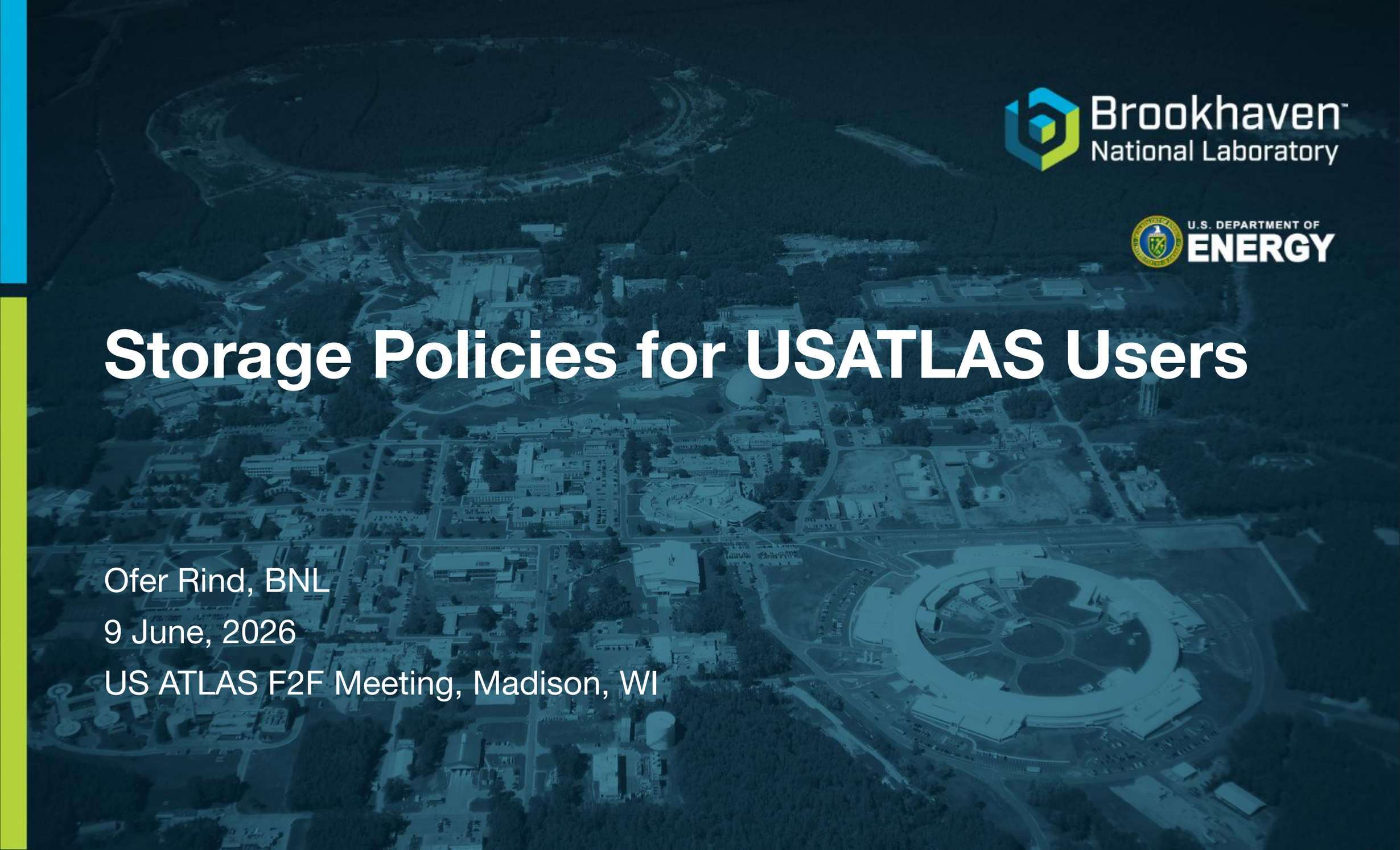


Storage Policies for USATLAS Users

Ofer Rind, BNL

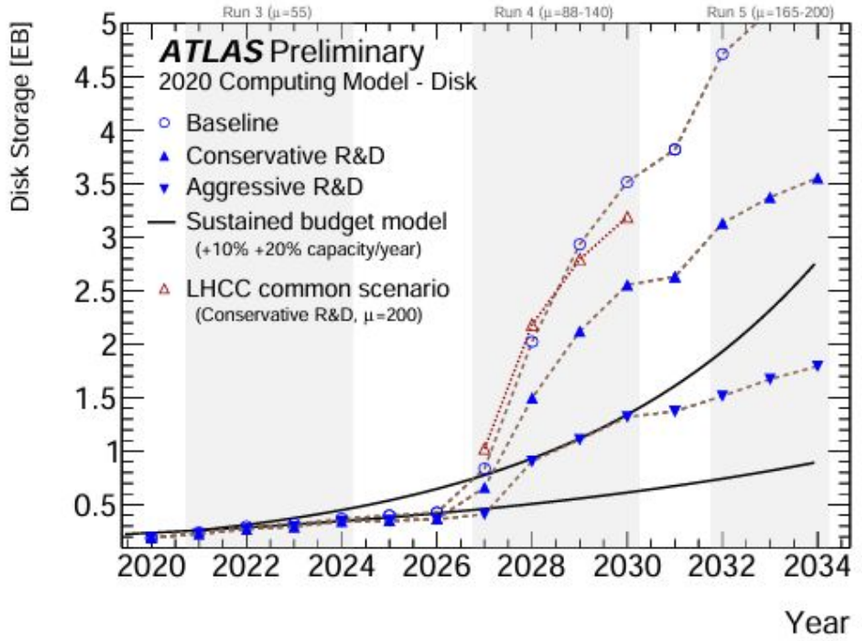
9 June, 2026

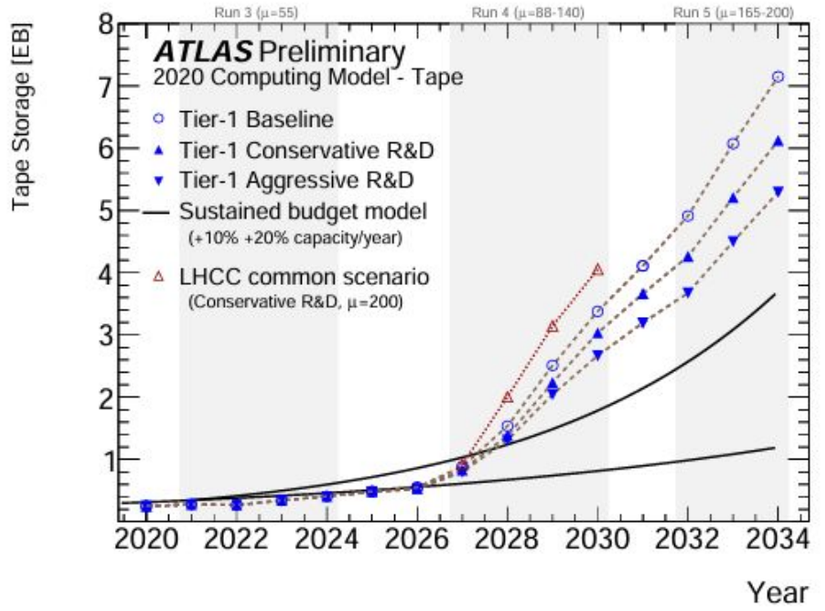
US ATLAS F2F Meeting, Madison, WI

Storage Needs for USATLAS Users

- Rucio-managed pledged storage
  - Effective access for running large scale analysis on the grid (e.g. derivations)
  - Can be accessed from local batch systems (BNL, UC)
- PHYSLITE
  - Reduced size physics analysis format
  - Significant and growing use
- Ntuples
  - Final stage physics analysis
  - Individuals and groups
  - Stored on non-pledged resources and shared via EOS, local shared storage, etc.
- Work files
  - Code, plots, publications
  - Stored on local home dirs or cloud tools (Git, overleaf, ggdocs,...)

Source: ATLAS CDR





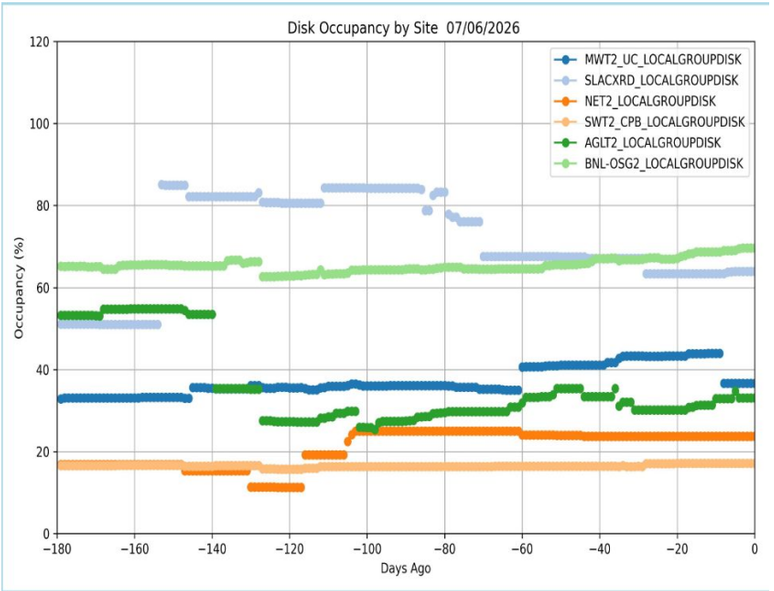
Brookhaven National Laboratory

The Current Landscape for User Storage in US

- Rucio-managed distributed storage, aka localgroupdisk, scratchdisk
  - Storage provided at six US sites (dCache, xrootd)
  - Centralized policy-based quota management (RAC, Mayuko)
  - Rucio subscription model - access via grid transfer tools or local mounts (xrdcp, root)
  - Generally Non-posix
- Local (“Tier-3”, Analysis Facility) user storage - BNL, UC, SLAC
  - Home directories, plus locally-managed data storage
  - Site-specific resources and policies
  - Generally posix (preferred by users)
- CERN EOS
  - Shared spaces directly accessible on lxplus
  - Accessible via grid tools (XRootD, WebDAV)
  - Kerberos-authenticated fuse mounts also available at sites (BNL, MWT2)
    - Convenient posix-like access, but low performance
- Other storage
  - XCache at sites and on ESNET - utilized by VP queues. Interactive use?
  - Globus at HPC sites
  - ServiceX

Localgroupdisk

- 7+ PB distributed across sites, but 80% at BNL and MWT2
  - Default quota is 15 TB/site/user (50 TB at BNL), “enforced” via email warnings
- Occupancy is low and stable
  - Why is this?



T1 T2 LOCALGROUPDISK Usage

LOCALGROUPDISK	Usage Size (TB) (Complete)	Usage Size (TB) (Incomplete)	Usage Size (TB) (Complete+Incomplete)	Disk Usage Size (TB)(SRM report)	Disk Allocation Size (TB)
SWT2_CPB	60.3	76.1	136.4	137.1	800.0
SLACXRD	67.5	5.1	72.6	76.7	120.0
NET2	43.5	1.5	45.0	45.0	189.6
MWT2_UC	602.6	127.3	729.9	732.8	2000.0
BNL-OSG2	2534.6	198.7	2733.3	2702.2	3881.7
AGLT2	104.2	9.2	113.4	115.7	350.0

Brookhaven National Laboratory

Analysis Facility Storage

- /home, /data, /work, /cold, etc….
- An array of solutions: NFS Netapp, GPFS, Ceph, dCache, …
- Access via local posix mount, local batch, jupyter notebook, grid, …
- #1 limitation reported in User Survey: **Availability of Disk Storage**
  - **Need both *more storage* and *easier accessibility* (including cross-site)**
  - ***"Users want more /data space and resist using LOCALGROUPDISK"***
- Throughput requirements? – 400 Gbps Challenge and NVME

Default Storage Quotas by Facility

Storage Tier	BNL	SLAC	UChicago
HOME	20 GB	30+200 GB	100 GB
DATA	500 GB	2 TB	5 TB
SCRATCH	No quota	No quota	No quota
LOCALGROUP DISK	15 TB	15 TB	15 TB

Brookhaven National Laboratory

Towards a Unified Storage Policy

- Draft Document started in conjunction with WBS 5 (link)
- Standardize storage allocations and performance expectations across sites
  - Default storage size and policy for extension requests
  - Localgroupdisk still expected to play a role for large, shared datasets
- Aim for as uniform a user experience as possible across sites
  - Standardize on directory structures
- Communication policies and methods
  - User notifications such as exceeding quota, expiration of storage extension requests, cleanup of inactive accounts and unused data, etc.
- Seamless cross-site data accessibility will be needed
  - Users run their analyses where they can access their data
  - Currently, no effective mechanism for providing this
  - Not limited to US

Questions

- What can we do now, in the short term?
  - Finalize user storage policy?
  - Determine a minimal standard for storage system performance requirements?
  - Begin to standardize user environment across sites?
- How to handle PHYSLITE?
  - Want a full copy in the US, including X versions.
  - Can regional network caches help with low latency, high throughput accessibility?
- What are realistic analysis throughput expectations and how do we meet them?
- Can we begin working on better cross site user data sharing mechanisms?
  - /data vs. /home - what do users really need? Does Ilija's LLM decide?
    - EOS mounts helpful, but not a great solution so far
  - Easier access to analysis teams shared storage?
  - Better integration with CERNBox for home directories?
- What will be the impact of new analysis workflows, e.g. AI/ML?
  - Payload offloading to clouds, HPC? Inference servers?
  - How does ServiceX fit in to the plan?