National Federated Compute Services NetworkPlus

Spring Conference
26-27 February 2026

Federated data movement across the DRI

Alastair Basden, George Beckett, John Taylor

Durham University

COSMA

DiRAC

epcc

StackHPC

UKRI

Extending data federation to direct shared (no-copy) data access across the DRI

National Federated Compute Services NetworkPlus

Data access without data transfer

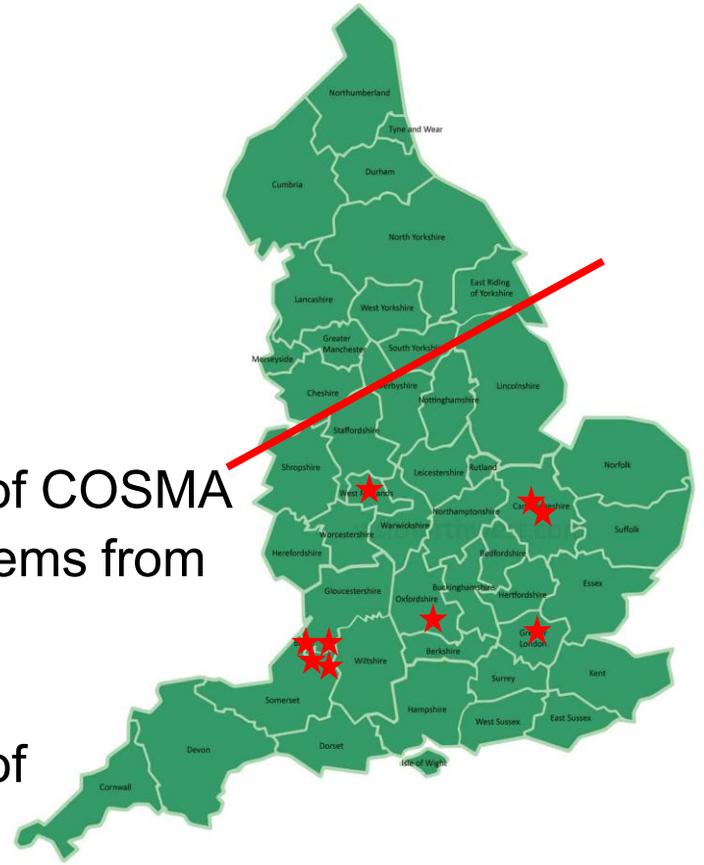
Between COSMA, DACF, Somerville

Existing practice, gaps and PoC

Survey of direct shared data access between DRI:
EBI, ELIXIR, Hartree, JASMIN, NFDI, EuroHPC, HDF5 streaming

Identify gaps where direct data access improves workflows

Develop use cases



Direct access of COSMA Lustre file systems from Somerville

Direct access of DACF project storage by COSMA

Related to a Phase-1 project: please fill in the questionnaire!

Project team

Project Leads:
Alastair, George, John

Co-leads:
Stig, Mark, Will

Project Team Members:
Billy

Quote or key statement:
Minimising the human-in-the-loop for efficient zero-copy data transfer

No-copy data access

A “pull” operation

No data staging (pushing)

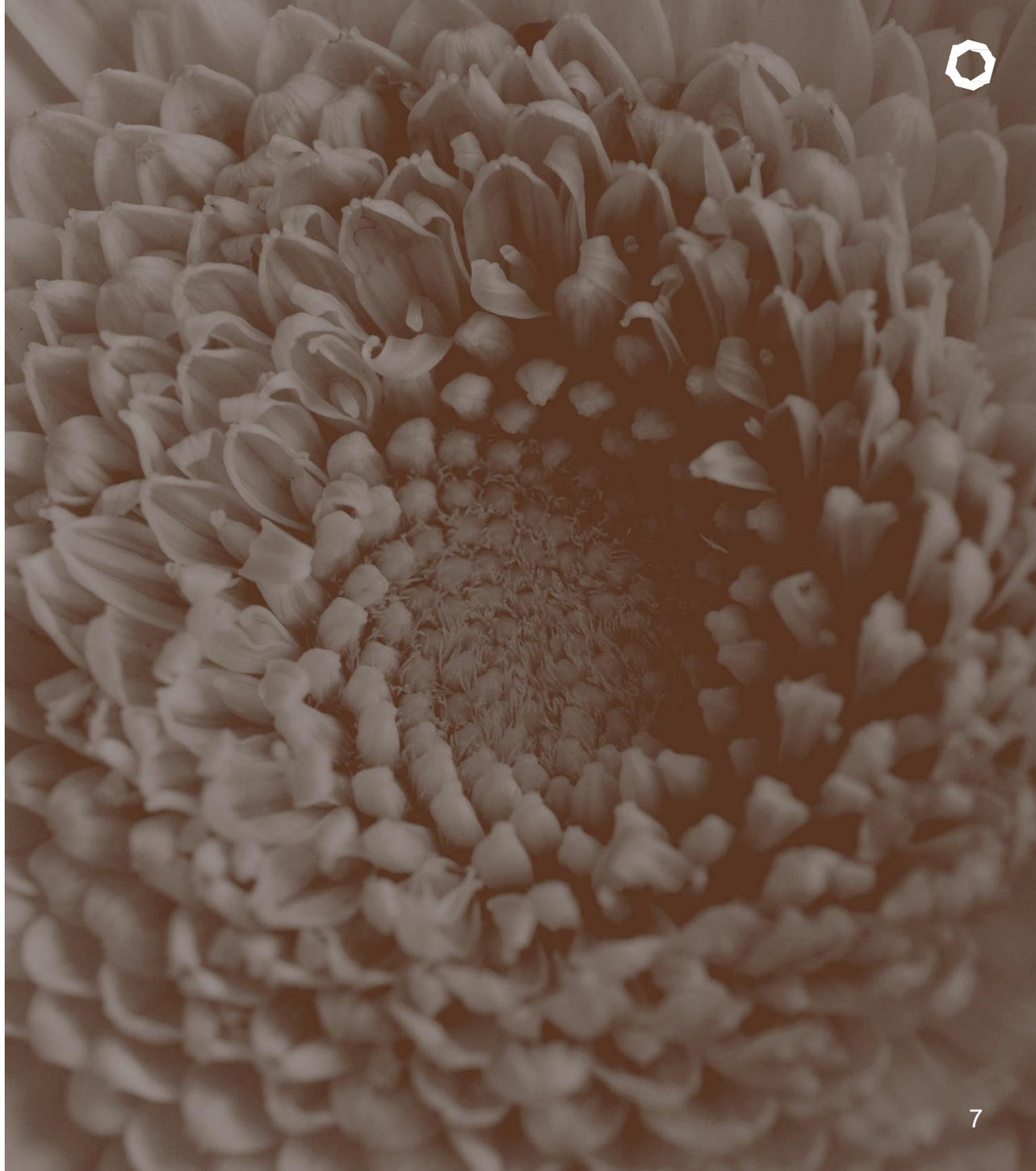
Partial dataset access: only access the necessary parts of a dataset

Use cases in astronomy, cosmology

Key tools: versitygw to provide an S3 interface to Lustre – to be posix mounted using s3fs/mountpoint/goofys

Gotchas and performance investigation

Appropriateness of solution



HPC-Cloud connection

Cross-mounting storage

- Prototype evaluation of Vera Rubin telescope workloads
  - Interactive analysis using direct data access
- Privacy and security implications
  - What precautions need taking?
- Instant access to large datasets without staging
  - But with higher latency
- StackHPC work package to import/export from Azimuth
  - From within the Azimuth cloud platform to HPC
  - User-friendly procedures, pre-configured applications

Community building

Data sharing workshop: May/June – knowledge exchange, training, methods

Interested? Let us know!

HPC Days 2026 conference @ Durham

NFCS+ roadmap building

Conclusions

Study of direct shared data access

Landscape survey

PoC implementations

Not a substitute for staged data transfer: HPC compute time is valuable

But has use cases for occasional data access: small fraction of large dataset – or one-off access