Optimising operation and use of JASMIN -
The UK's data analysis facility for environmental science

Fatima Chami and James Hannah

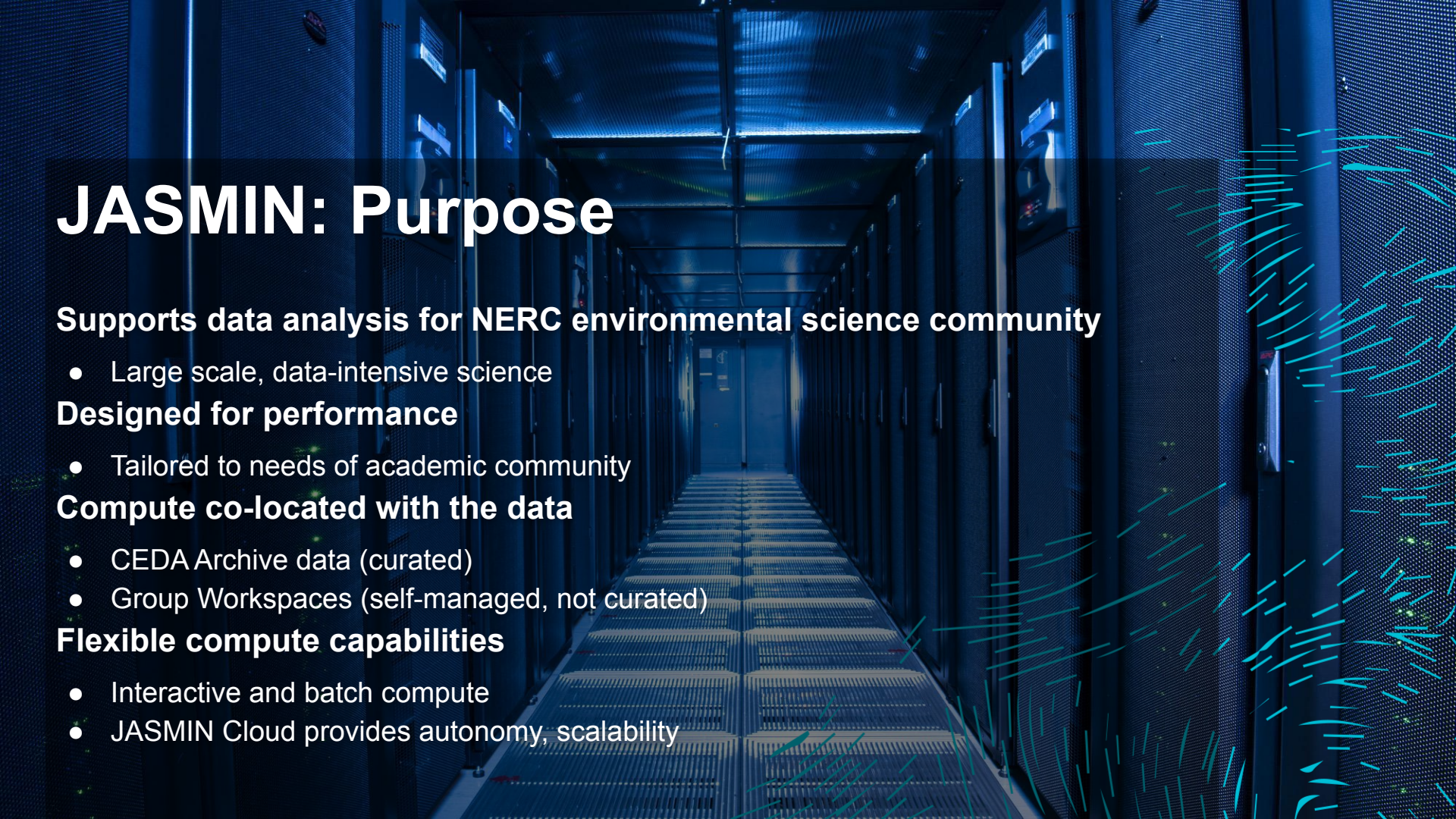
IRIS Collaboration meeting, 1st July 2025, Durham

UKRI Science and Technology Facilities Council
Natural Environment Research Council

UK Research and Innovation

Outline

- Overview of JASMIN
- Bringing users to the data & compute
- Efficient collaborative data processing space
- Data management services
- Automated data migration -NLDS
- Batch processing LOTUS- new cluster/configuration
- JASMIN future plans



JASMIN: Purpose

Supports data analysis for NERC environmental science community

- Large scale, data-intensive science

Designed for performance

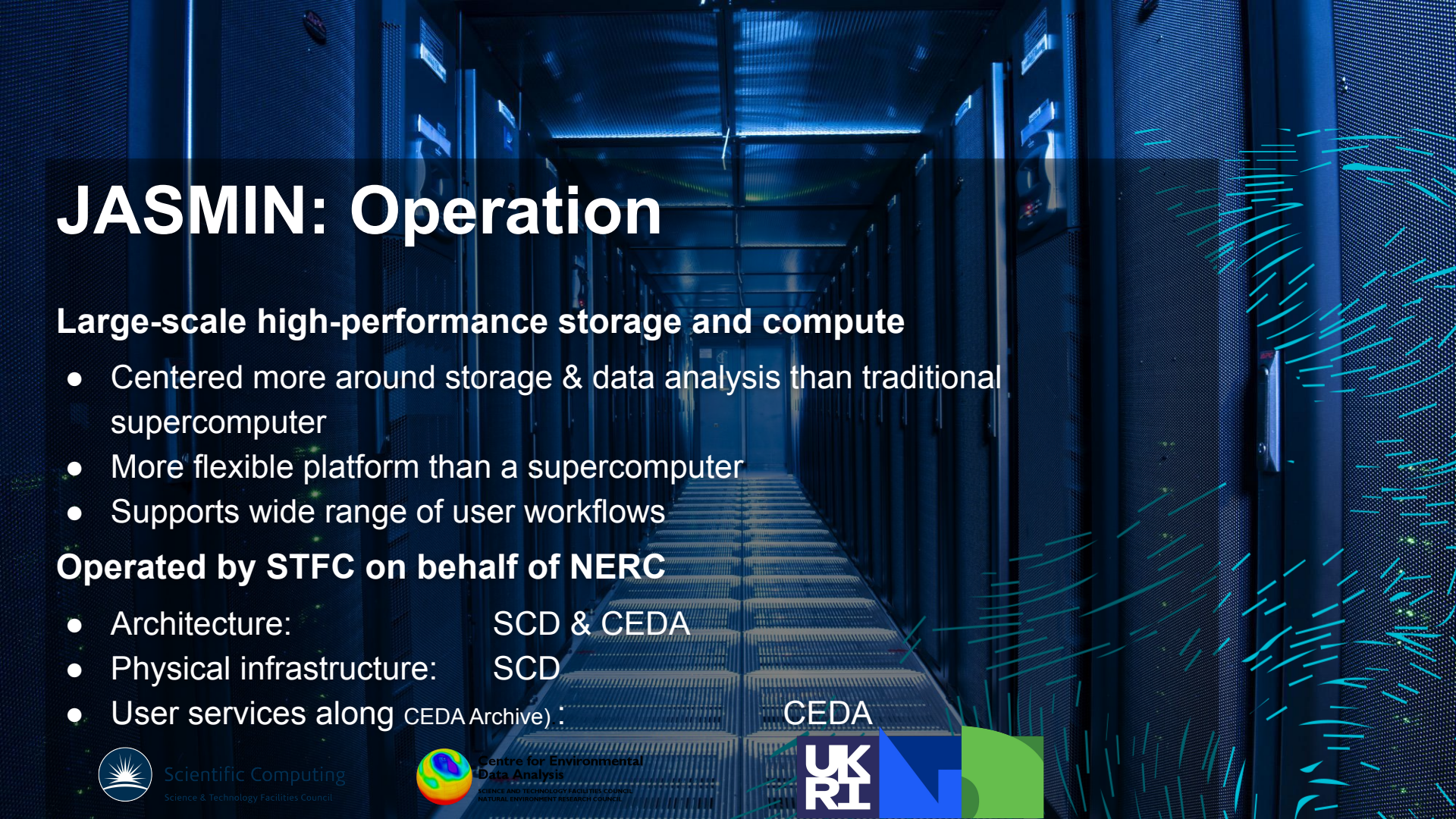
- Tailored to needs of academic community

Compute co-located with the data

- CEDA Archive data (curated)
- Group Workspaces (self-managed, not curated)

Flexible compute capabilities

- Interactive and batch compute
- JASMIN Cloud provides autonomy, scalability



JASMIN: Operation

Large-scale high-performance storage and compute

- Centered more around storage & data analysis than traditional supercomputer
- More flexible platform than a supercomputer
- Supports wide range of user workflows

Operated by STFC on behalf of NERC

- Architecture: SCD & CEDA
- Physical infrastructure: SCD
- User services along CEDA Archive): CEDA

Scientific Computing
Science & Technology Facilities Council

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

UKRI

Science and Technology Facilities Council

Natural Environment Research Council

Part 1

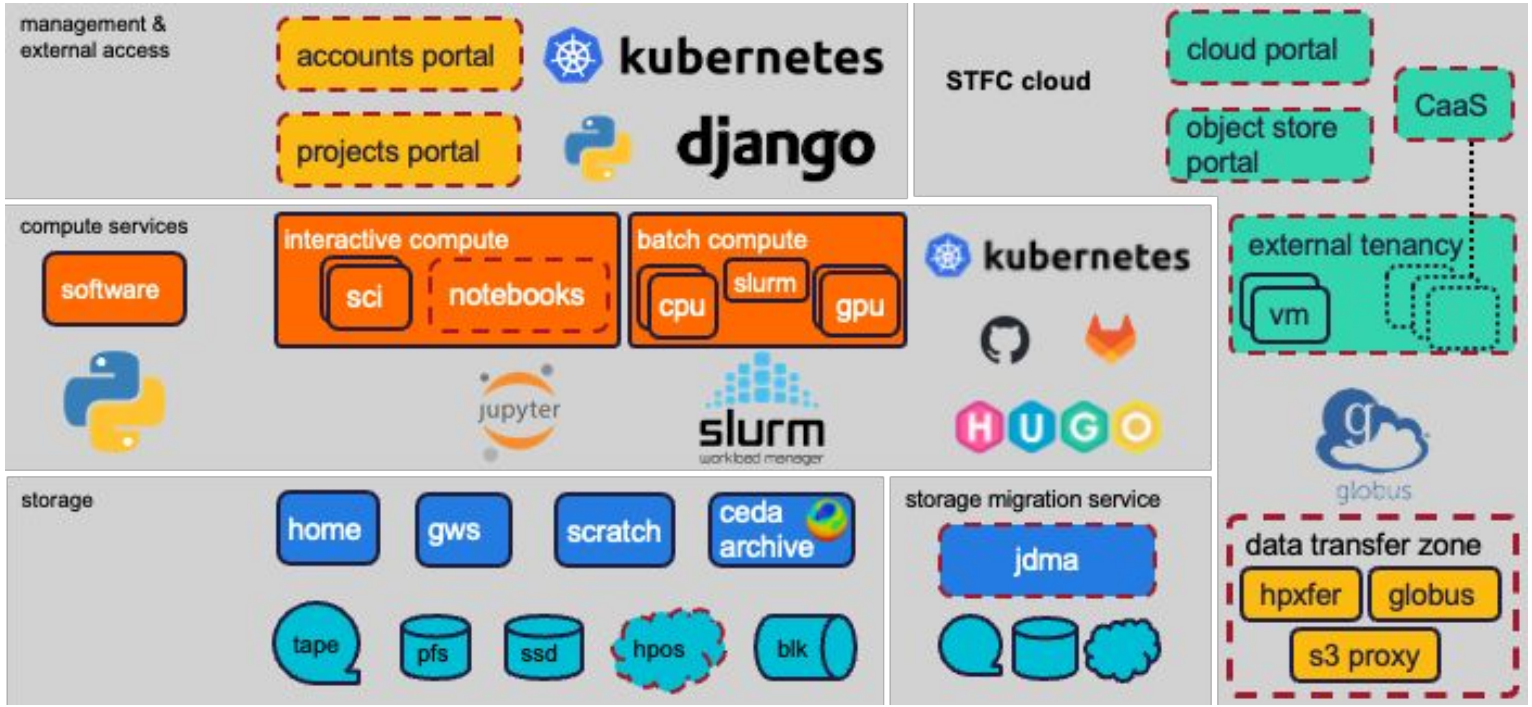
- Bringing users to the data & compute
- Efficient collaborative data processing space
- Data management services

Bringing users to data & compute - Why?

- Increasing the efficiency of all of NERC's environmental science compute
- Reducing the duplication of compute resources, data movements and storage
- Bringing all users to one place
- Enabling collaborative research

UKRI Science and Technology Facilities Council Natural Environment Research Council

Bringing users to data & compute - How?



UKRI Science and Technology Facilities Council Natural Environment Research Council

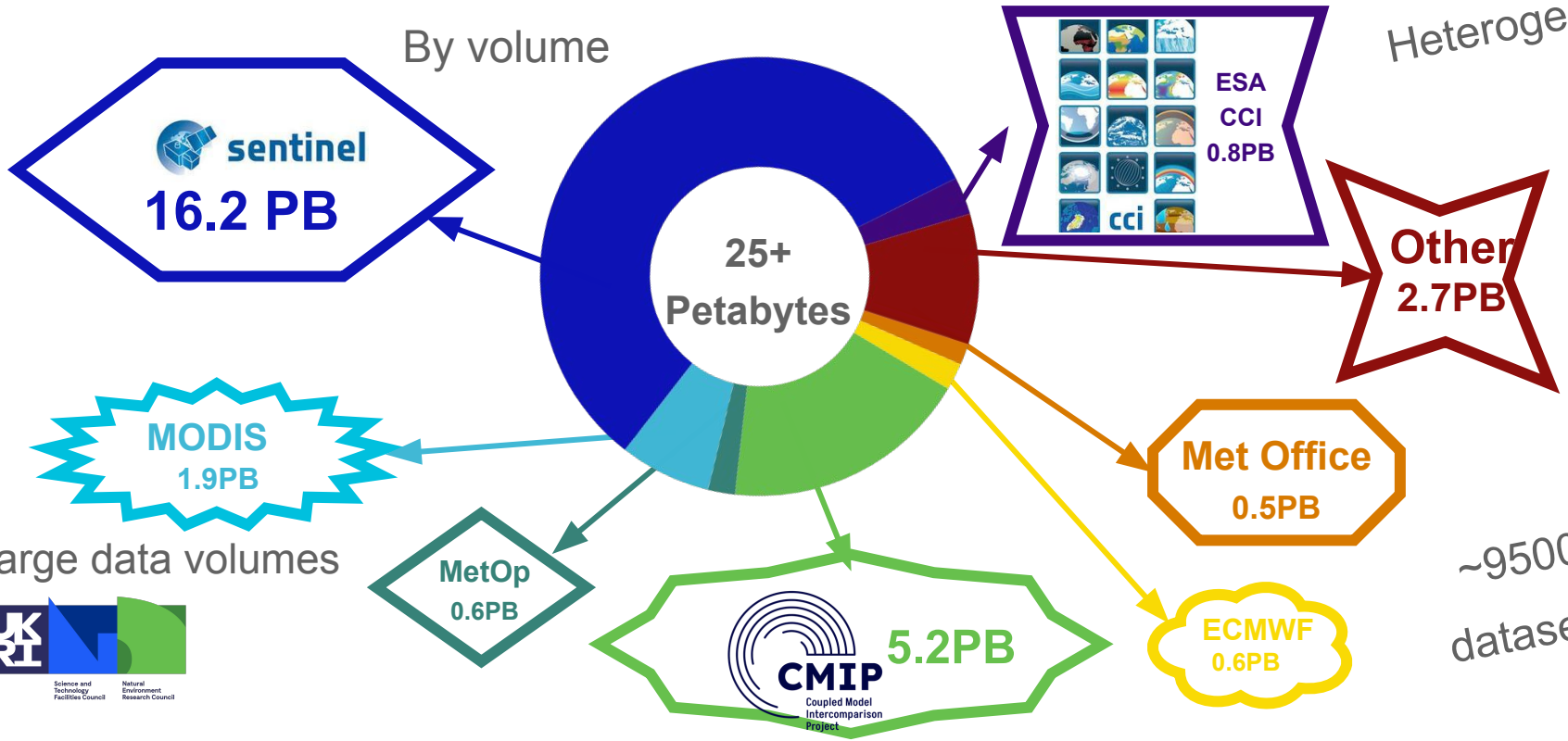
CEDA Archive: by volume

CEDA Archive

NERC Environmental Data Service

Long term archive for Atmospheric and Earth Observation data

By volume



Heterogenous

Large data volumes

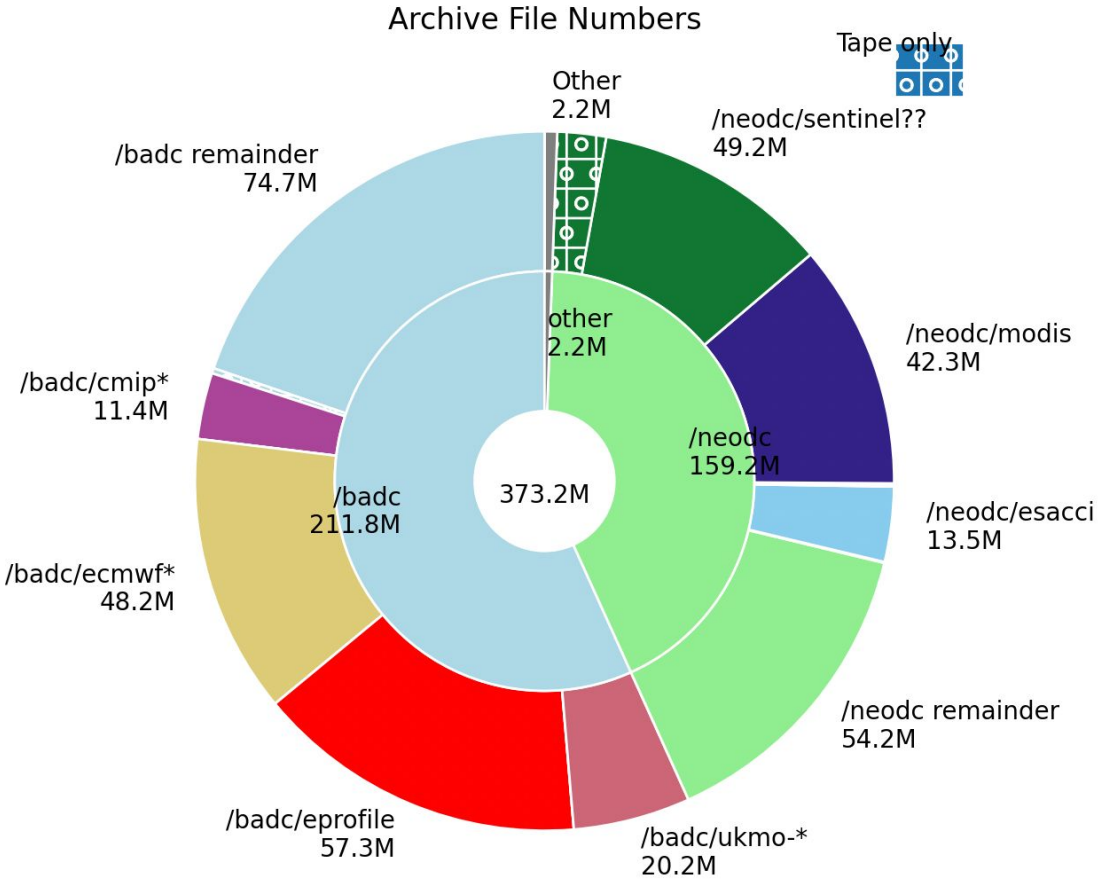
~9500 datasets

UKRI Science and Technology Facilities Council / Natural Environment Research Council

Big *Numbers* of files as well as big volumes

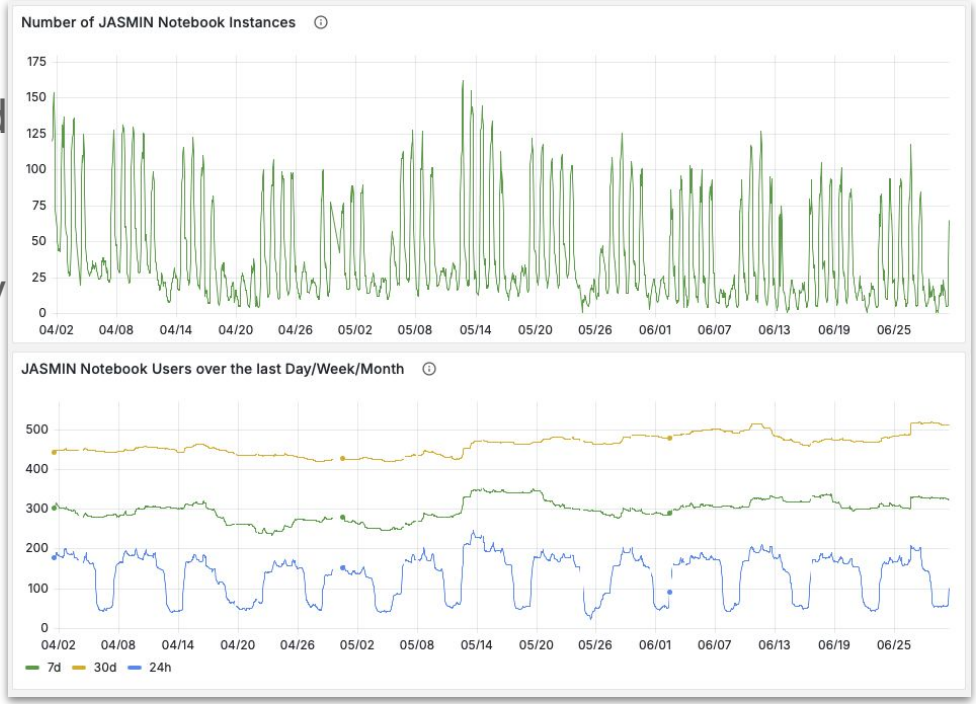
It's not only big volumes that we need to deal with:

- Archive is 350+ million individual files
- eprofile is just a few TB but its 60M files!



Bringing users to data & compute

- Easy access to JASMIN
- No data download, data is collocated
- CEDA archive data (~30PB)
- The Met office tape store (300PB)
- Different modes of working efficiently
  - -Interactive sci platform
  - -Graphic -remote desktop
  - -Jupyter notebook service
  - -Batch cluster



UKRI Science and Technology Facilities Council Natural Environment Research Council

JASMIN users - Multiple account types

- Standard
- Shared
- Service
- Training
- JASMIN accounts portal
  -easy to manage access

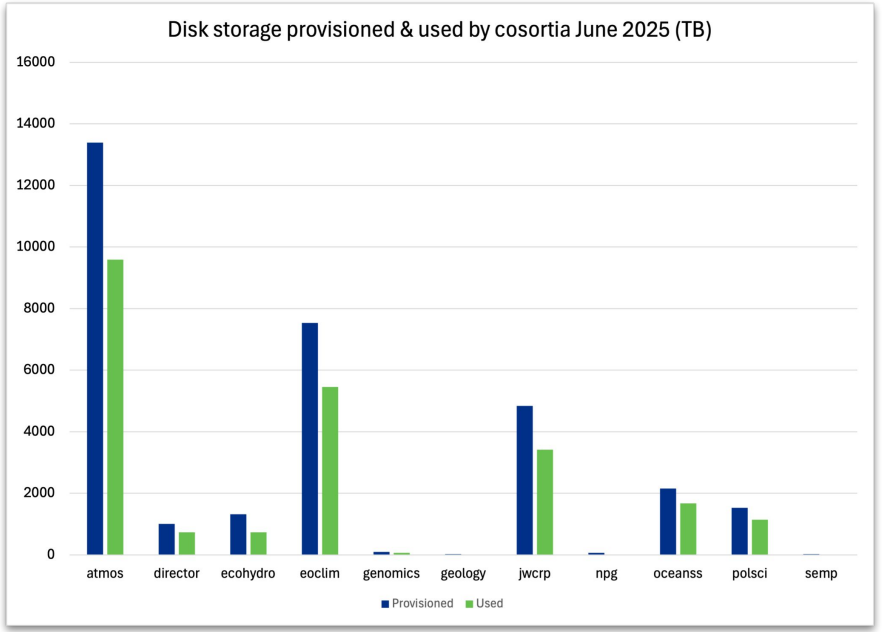
Fig 1: Number of active JASMIN login grant holders over the last 3 months

UKRI Science and Technology Facilities Council Natural Environment Research Council

Efficient collaborative data processing space

GWS - Group workspaces

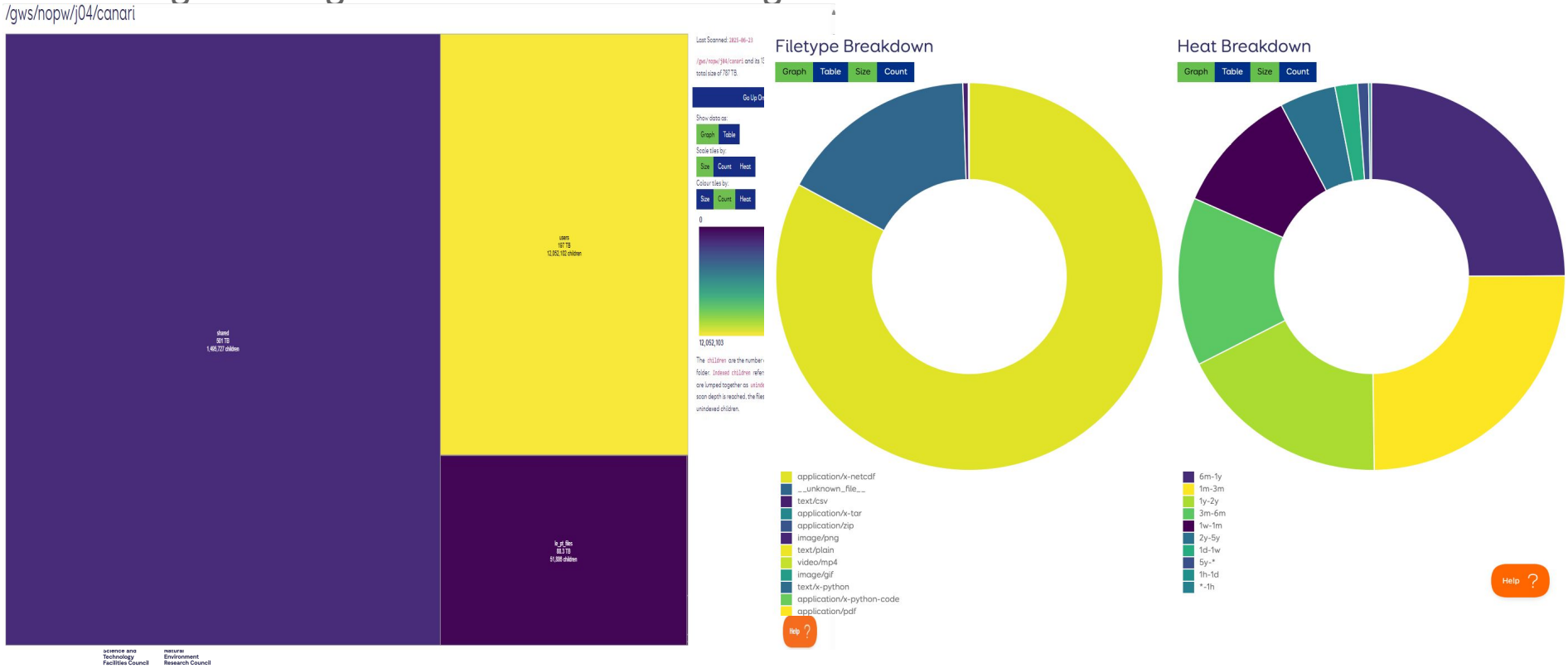
- Self-managed, not curated
- Projects portal - efficient resources management interface with consortia & infrastructure
- Enabling data/code sharing
- Efficient disk usage management- GWS scanner tool



UKRI Science and Technology Facilities Council Natural Environment Research Council

Efficient collaborative data processing space

Disk usage management tool- GWS managers



Data movement

-Globus (last 6 months)

- ~2.9PB of data moved (combination of to/from JASMIN)
- About 12 million data transfer tasks

-High-performance data transfer server

-New automated data migration service NLDS (Part 2)

Science and Technology Facilities Council

Natural Environment Research Council

Part 2

- Automated data migration -NLDS
- Batch processing LOTUS- new cluster/configuration
- Plan for the future

The Near-line Data Store

Issue: Users retain files in group workspaces (on POSIX disk) for longer than necessary, beyond the life of their projects, after they are useful
Solution: Development and deployment of NLDS
Outcome: Users' data can be moved to tape easily, retrievable as required

NLDS is an (STFC-developed) Hierarchical file management system

- Replacement for older JASMIN tape services, JDMA & Elastic Tape
- Users ingest files from hot (disk) to warm (S3 object) storage
- Data is catalogued upon ingest
- Once ingested to the cache, data is automatically backed up to cold (tape) storage
- Data may later be removed from warm storage via policies
- Retrieval from cold/warm storage is via same command
    - Difference from user POV is time lag if data needs to be moved out of cold storage
- Ingest and retrieval are asynchronous
- No proprietary formats for data storage

Deployment status: users testing now with live data, rolling out at the moment.

UKRI Science and Technology Facilities Council Natural Environment Research Council

LOTUS Efficiency improvements

Building on what we've learnt from operating LOTUS[1] (18k cores) over the past 10 years to better operate LOTUS2 (55k cores)

Issue: Nodes are split into different Slurm partitions (based on job size/max-wall-clock-time), leading to some nodes sitting idle while other partitions are oversubscribed
Solution: Switch from using partitions to QOS in Slurm, meaning that nodes are kept busier by the Slurm scheduler (instead of being dedicated to specific workload types), l
Outcome: Greater node efficiency, and some jobs/workloads to scale to much larger node counts

Issue: Nodes sometimes encounter issues which have to be either detected manually or
Solution: Use `HealthCheckProgram=nhc` to have nodes automatically be drained in the event of issues - checking for mount status, network issues etc
Outcome: Fewer jobs started on nodes with issues, leading to fewer wasted CPU hours

UKRI Science and Technology Facilities Council | Natural Environment Research Council

LOTUS Efficiency improvements

Issue: Spiky user workloads/periods of quieter usage mean that energy is wasted on running nodes which do not need to be powered on at the current time

Solution: Enable Slurm node power management, powering down nodes after a period of idle time, and powering them back on using IPMI when needed

Outcome: Fewer nodes sitting idle, power reduction of >60-100kW (at the expense of some jobs taking a few minutes to start if required to wait for nodes to power on)

T1 vs T2:
Similar amount of allocated nodes , but powering off ~40% of nodes (when idle) reduces platform power usage by ~60kW

Science and Technology Facilities Council
Natural Environment Research Council

LOTUS Future plans

- Net zero accounting & carbon budgets?
- Carbon-aware queues
- Pre-emptible jobs
- Improved data on workload efficiency for users (LLView?)
- Working directly with heavy LOTUS users to improve job performance/efficiency by changing their workloads/submission scripts
- ARM POC?

UKRI Science and Technology Facilities Council Natural Environment Research Council

JASMIN Future Plans

- Reducing storage server footprint, moving to flash-based systems (Pure FlashBlade, others?)
- Encouraging users to reduce data stored on group workspaces and move data to tape with NLDS
- Collaborations with new communities across UKRI to encourage efficient usage of data/compute already part of JASMIN
- Investigating opportunities for TRE deployment alongside JASMIN to make use of large datasets without data transfer

UKRI Science and Technology Facilities Council Natural Environment Research Council

JASMIN: support@jasmin.ac.uk
CEDA: support@ceda.ac.uk

Website - www.jasmin.ac.uk