - data formats and lifecycle, intermediate products
  - resource reporting
    - monitoring usage
    - eliminating dark data
  - data "enrichment"
    - popularity
    - caching, avoiding unused data
      - promoting locality in workflows
  - trading disk for...
    - tape
      - data parking
    - CPU
      - maintain metadata enabling regeneration of data on demand