Provider's intro

Alastair Basden

Durham University
Institute for Computational Cosmology

DiRAC
High Performance Computing Facility

Alastair Basden
DiRAC / Durham University

IRIS compute providers

- RAL cloud

- Cambridge: CSD3/Arcus

- Imperial

- GridPP

- Somerville

- DiRAC

- Durham/COSMA (new)

Alastair Basden
DiRAC / Durham University

# Recent(ish) developments

- The past decade has seen:

  - Some shift from HPC to private cloud

    - And back again in some cases

- User approach in computing is changing

  - A fear of technical compute

    - "Is there an app for that?"

- New user communities

  - Increasing demand for capacity

Alastair Basden
DiRAC / Durham University

And more recently...

- Fast increases in costs
  - 50% increases within a month
    - And still going up

- Very short quotation validity periods
  - All because of AI

Alastair Basden
DiRAC / Durham University

Compute at Durham

- Traditional HPC: COSMA - DiRAC

  - HPC Hardware Lab (CPU, GPU, Networks, storage, composability, etc)

  - Disk and tape storage

  - Reinstalled in 2024 on an OpenStack backend
    - Due to end-of-life of CentOS7
    - With help from StackHPC

- New IRIS Azimuth cloud
  - Up and running, ready for test users
  - More hardware being added
    - In our immersion tank (ICHS)
      - Feel free to arrange a visit if interested

Alastair Basden
DiRAC / Durham University

Alastair Ba...
DiRAC / Durham University
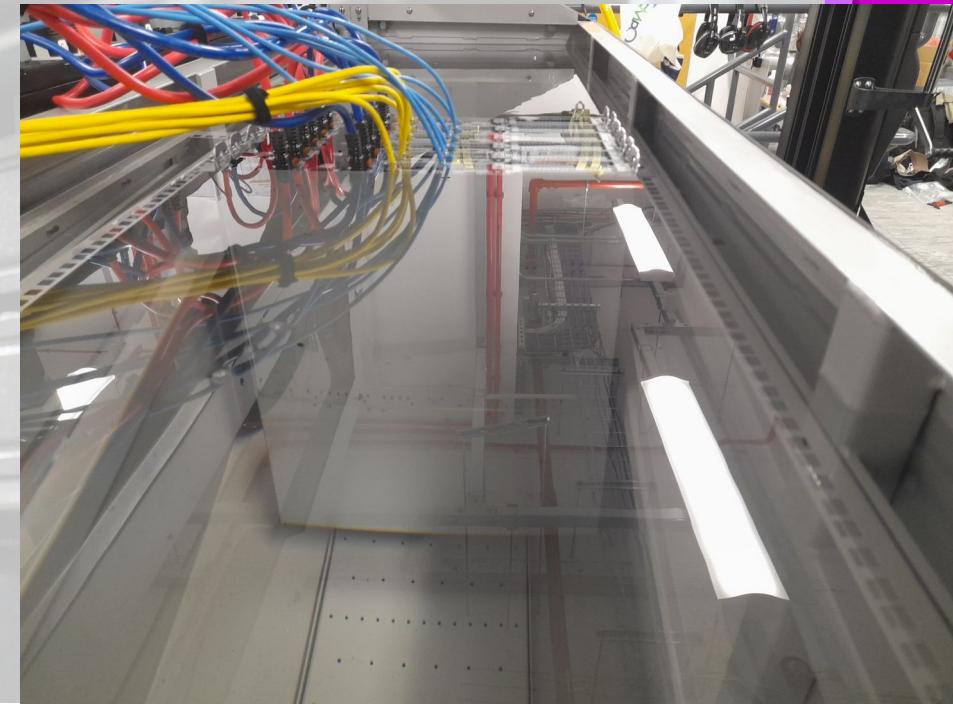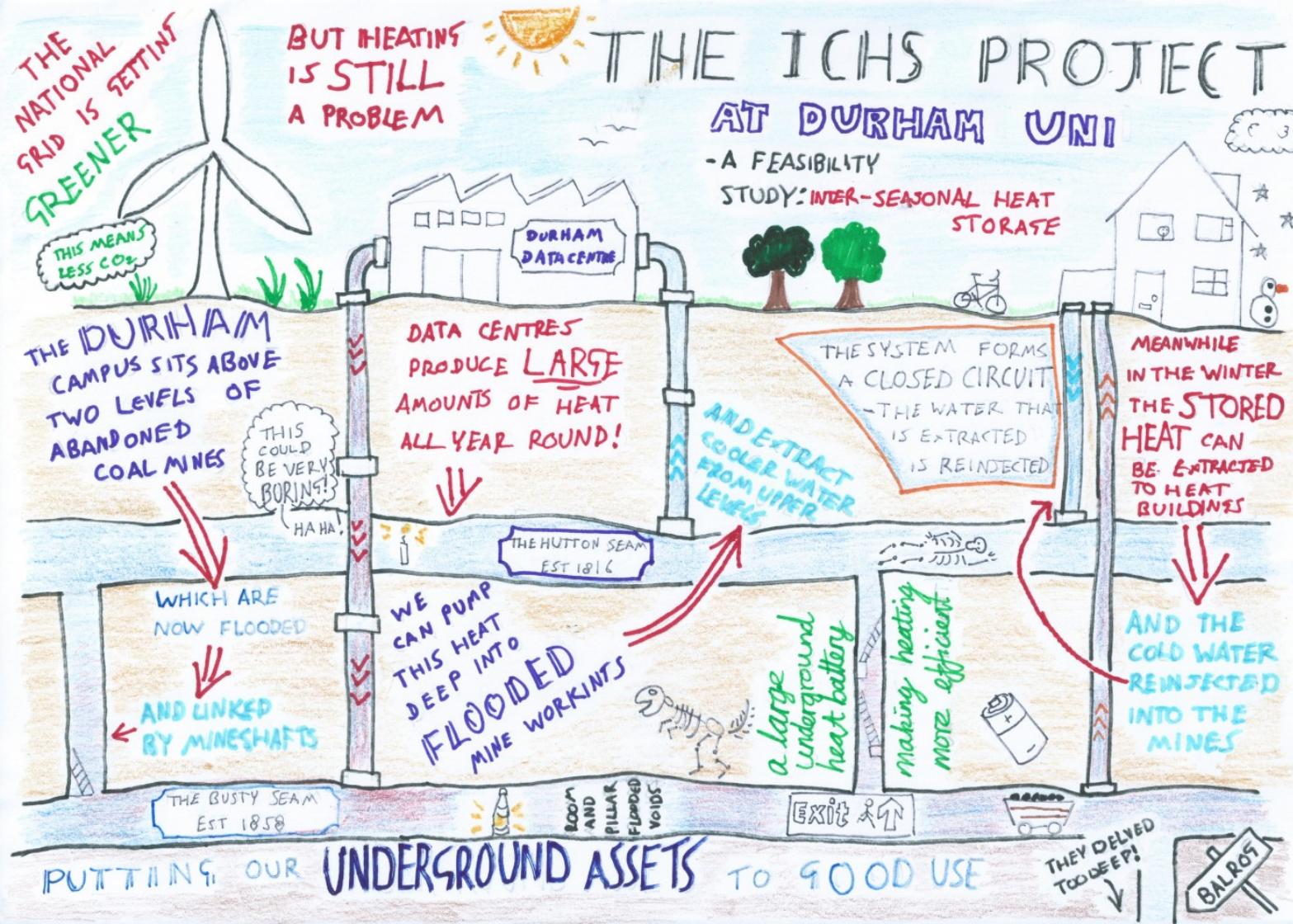
Azimuth

- Should be a familiar interface for some
  - Giving root access to virtual servers and slurm clusters
  - And Kubernetes clusters for Jupyter etc

Alastair Basden
DiRAC / Durham University
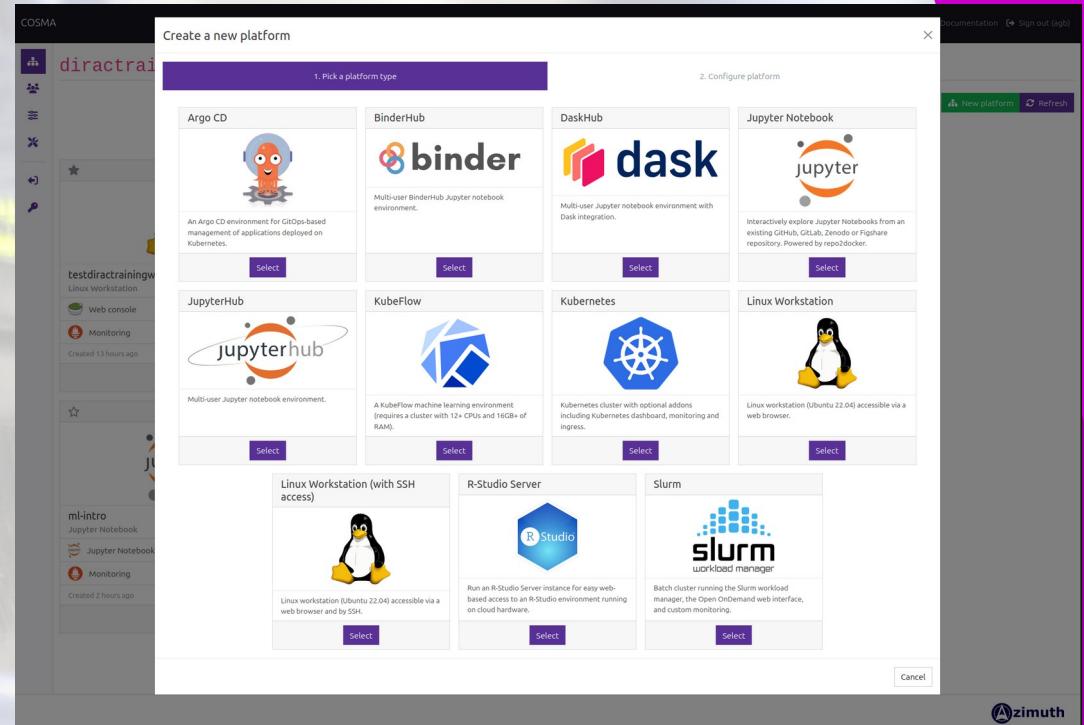
Knowing what you get...

- Is important

(Galaxy Classification CNN) across the systems:

Metric	Google Colab CPU (AMD EPYC 7B12)	IRIS Cloud CPU (AMD EPYC-Rome)	Google Colab GPU (Tesla T4)
Time per Epoch	~33 seconds	~6.5 seconds	~1.2 seconds
Step Time	~136ms / step	~32ms / step	~6ms / step
Total Training Time	27 min 30 sec	5 min 30 sec	59 seconds
	(1,649s)	(327s) 5x Faster	(59s) 28x Faster

Panel questions (to think about)

- What do you need from IRIS (apart from funding)

- In the next 2-3 years, what should IRIS prioritise?

- What do you use for forward planning? (roadmaps, usage, growth curves, etc)

  - Where are you currently planning blind?

- What user problems does federation solve today? (SSO, portability, data, accounting?)

  - Are users calling out for federation?

- Which policies could have a federated baseline? (data, encryption, quotas, acceptable use, incident response)

  - Which should remain autonomous?

- Have you registered for HPC Days?

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University

Alastair Basden
DiRAC / Durham University