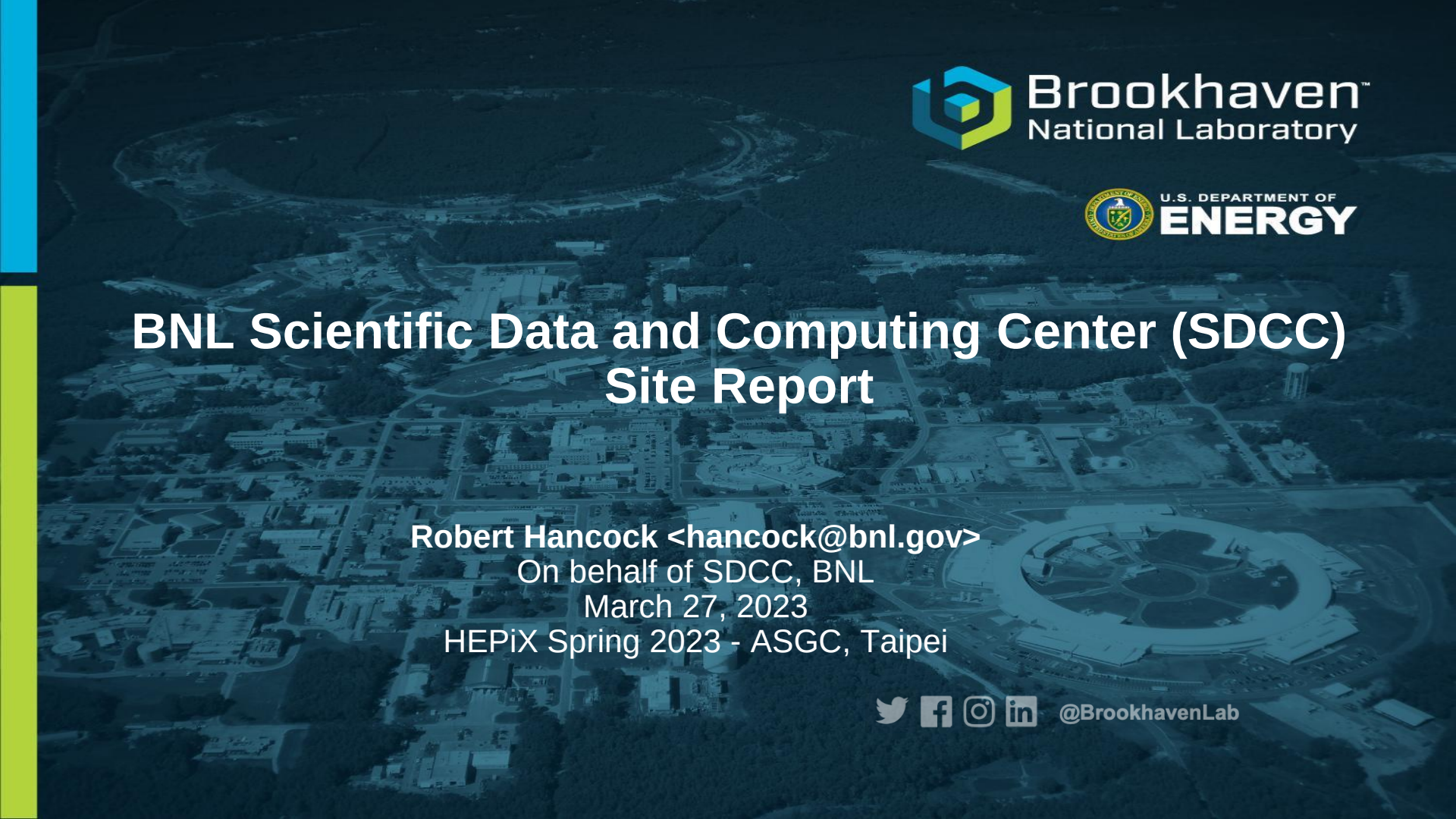


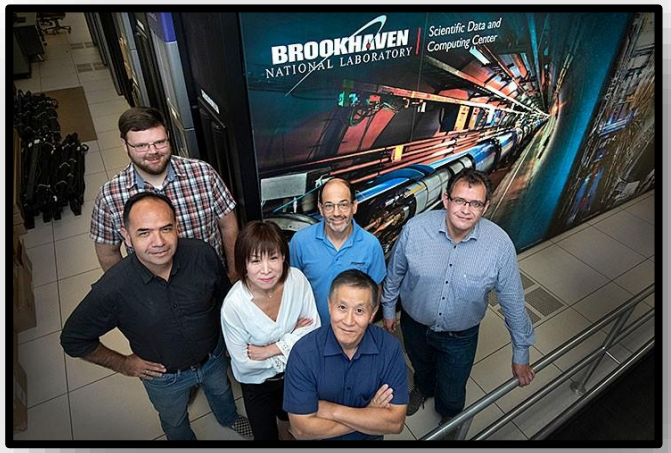
BNL Scientific Data and Computing Center (SDCC) Site Report

Robert Hancock <hancock@bnl.gov>
On behalf of SDCC, BNL
March 27, 2023
HEPiX Spring 2023 - ASGC, Taipei

@BrookhavenLab

SDCC: The Scientific Data and Computing Center

- Located at Brookhaven National Laboratory (BNL) on Long Island, New York
- SDCC was initially formed at BNL in the mid-1990s as the RHIC Computing Facility



Shared multi-program facility serving ~2,000 users from more than 20 projects

Scientific Data and Computing Center Overview

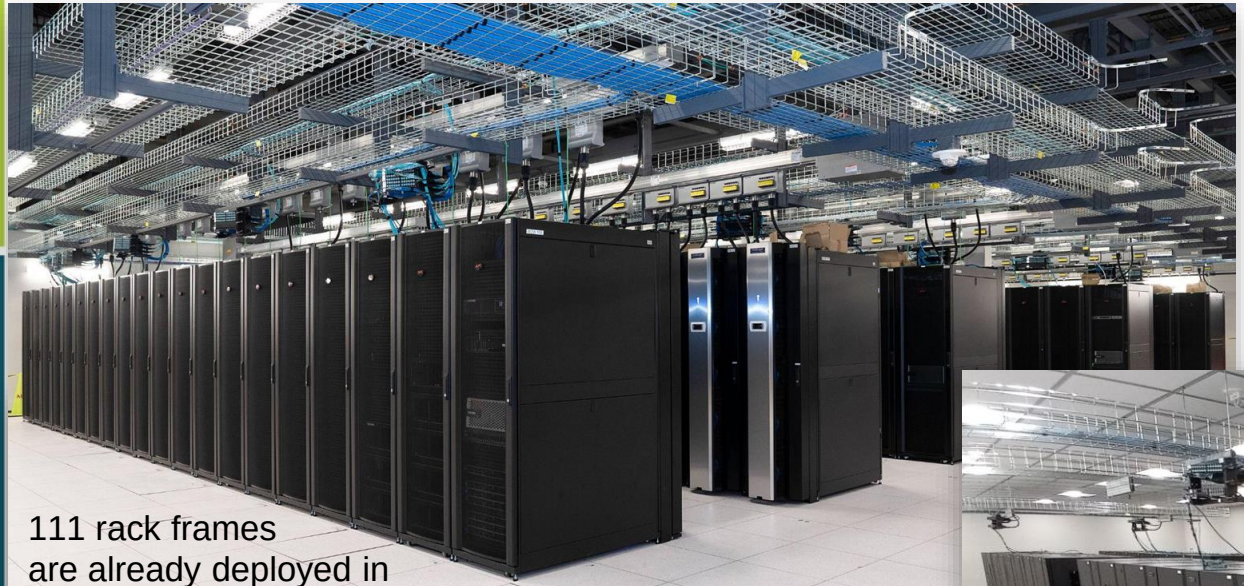
- Tier-0 computing center for the RHIC experiments
- US Tier-1 Computing facility for the ATLAS experiment at the LHC，also one of the ATLAS shared analysis (Tier-3) facilities in the US
- Computing facility for NSLS-II
- US Data center for Belle II experiment
- Providing computing and storage for proto-DUNE/DUNE along w/ FNAL serving data to all DUNE OSG sites
- Also providing computing resources for various smaller / R&D experiments in NP and HEP
- Serving more than 2,000 users from > 20 projects
- Developing and providing administrative/collaborative tools:
  - Invenio, Jupyter, BNL Box, Discourse, Gitea, Mattermost, etc.
- BNL was selected as the site for the upcoming major new facility Electron-Ion Collider (EIC/eRHIC)
- sPHENIX - scheduled to start taking data in May

Brookhaven National Laboratory

BNL Core Facility Revitalization (CFR) Project: New Data Center

New Data Center (Building 725) — 2023Q1: 1.5 Years of Production Operations

- CFR project finished the design phase in the first half of 2019 and completed the construction phase by the end of FY21
- The occupancy of the B725 data center for production CPU and DISK resources for all programs started in 2021Q4 and ramped up in 2022Q1-2023Q1 to the level of 62 racks populated with equipment in the B725 Main Data Hall (MDH)
  - 10 more storage / infrastructure racks are in the process of being configured as of 2023Q1
  - 20 more new HTC CPU racks and 1 more HPC CPU rack are expected to be added to B725 in 2023Q2
- Currently we have two diesel generators installed in B725 diesel generator yard providing covering up to 1.2 MW of total IT payload with N+1 redundancy
  - Two more diesel generators are planned to be added in FY24-25 to provide all IT payload in the B725 data center as it scales beyond 1.2 MW and 2.4 MW thresholds for combined IT payload deployed
- Two library rows in B725 Tape Room are populated with IBM TS4500 tape libraries to serve ATLAS and sPHENIX experiments (4 libraries, 128 tape drives in total).
  - One more library row is expected to be populated in FY24 (2 more IBM TS4500 sPHENIX libraries)
- The completion of the transition of the majority of CPU and DISK resources deployed in SDCC environment to the new B725 datacenter is still expected to be achieved by the end of FY23
  - The vast majority of equipment purchased by SDCC starting from 2021Q3 is being placed in the new data center in preparation for the retirement of the oldest areas of B515 datacenter by Sep 30, 2023



111 rack frames
are already deployed in
B725 Main Data Hall MDH

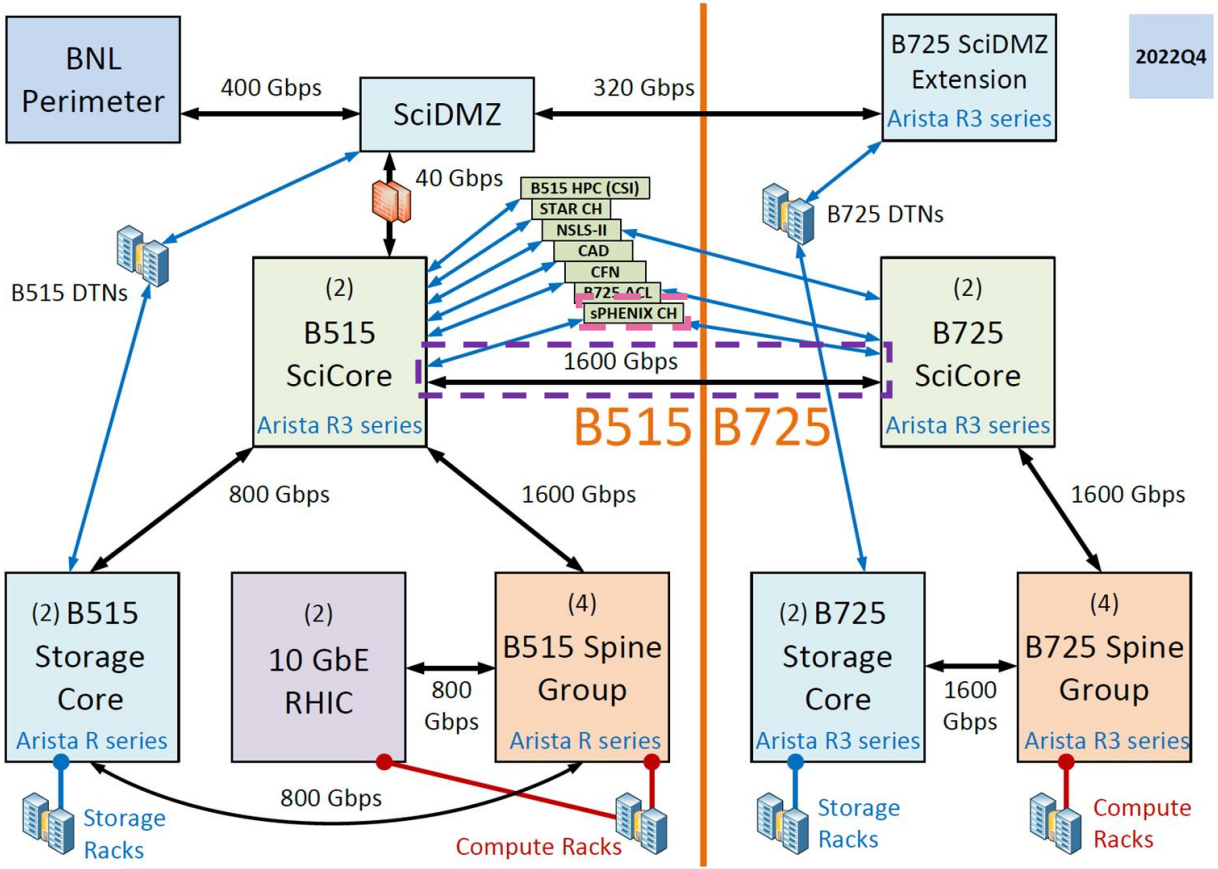
84 RDHx units deployed in B725 MDH, out of which 59 are on racks with equipment while 25 are deployed for the future growth

B725 Central Network Equipment Is Deployed & Active
(10x 400 GbE ready Arista modular chassis with 48x line cards slots in total)

4x 8-frame IBM TS4500
tape libraries are installed
in the B725 Tape Room

5

Network Systems of B515 and B725 Data Centers (since Oct'22)



- 1.6 Tbps (LR) B515/B725 interbuilding link is functioning as expected
- sPHENIX Counting House (CH) uplinks to both B515 & B725 data centres are activated at 200 Gbps to B515 plus 400 Gbps to B725

High Throughput Computing

- Providing our users with ~1,900 HTC nodes:
  - ~90,000 logical cores
  - ~1050 kHS06
  - Managed by HTCondor
- Purchased 648 Supermicro SYS-610C-TR nodes for ATLAS and the RHIC experiments (~62k logical cores total)
  - Expected delivery April 2023
  - Housed in 20 racks
  - System specs:
    - Dual Intel Ice Lake Xeon Gold 6336Y 24-core processors
    - 12x32 GB 3200 MHz ECC DDR4 RAM (384 GB total)
    - 4x2 TB SSD drives
    - 1U form factor
    - 10 Gbps NIC
- Will be purchasing some Supermicro ARM test nodes in April
  - With Ampere Altra CPUs
- All nodes running still running Scientific Linux (SL) 7
  - Preparations for an OS upgrade to Alma Linux in progress
- HTCondor 10.0 fully tested, and a rolling upgrade has begun

Supermicro SYS-6019U-TR4 Servers

Brookhaven National Laboratory

High Performance Computing

Currently supporting 5 HPC clusters

- **Institutional Cluster gen1 (IC)**
  - 216 HP XL190r Gen9 nodes with EDR IB
  - 108 nodes with 2x Nvidia K80
  - 108 nodes with 2x Nvidia P100
- **Skylake Cluster**
  - 64 Dell PowerEdge R640 nodes with EDR IB
- **KNL Cluster**
  - 142 KOI S7200AP nodes with dual rail Omnipath Gen.1 interconnect
- **ML Cluster**
  - 5 HP XL270d Gen10 nodes with EDR IB
  - Each node has 8x Nvidia V100
- **NSLS2 Cluster**
  - 32 Supermicro nodes with EDR IB
  - 13 nodes with 2x Nvidia V100

New IC Gen2 Cluster

Institutional Cluster gen2 (IC)

IC gen2 was just delivered last week. Specs:

- 2x Intel Xeon (Ice Lake)
- 512GB DDR4-3200 on CPU nodes
- 1TB DDR4-3200 on GPU nodes
- NDR200 InfiniBand interconnect (200Gbps per uplink)
- 4x Nvidia A100 80GB on GPU nodes

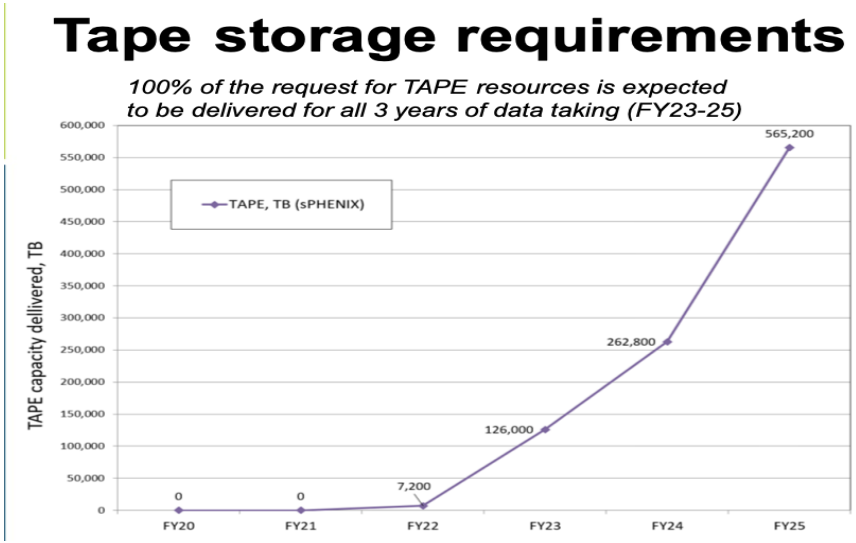
We are looking at a performance of 3x from current IC node: ~7.9 TF to ~25TF IC gen2 node.

Brookhaven National Laboratory

Tape System

- Currently ~220 PB of data in HPSS with ~75k tapes
- New HPSS hardware in the newly commissioned data center for sPhenix experiment
  - 10GB/sec data injection requirement
  - High performance/capacity disk cache (2.1PB, 330 HDDs)
  - Two new IBM TS4500 tape libraries
  - Total of 64 new LTO-9 tape drives in two TS4500 libraries

See Tim Chou's sPHENIX Archival Storage talk later in the week



sPHENIX Tape Storage Space Requirements

Brookhaven National Laboratory

Storage: Lustre, dCache & XROOTD

dCache.org

Total ~74 PB in dCache

- ATLAS (v8.2.15), Belle II (v7.2.19), PHENIX (v5.2.9), DUNE (v8.2.2)

XROOTD

~11 PB total storage for STAR

- Mix of central and farm node storage

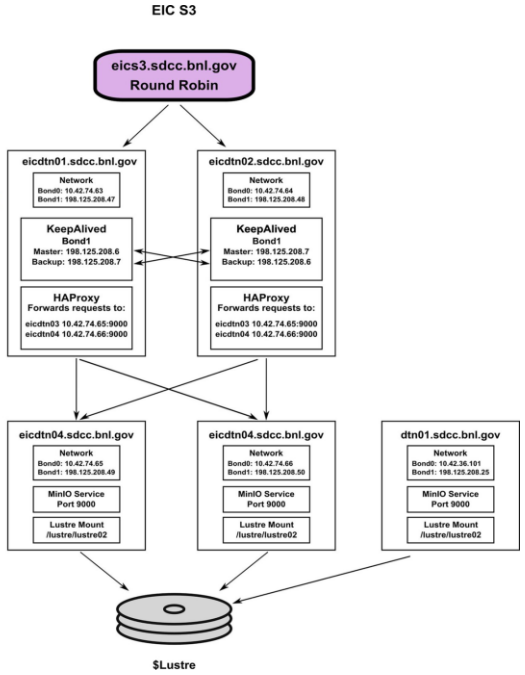
lustre®

Total ~50 PB in Lustre

- ATLAS, EIC, LQCD, NSLS-II, sPHENIX, STAR
- Growing footprint for Lustre (2.12.8)
- Added 25PB to sPHENIX
- Excellent streaming sequential performance with aggregate throughput of 210 GB/s

Support for home-hosted S3 storage

- Used by the Electron Ion Collider - Allows for seamless access from anywhere, anytime - well suited for the EIC model (unique dual-site model)
- Highly Available configuration using HAproxy and two servers
- Initial deployment used S3 over Lustre (deprecated) - 3 PBytes
- New storage (now in-house) will use native Object Storage (CEPH) & Federated ID access



Brookhaven National Laboratory

Redhat Virtualization

- Redhat Virtualization is end of life in 2024.
- Evaluating Openshift Virtualization and VMware.
  - Leaning towards VMware due to product maturity, features and pricing.
- Moving 600 VMs from RedHat Virtualization to VMware or Openshift will take time.
- Since RHEL7 is also EOL in 2024, we can rebuild RHEL7 VMs as RHEL8 VMs in the new virtualization platform.

SDCC Web Presence

SDCC Hosted Drupal Sites

- SDCC - https://www.sdcc.bnl.gov/
- sPHENIX - https://www.sphenix.bnl.gov/
- US ATLAS - https://www.usatlas.org/
- Quantum Astrometry - https://www.quantastro.bnl.gov/
- Cosmology & Astrophysics - https://www.cosmo.bnl.gov/
- US Belle II - https://www.usbelleii.bnl.gov/ (newest addition)

The newly added US Belle II site is the first of our sites to utilize our new COmanage service for Federated Logins

All sites are set for a major version upgrade in Nov 2023 including an OS upgrade

Brookhaven National Laboratory

SDCC Web Presence

SDCC MediaWiki Service

- sPHENIX MediaWiki is now hosted by the SDCC: https://wiki.sphenix.bnl.gov/
  - Wiki is set for private and public content
  - Upon logging in users are authenticated against SDCC Keycloak
  - Users currently in the sPHENIX group will automatically receive higher privileges to access private group only information
  - When a user is removed from the experiment the elevated role is removed as well this check is made on every login to accurately reflect user status
- Planned upgrade for service and OS for late 2023
- Automated deployment provides future experiments a quick setup
- Will be added to use SDCC COmanage service at a future date
  - COmanage service will provide for a wider range of institutions to participate with the added bonus for a more finely managed roles based privileges

Thanks to the following people at BNL for contributing to this presentation:

Costin Caramarcu, Tim Chou, Joe Frith, Vincent Garonne, Chris Hollowell, Jerome Lauret, Louis Pelosi, Alex Zaytsev, [...]

Questions?