



Science and
Technology
Facilities Council

The Role of the UK in the SKA Deployment of SRCNet

Ian Collier (RAL-STFC), John Garbutt (StackHPC), James Walder (RAL-STFC)
On behalf of the UKSRC Community

CHEP 2024, Kraków, Poland
21–25 October 2024



The Square Kilometre Array (SKA)

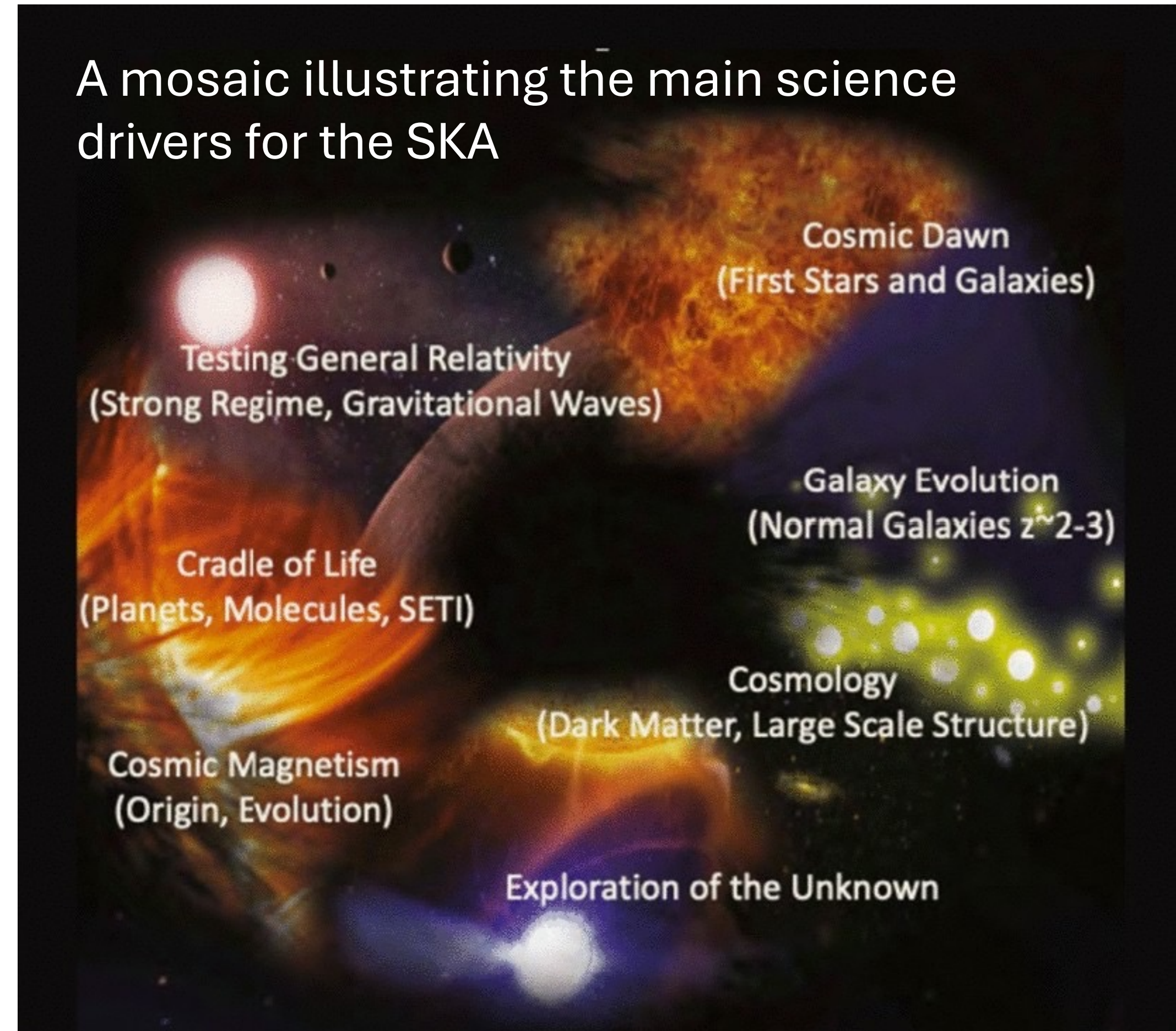
The Square Kilometer Array (SKA) Observatory (SKAO) is a next-generation radio astronomy facility which will cover the frequency range from 50 MHz to 15 GHz.



Composite image of the SKA telescopes, blending real hardware already on site with artist's impressions. credit: SKA Observatory

- More details: See talk [Ian Collier](#)

A mosaic illustrating the main science drivers for the SKA



Credit: SKA Observatory

SKAO HQ and Construction Efforts

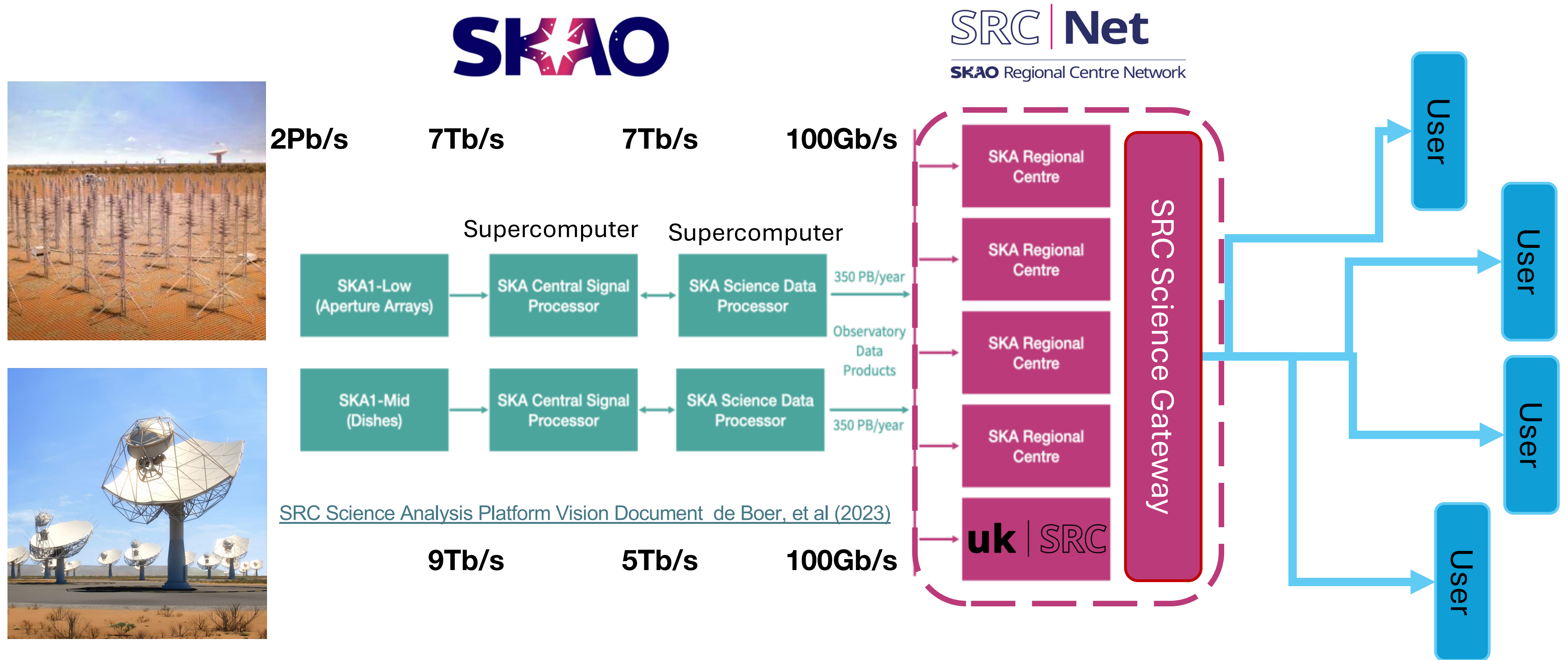


- SKA Observatory's international headquarters
 - Jodrell Bank, near Manchester
- Involvement across UK institutions in construction efforts; ~ 15% of the cost

- Contributions including:
 - Signal and Data Transport
 - Science Data Handling and Processing
 - Observatory Management and Control Cryostats
 - LOW Signal Processing System

Role of SRCNet

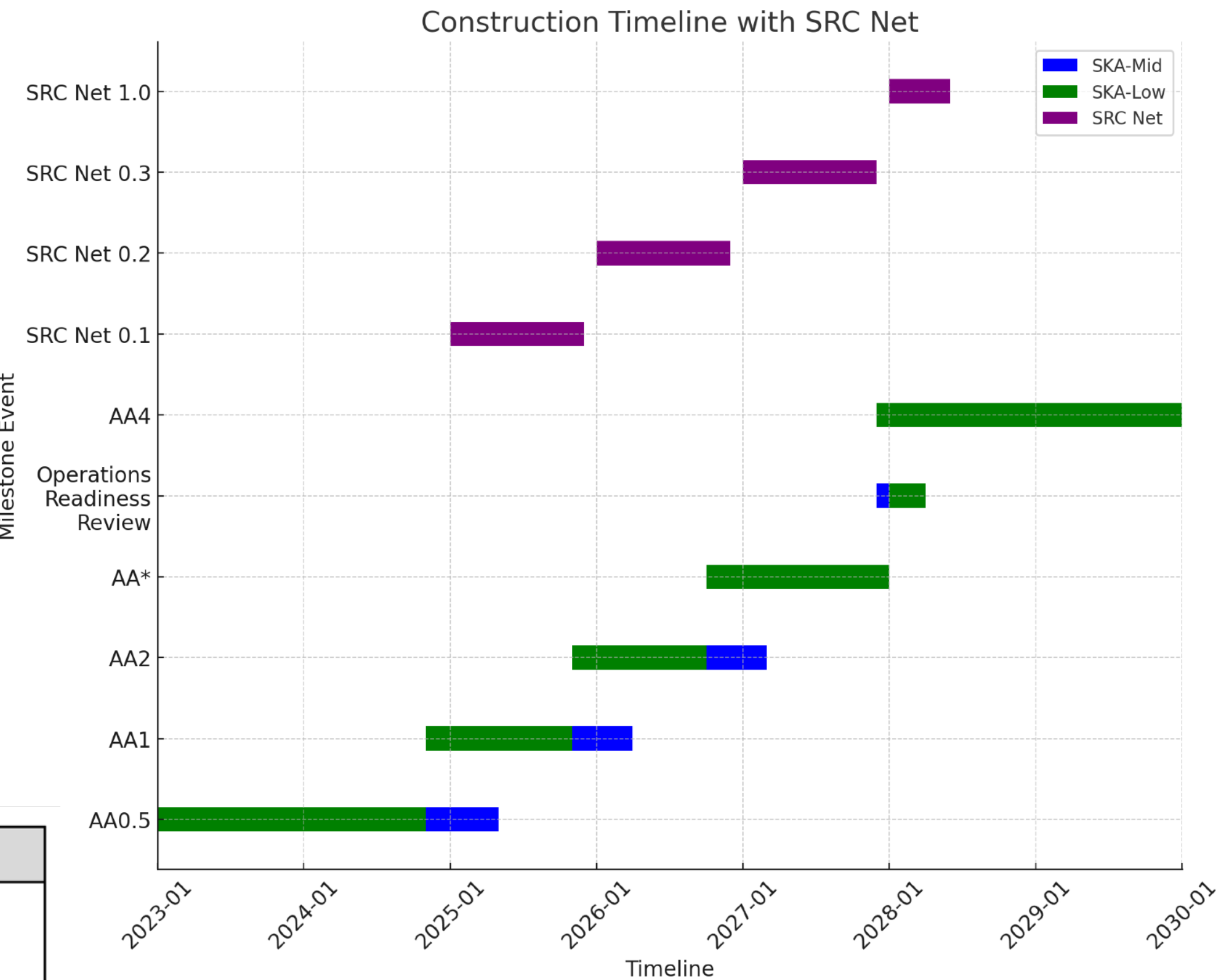
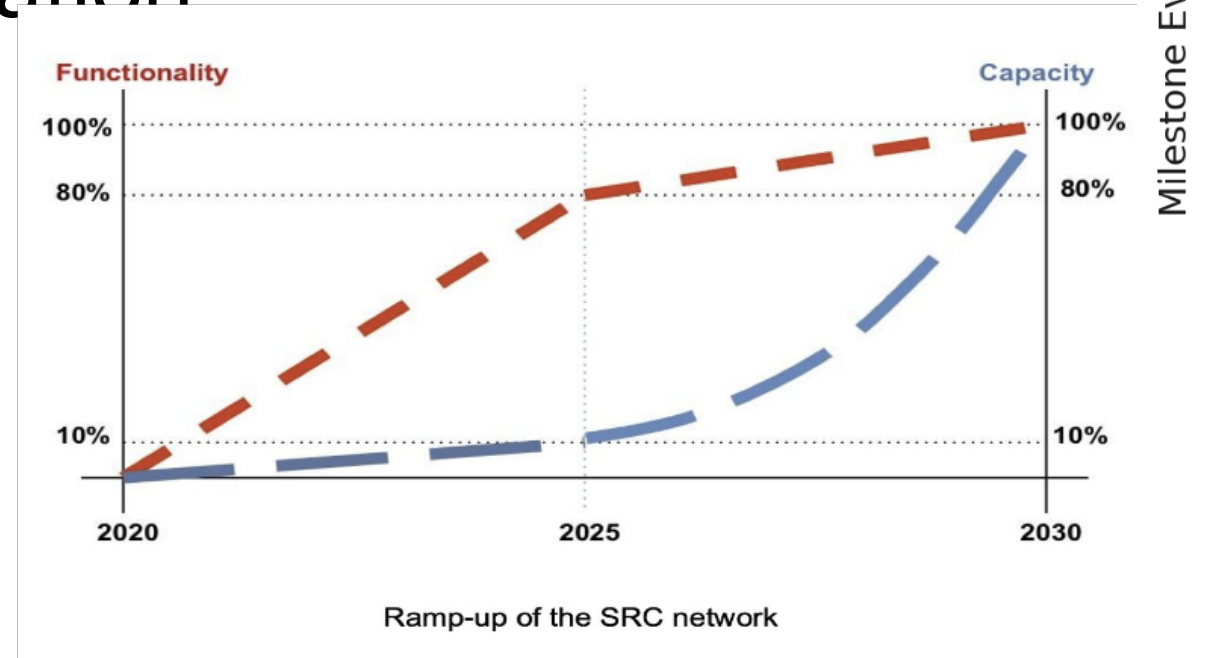
- SKA Regional Centre Network (SRCNet) (c.f. ~ WLCG in the context of HEP)



....SRCNet is the gateway for the science user communities to access the SKAO data and do science...

SRCNet Timeline

- SRCNet deployments: v0.1 up to v1.0 full operations
 - Increased capabilities at each phase,
 - Scaling up in Capacity closer to final readiness
- Represents first deployment of SRCNet,
 - concentrating on data dissemination
 - query, movement, and access via 'local' resources
- Initial deployment Dec '24
- Building to ~ 1EB data (on disk)



Milestone	Description	SRC Net Functionality	Scope (users)
SRCNet v0.1 First quarter of 2025	First version of SRCNet sites deploying common services and connecting via SRCNet APIs. Enable technical tests of the architectural implementation. [Added c.f. document] (Potentially Opportunity to engage SRCNet with AA0.5 data transfer and access.)	<ul style="list-style-type: none"> • Test data (and some precursors data) disseminated into a prototype SRC Net • Data can be discovered through queries to the SRC Net • Data dissemination to SRC nodes • Data can be accessed through a prototype data lake • Data replication. Data can be moved to a local SRC area where non-connected local interactive analysis portals (notebooks) could allow basic analysis • Unified Authentication System for all the SRCs • Visualisation of imaging data 	SRC ART members Members of SKA Commissioning team (potentially, but not required)



UK SRC infrastructure and services:
Supporting and facilitating UK science

Global SRC Network:
Developing and delivering the global SRCNet.

uk | SRC

SKAO Regional Centre United Kingdom



THE UNIVERSITY of EDINBURGH



Durham University

University of Hertfordshire **UH**



The University of Manchester



Summary: Delivering STFC's UK SKA Regional Centre Strategy

The UKSRC Strategy covers the SRC construction phase and early operations phase (2022 to 2030). The project timing is aligned with SKAO Array Assemblies and the global SRC network. This project is funded from January 2023 – December 2025.

3 Pillars:

UK SKA Regional



Elchinator, Pixabay

Developing digital research infrastructure

Bespoke UK-based computational and data facilities, tools, and services will contribute to the analysis of 700PB of data generated per year by the SKA telescopes.

UK Science Community



Robert Braun

Strengthening the UK astronomy community

UK astronomers will have opportunities to inform the UKSRC's development and to enhance their skills in preparation for the deployment of the SKA telescopes.

Global SRC Network



SKAO

Collaborating internationally

The UKSRC team working with a global network of 14 nations and the SKA Observatory to develop interoperable functionalities to find, access, manipulate and visualise SKA Data products.

Components of UKSRC

- SKA adopted SAFe agile methodology for the management of this complex project.
- SRCNet Agile Release Train (ART):
 - Cross-functional teams of SRCs working towards developing infrastructure and tools for SKA data handling.
 - Program Increment (PI) Planning: Ensures alignment of goals, planning of tasks, and addresses dependencies.
- UK Team's Role:



Data Access & Compute

Cloud & Data metadata & archive



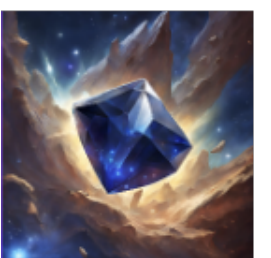
Purple

AAI, data logistics, policy, PerfSONA



Teal

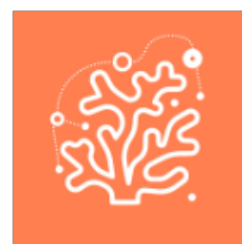
Science Platform and workflow development



Sapphire

Science user support, training, and community engagement

International teams



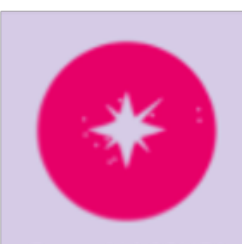
Coral

Tests node deployment and support the tech development to build a performant SRCNet.



Tangerine

To deliver the SRCNet Science Gateway which provides users with access to SRCNet services



Magenta

SRCNet Rucio data management, data management APIs



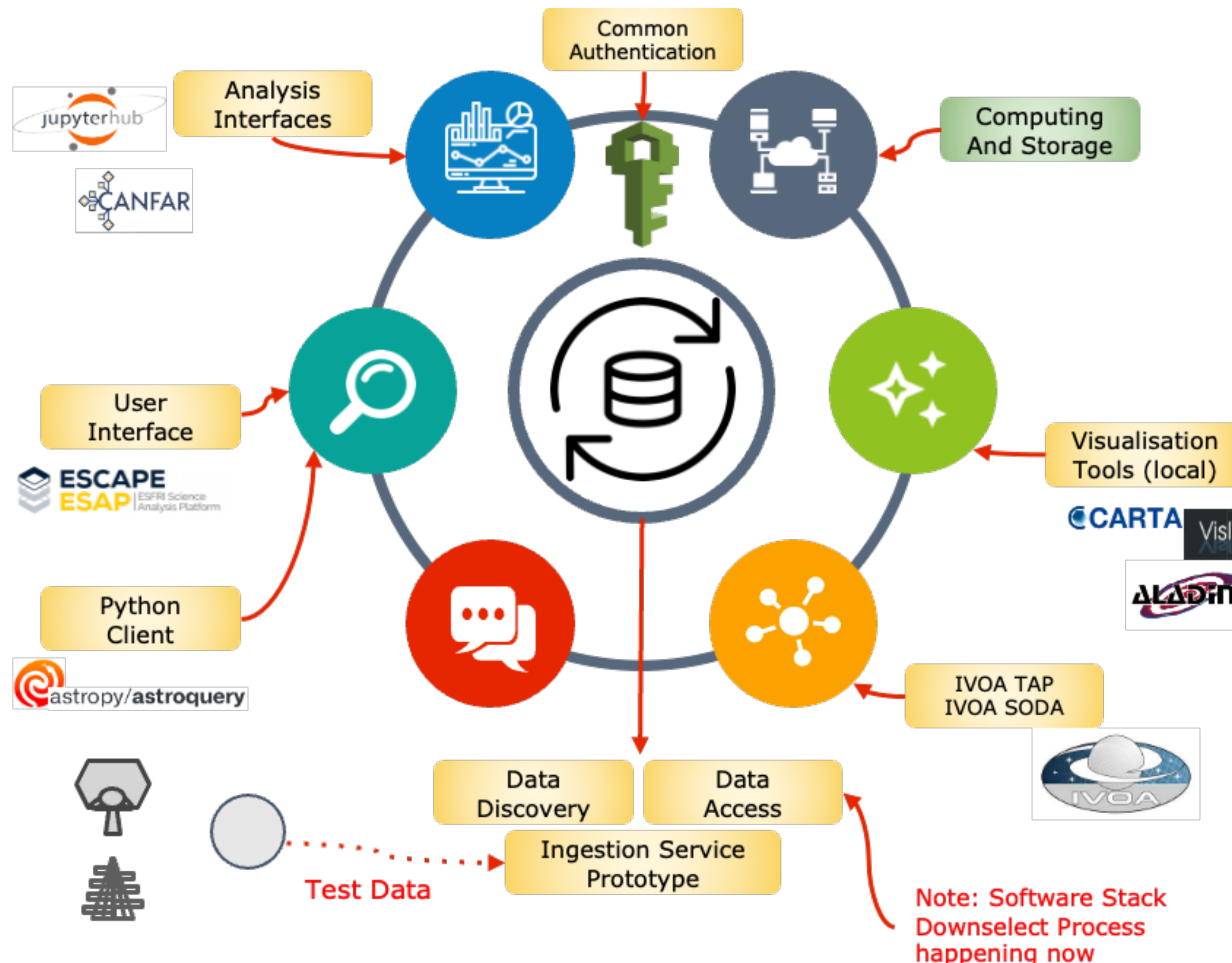
Program team

Responsible for the running of the ART

- *The Scaled Agile Framework (SAFe) is a methodology designed to help large, cross-functional teams work in an agile manner, allowing for efficient collaboration, synchronization, and value delivery.*

Software Stack (v0.1)

- SRCNet v0.1 defines a common set of tools / services for initial prototyping and to allow demonstration of:
 - Ingestion, Data Movement, Data life-cycle, Execution of “known” science use cases on SKA test data



UK Deployment Plan: SRCNet v0.1

- For v0.1 concentrate deployment at Rutherford Appleton Laboratory (RAL) STFC, near Oxford, UK (i.e.. same location as the WLCG UK Tier-1).
- Deployment teams from RAL, Cambridge, Manchester StackHPC contributing.
- GitOps style approach recommended (e.g. ArgoCD/FluxCD, k8s):

D Deployments

Per site deployments.
[Read more](#)

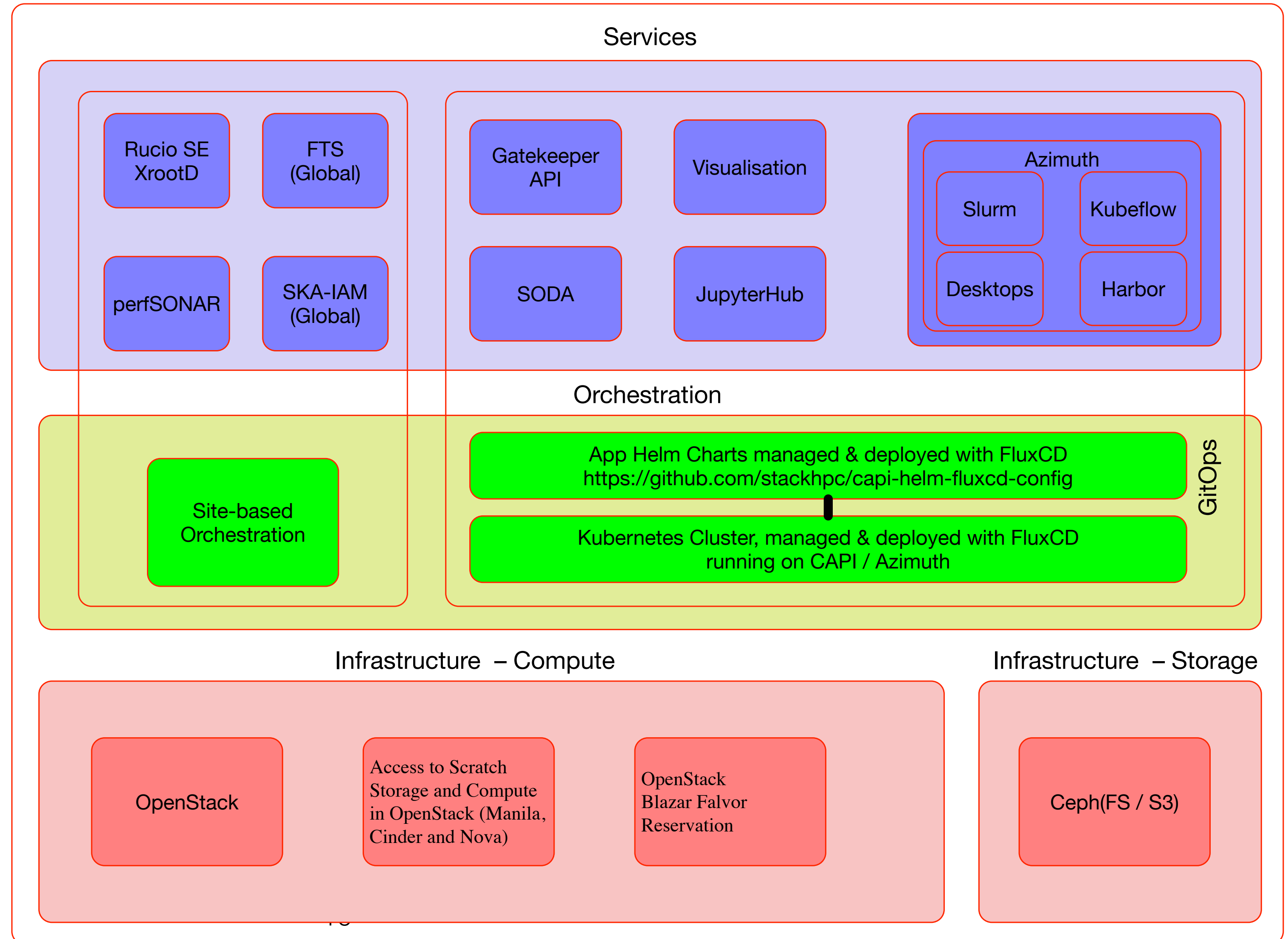
Subgroups and projects Shared projects Inactive

Search (3 character minimum)

Name

Name	Last update
UKSRC	
apps	5 hours ago
bin	3 weeks ago
clusters	8 hours ago
components	8 hours ago
infra	1 day ago
sites	2 hours ago
.gitignore	1 week ago
CODEOWNERS	2 weeks ago
LICENSE	4 months ago
README.md	3 months ago
requirements.txt	4 months ago

- > CHSRC
- > CNSRC
- > ESPSRC
- > ITSRC
- > SKAOSRC
- > SWESRC
- > UKSRC



STFC-cloud

- STFC Cloud is a dedicated cloud infrastructure which provides access to elastic compute resources for users across the facilities provided by STFC and partner organisations;
 - CPU and GPU (plus disk) available through IRIS (<https://www.iris.ac.uk>):
 - IRIS: Provides hardware & software to scientific computing communities, and supports, represents these communities and scrutinises resources allocated to them
- Deployment of 'local' services to SRCNet through GitOps approach on k8 via OpenStack
- Other services: IAM, FTS, Rucio Storage Endpoint managed via existing RAL infrastructure and orchestration

OpenStack to the rescue

- OpenStack is an open-source cloud platform
- Strong multi-tenancy guarantees
- APIs for compute, network and storage
 - Dashboard and command-line interface
- DevOps tools for OpenStack and platforms
 - Code reviewed changes
 - Continuous integration and delivery



{ REST }





- “Azimuth provides a self-service portal for managing long(er)-lived cloud resources - "science platforms" - with a focus on simplifying the use of cloud for scientific computing and artificial intelligence (AI) use cases. It is currently capable of targeting OpenStack clouds”

- Runs on OpenStack

- Integrates with indigo-IAM (and others) AAI

- Built-in monitoring (Prometheus, Grafana)

- StackHPC develops OpenStack capabilities for research computing use cases.

- Through extensive experience, we understand HPC and cloud. We know the needs and the shortcomings of each paradigm.

- StackHPC

- Looking at solutions e.g. Blazar for resource reservation / allocation

The screenshot displays the 'Create a new platform' dialog in the Azimuth self-service portal. The dialog is titled 'Create a new platform' and has a close button (X) in the top right corner. It is divided into two main sections: 'Pick a platform type' and 'Configure platform'. The 'Pick a platform type' section contains seven platform options, each with a logo, a brief description, and a 'Select' button:

- Jupyter Notebook:** Interactively explore Jupyter Notebooks from an existing GitHub, GitLab, Zenodo or Figshare repository. Powered by repo2docker.
- JupyterHub:** JupyterHub provides a multi-user environment where authorized users can have their own notebook server.
- Kubernetes:** Kubernetes cluster with optional addons including Kubernetes dashboard, monitoring, ingress and application dashboard.
- Linux Workstation:** Linux workstation (Ubuntu 20.04) accessible via a web browser.
- Linux Workstation (with SSH access):** Linux workstation (Ubuntu 20.04) accessible via a web browser and by SSH.
- Pangeo:** Pangeo provides a multi-user environment where authorized users can have their own notebook server.
- Slurm:** Batch cluster running the Slurm workload manager, the Open OnDemand web interface, and custom monitoring.

The 'Configure platform' section is currently empty. At the bottom right of the dialog, there is a 'Next' button with a right-pointing arrow. The background shows the main portal interface with a sidebar menu containing 'Platforms', 'Quotas', 'Project metrics', 'Advanced', 'Switch tenancy', and 'SSH public key'. The user profile 'matt-dev' is visible in the top right corner.

Azimuth

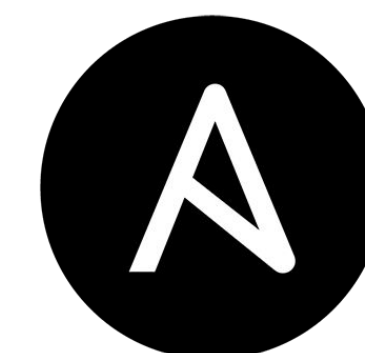
- Web portal for self-service platforms
- Configurable catalogue of curated platforms
 - StackHPC reference platforms
 - Site-optimised platforms
 - Automation using standard tools
- Platform services exposed using Zenith
 - Tunneling application proxy
 - No public IP required
 - SSO and TLS
- Manage platform users with Keycloak



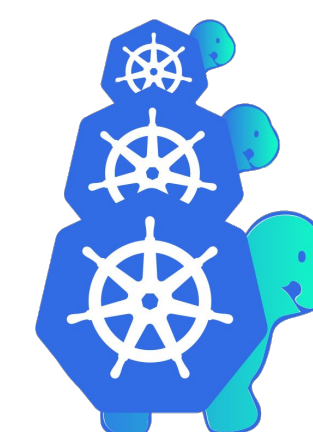
StackHPC



azimuth



ANSIBLE



**Kubernetes
Cluster API**



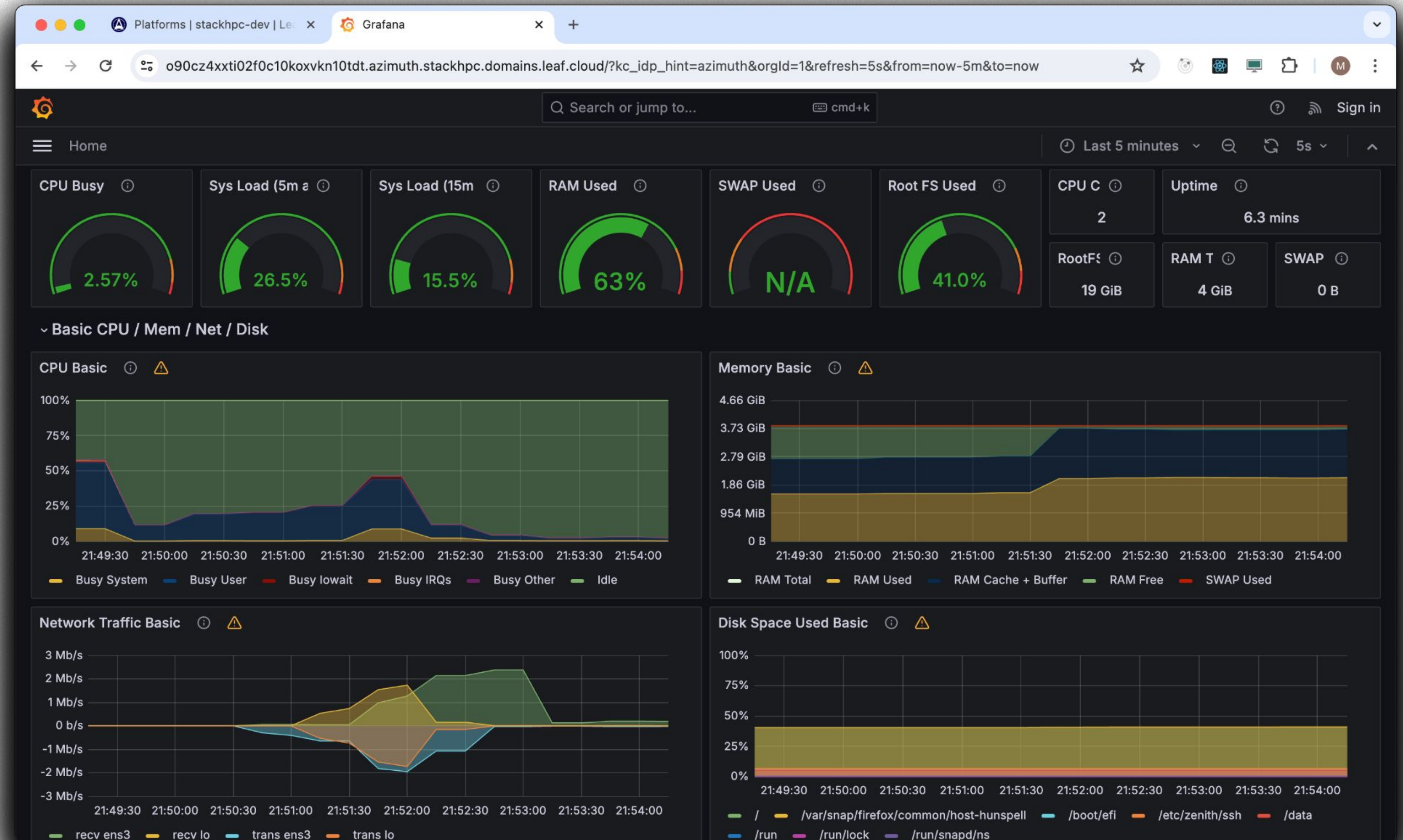
OpenTofu

Workstation

- Web-based shell and desktop
- Secure access via Zenith
- Monitoring stack
- Platform lifetime
- User gets sudo
- Apptainer and podman
- Optional SSH with public IP
- Access to project share



StackHPC

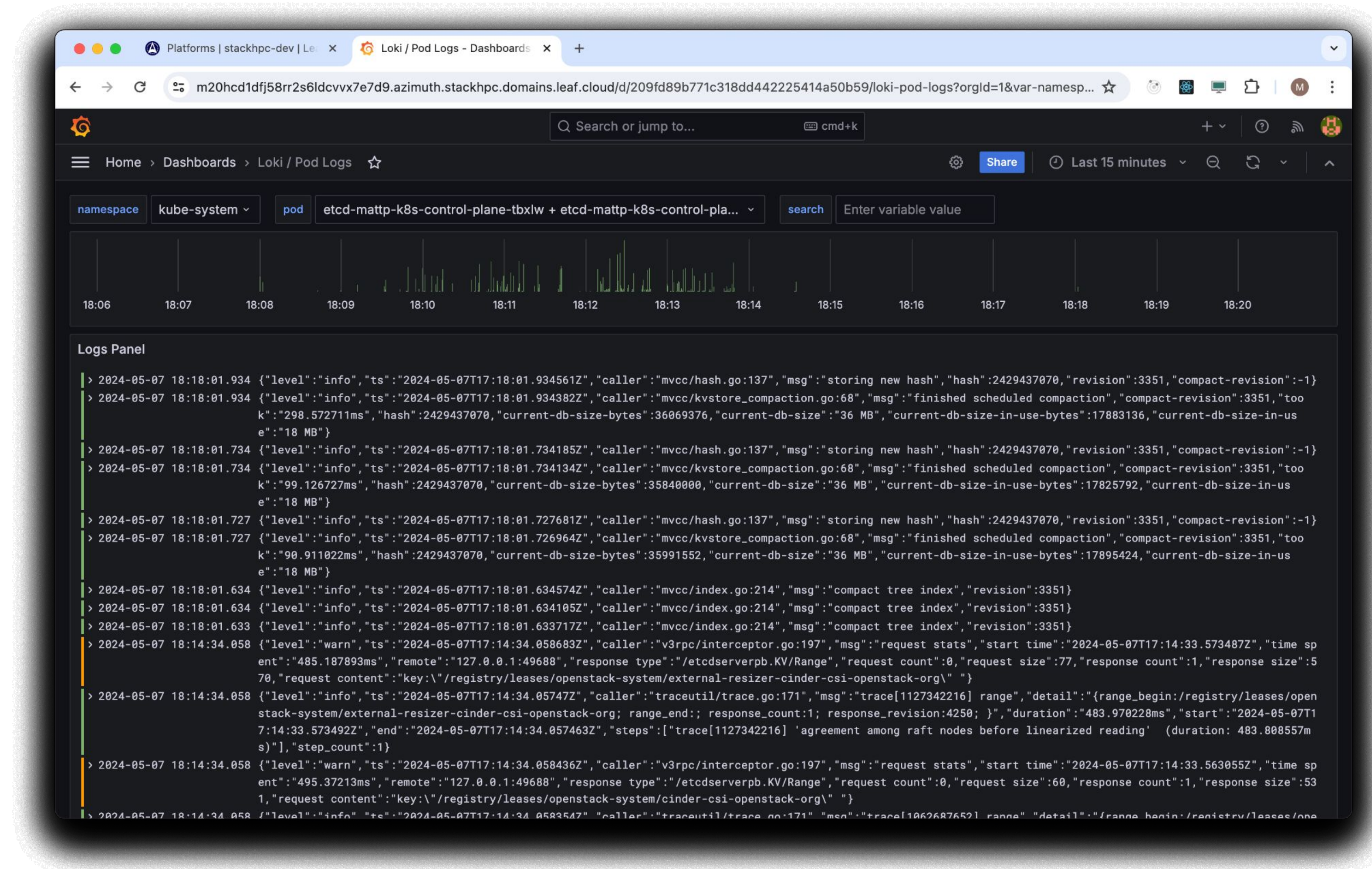


Kubernetes



StackHPC

- Built on Cluster API
- HA control plane
- Multiple node groups
- Download kubeconfig
- Autoscaling, autohealing
- Rolling upgrades
- NVIDIA GPU + NIC support
- Kubernetes dashboard
- Monitoring and logging
- Secure access via Zenith

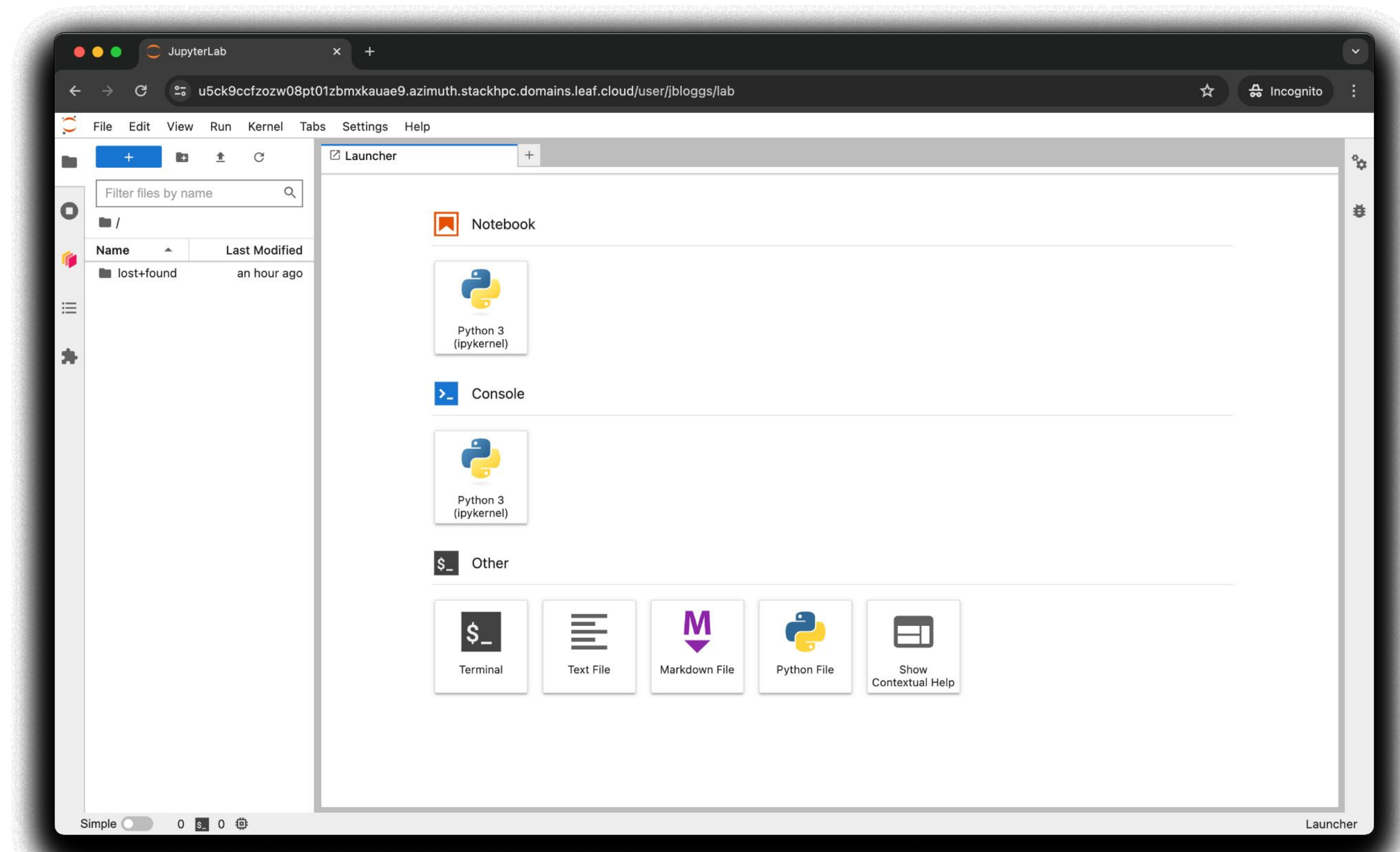


DaskHub



StackHPC

- Runs on Kubernetes cluster
- Each user gets their own notebook server
- Secure access via Zenith
- Grant access to external users using tenancy Keycloak realm
- Dask clusters for parallel computing using Dask Gateway

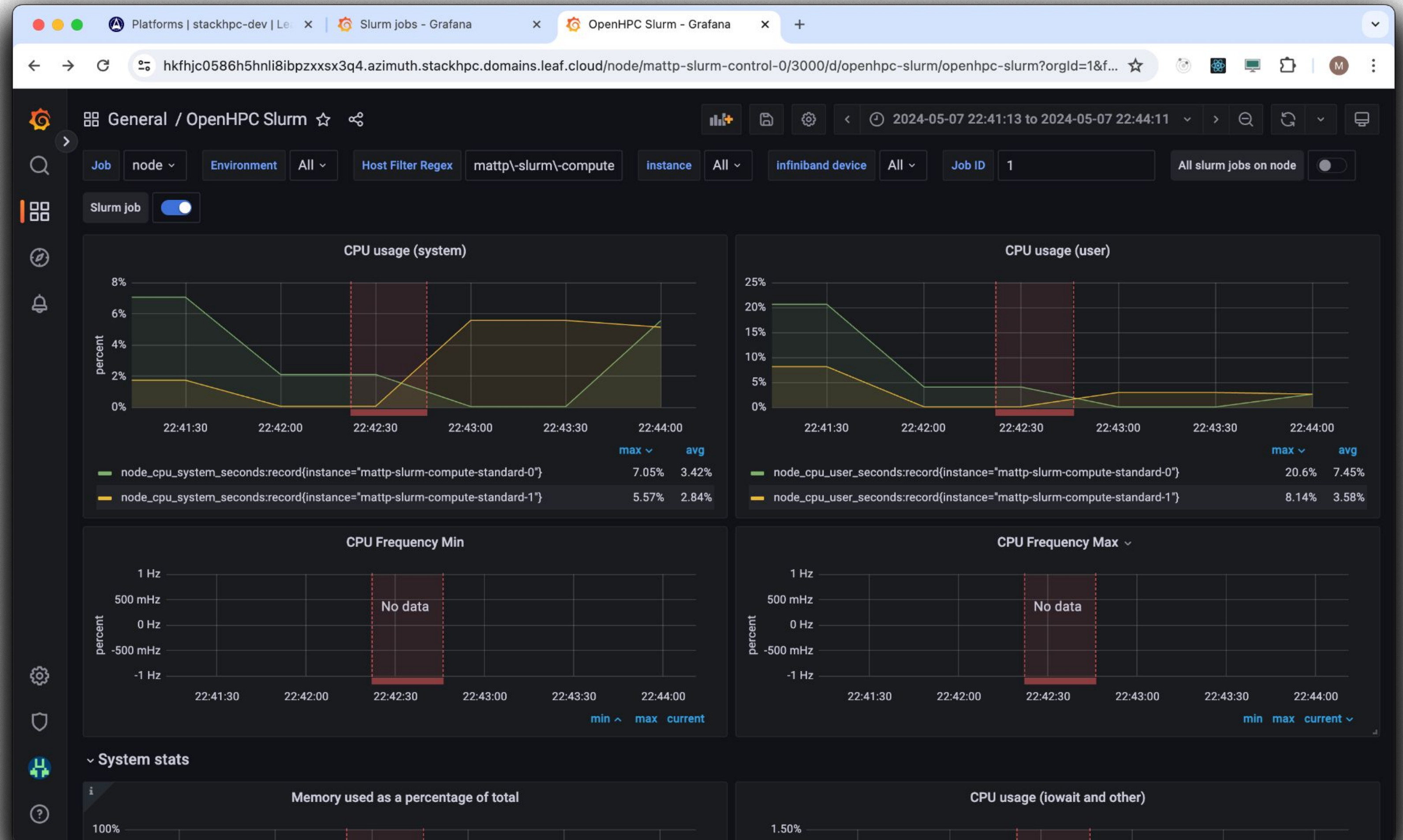


On-demand Slurm



StackHPC

- Single-user Slurm cluster
- No waiting for queues
- Image-based updates
- OpenHPC, Apptainer, EESSI
- Open OnDemand UI
- Job aware monitoring
- Access to project share



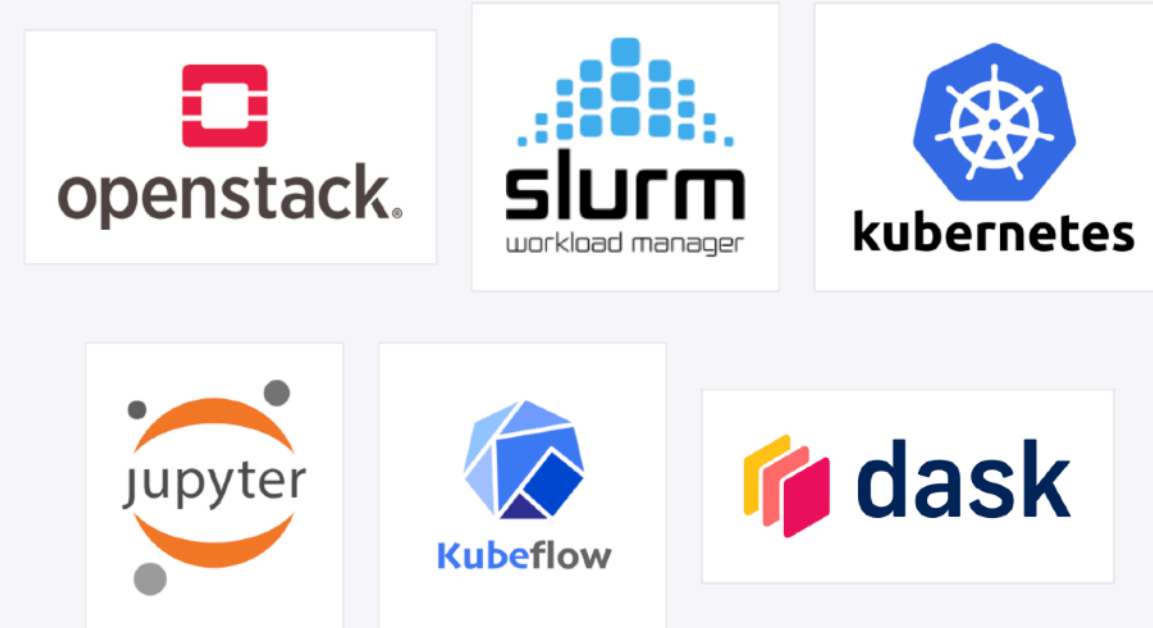
Arcus @ Cambridge

Welcome to Azimuth, a portal to help you access the platforms and storage that you need to get science done.

Using Azimuth, you can quickly create the platforms you need for your science. Advanced users can also manage virtual machines directly, provision and attach cloud storage and expose machines to the internet by connecting external IP addresses.

For help getting started, check out the [documentation](#).

[My Tenancies](#)



Create a new platform

1. Pick a platform type 2. Configure platform

JupyterHub
Multi-user Jupyter notebook environment.

Platform name
Platform name
Must contain lower-case alphanumeric characters and dash (-) only.

Kubernetes cluster
Select a Kubernetes cluster... +
The Kubernetes cluster to deploy the platform on.

App version
0.4.0
The version of the application to use.

Notebook CPUs
1
The number of CPUs to allocate to each user notebook.

Notebook RAM
2 GB
The amount of RAM to allocate to each user notebook.

Notebook storage
10 GB
The amount of storage to allocate to each user notebook. This is dynamically allocated from the tenancy quota.

Cancel Back **Create platform**

ska-src-cambridge

Access to platforms is managed using the [identity provider](#) for the tenancy.

UNHEALTHY

- ju-test-ruicio
Kubernetes
Kubernetes Dashboard
Monitoring
Created 20 days ago
Expires in 3 days

Platform	Status	Created	Expires
slurm bb23-lapcat-4mm	READY	Created 3 months ago	Expires in 8 years
slurm bo307-k8s	READY	Created 4 months ago	Expires in 27 years
slurm bo307-k8s-test01	READY	Created 2 months ago	Expires in 24 days
slurm bo307-test02	READY	Created 29 days ago	Expires in 1 month
slurm fq-test-slurm	READY	Created 2 months ago	Expires in 9 months
slurm hy297-singularity	READY	Created 7 months ago	
slurm hy297-test	READY	Created 7 months ago	
slurm iul-srcw	READY	Created 7 days ago	Expires in 22 days
slurm iul-srcwmw	READY	Created 7 days ago	Expires in 22 days
slurm lofar-slurm-demo	READY	Created 16 days ago	Expires in 2 months
slurm mjh-carta	READY	Created 5 months ago	Expires in 6 months
slurm mjh-slurm-test	ERROR	Created 2 days ago	Expires in 27 days
slurm ylb71-file-transfer	READY	Created 1 day ago	Expires in 9 months

Create a new platform

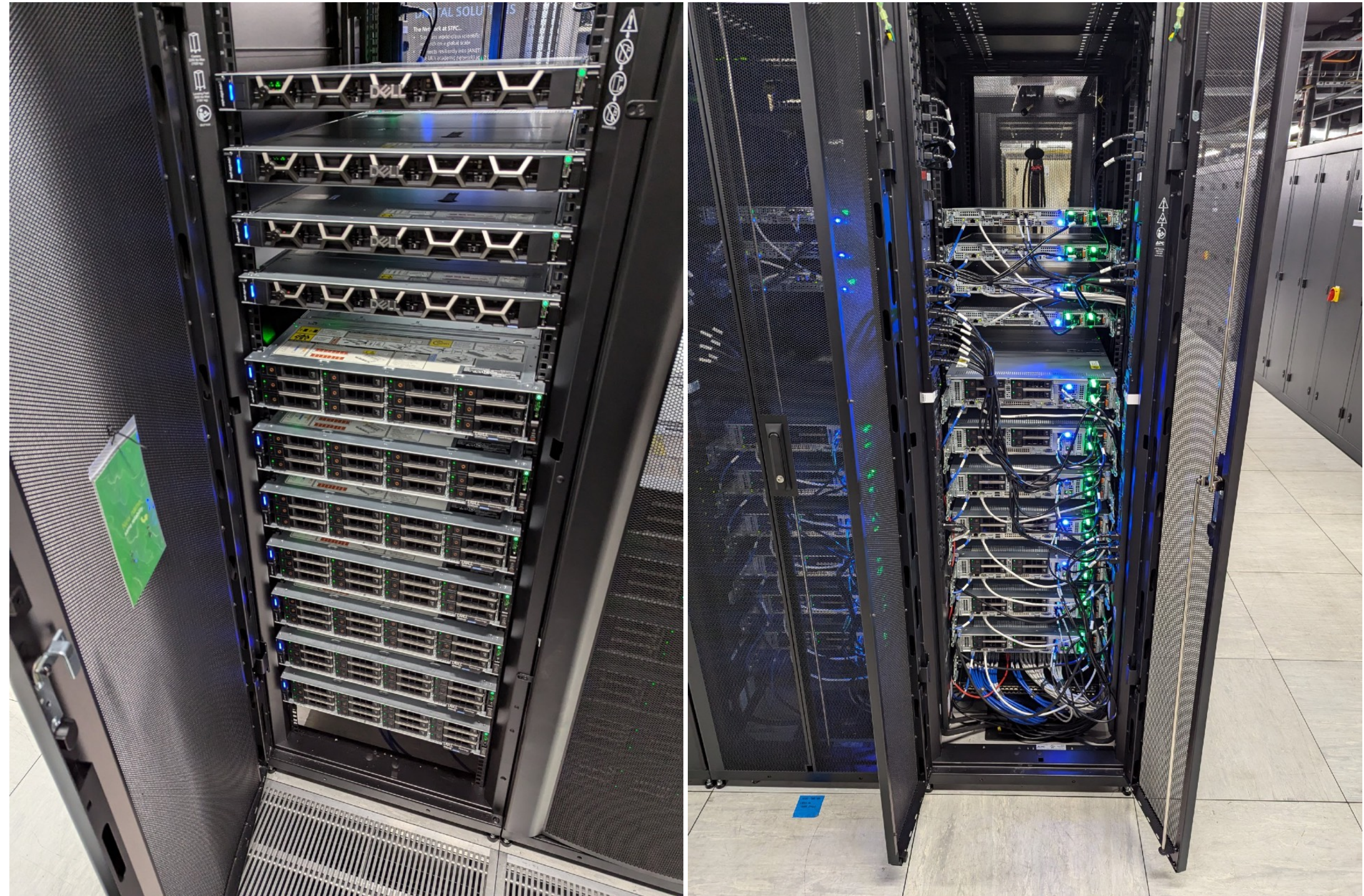
1. Pick a platform type 2. Configure platform

Argo CD An Argo CD environment for GitOps-based management of applications deployed on Kubernetes.	BinderHub Multi-user BinderHub Jupyter notebook environment.	DaskHub Multi-user Jupyter notebook environment with Dask integration.	HuggingFace LLM A generative AI chatbot service backed by a HuggingFace large language model (requires a GPU node group on the target cluster).
Jupyter Notebook Interactively explore Jupyter Notebooks from an existing GitHub, GitLab, Zenodo or Figshare repository. Powered by repo2docker.	JupyterHub Multi-user Jupyter notebook environment.	KubeFlow A KubeFlow machine learning environment	Kubernetes Kubernetes cluster with optional addons including Kubernetes dashboard, monitoring and ingress.
Linux Workstation Linux workstation (Ubuntu 22.04) accessible via a web browser.	Linux Workstation (with SSH access) Linux workstation (Ubuntu 22.04) accessible via a web browser and by SSH.	R-Studio Server Run an R-Studio Server instance for easy web-based access to an R-Studio environment running on cloud.	Slurm (CephFS home) Batch cluster running the Slurm workload manager, the Open OnDemand web



Initial Storage Deployment

- Initial 4PB (usage) storage @ RAL
 - New Ceph-based cluster
- To be provisioned with CephFS, and explore also S3, XrdCeph, etc
- Connect to Rucio SRC Network using XRootD, each server with 100Gb/s NICs
 - Installed within the sites' DMZ
 - perfSONAR physically close for monitoring
- Connecting the data in the Rucio managed storage to the User areas within Openstack is ongoing work:
 - R/O access
 - Authorisation (tokens)
 - Embargos
 - Performance
- SRCNet v0.1 gives a platform to test and explore this; use of benchmarking tests to validate performance.

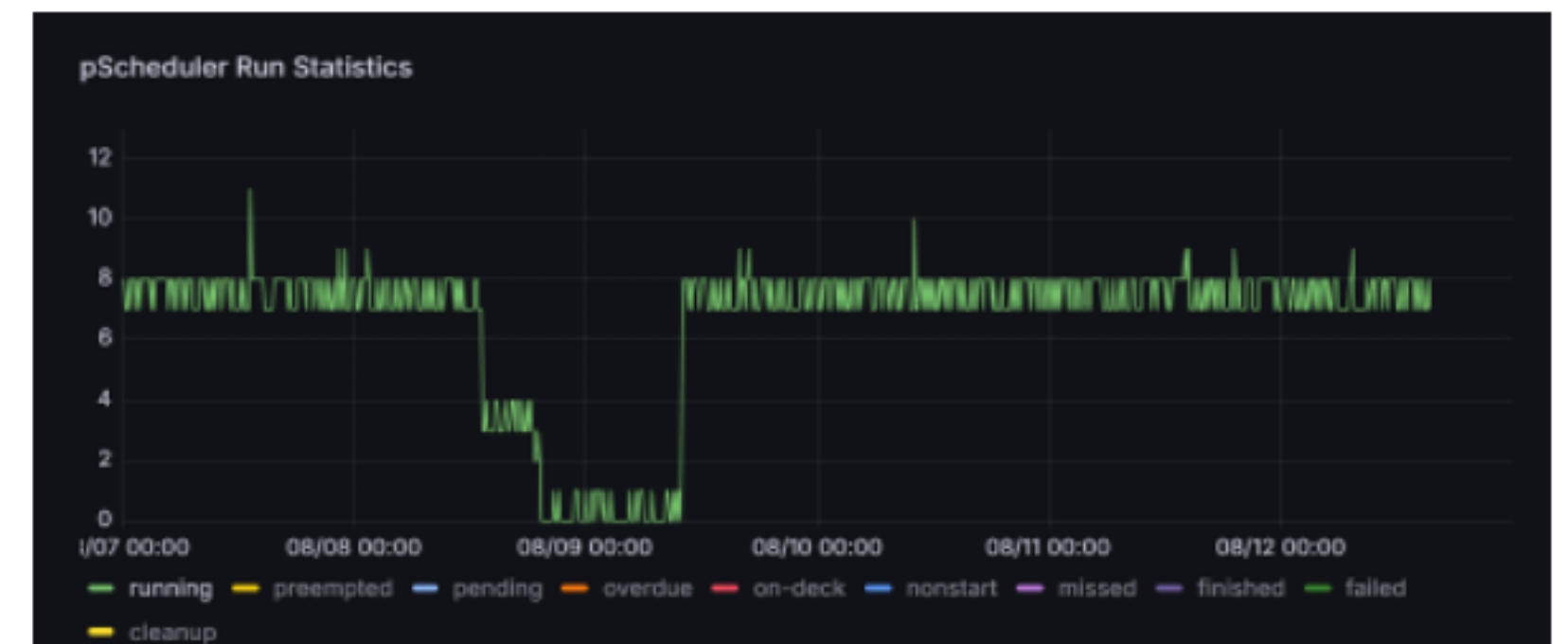
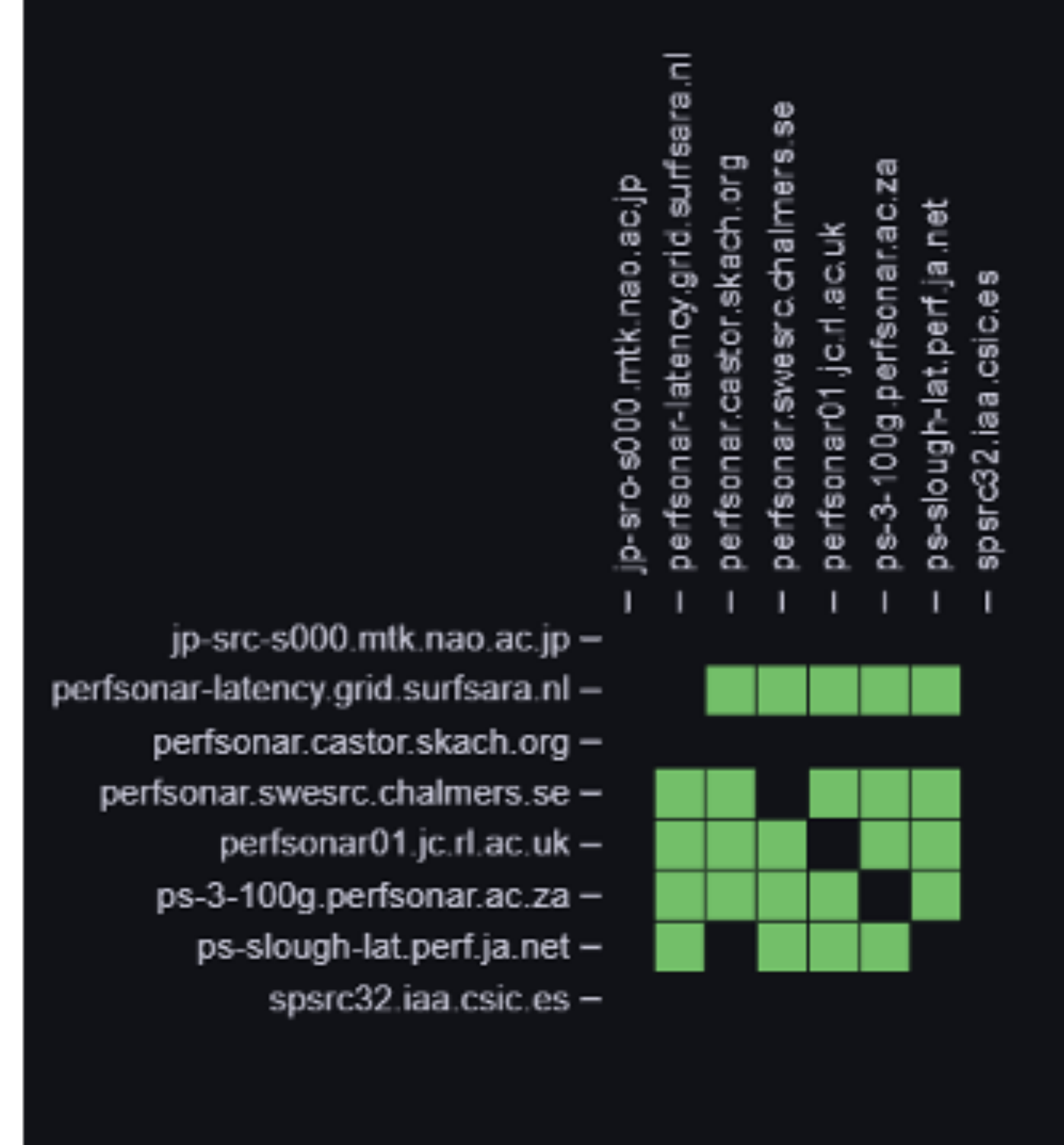
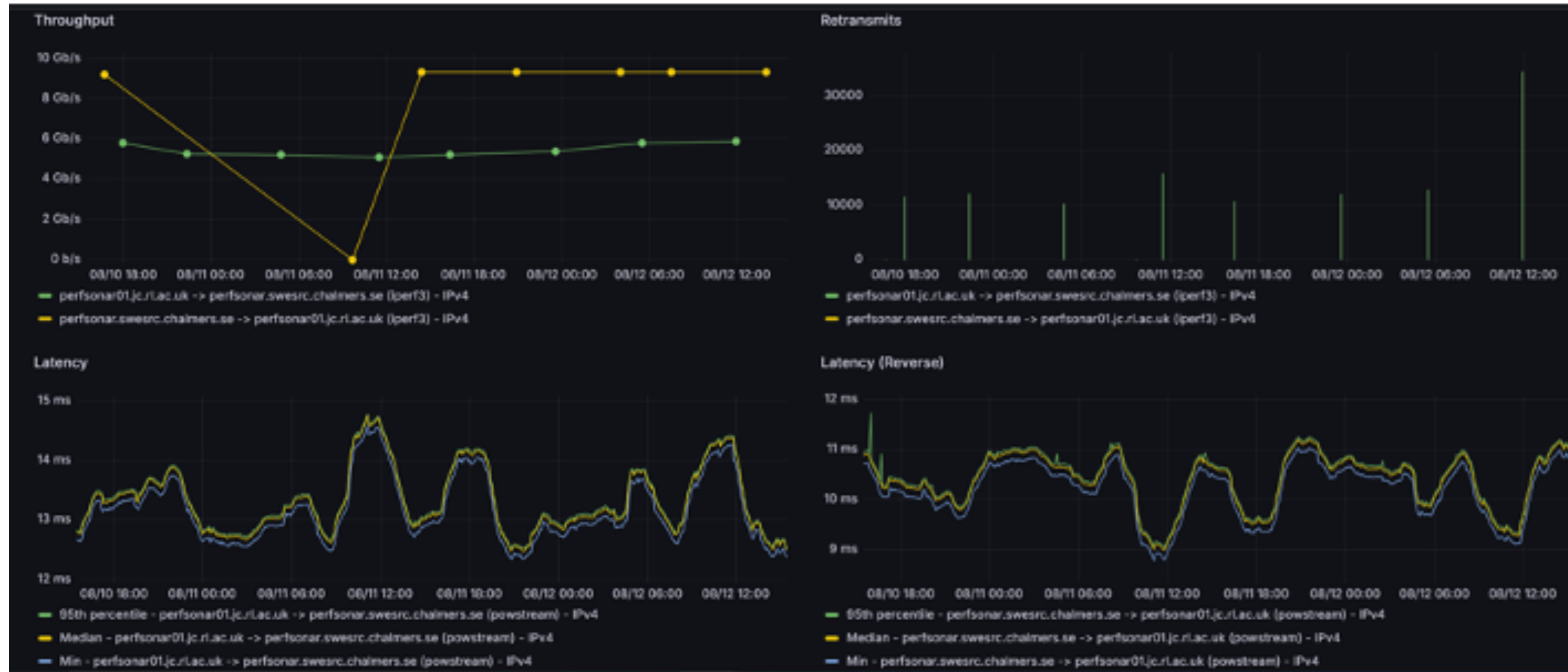


Photos: W. Mayers; [Alces Flight Ltd](#)

Additional cable length required to slide servers forwards on for the rack-rails to change disks

perfSONAR

- Network measurement toolkit establishing federated coverage of end-to-end networking paths
- Continued use of tools developed for / used by HEP communities
- UK (RAL) running the scheduler and providing expertise to other SRCNet Countries in deployment
- Aim: Establish a mesh for all SRCNet v0.1 Nodes
 - And including locations in South Africa and Australia.



Scheduler error identified

DDM & AAI

- UK runs the SKA Indigo IAM instance
 - SRCNet – No user x509 VOMS proxies; all tokens!
- Rucio Instance run by SKAO team on the RAL STFC-Cloud



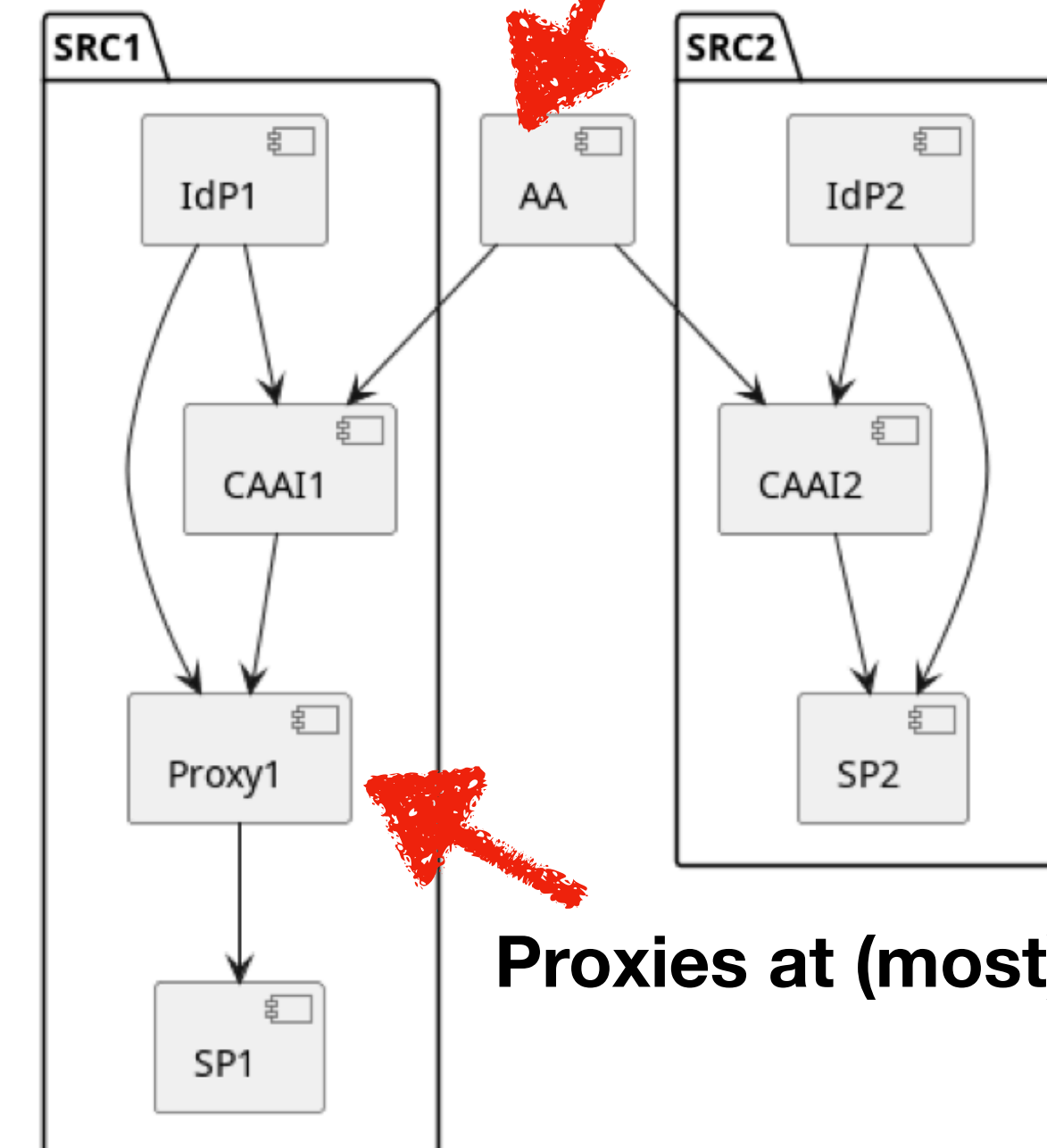
- FTS instance for SKA running at RAL



- See talk by [Rose Cooper](#)

- Work ongoing to specify the final AAI design model

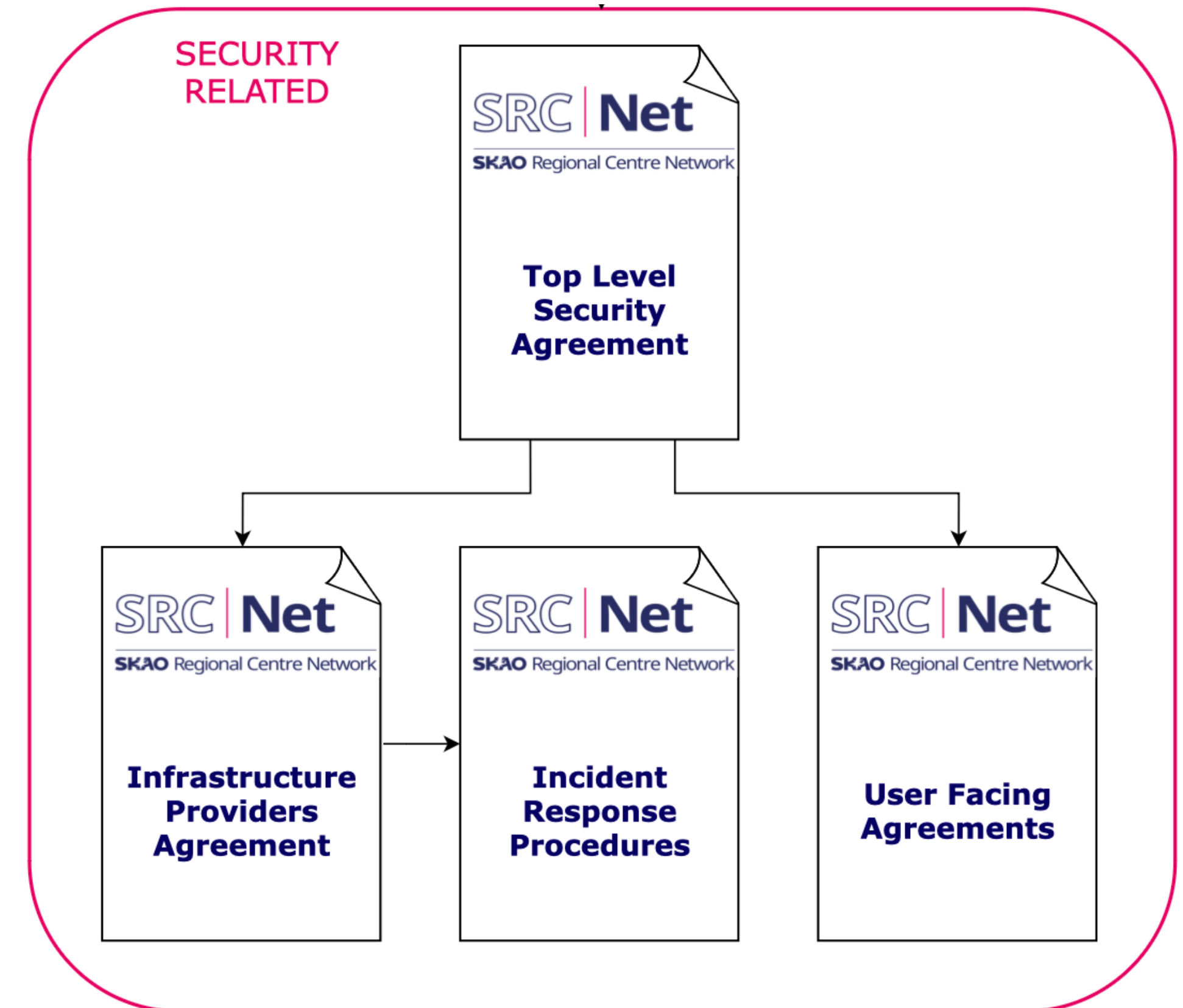
Central Attribute Authority (SKAO)



Proxies at (most) SRCNet nodes

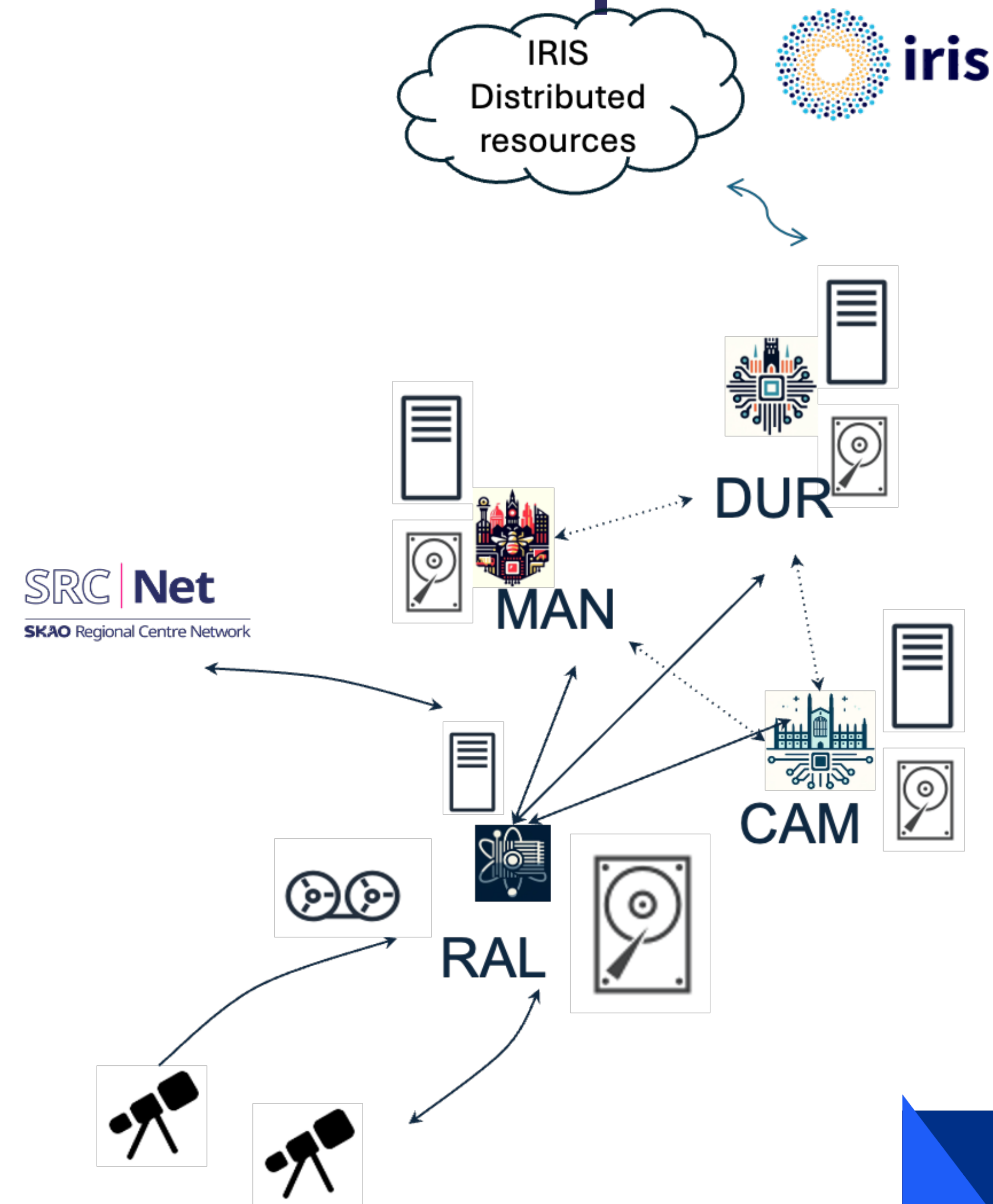
Policies

- As SRCNet pivots from development testbed to prototyping deployments:
 - No “Users” yet, but must develop, socialise and endorse agreements.
 - UK leading the defining and engagement effort on policy
- Core Baseline set of policies for 0.1
 - Built on the AARC Blueprints
 - Familiar territory for anyone involved in WLCG
- Augments existing site policy
 - 2 User Facing Policies
 - AUP and Privacy Policy
 - 1 Service Owner Facing Policies
 - Service Operations
 - 2 Infrastructure Wide Policies
 - Infrastructure Security and Incident Response
- Some Sites with long heritage on federated large-scale projects
 - Some just starting their adventures



UKSRC: proto-architecture and future plans

- 0.1 initial deployment, concentrate at RAL
 - (Utilise teams from across the UK).
- Beyond v0.1;
 - Build up infrastructure at other UK sites (e.g. Durham, Cambridge, Manchester)
 - Develop and test data movement strategies between UK sites
 - Aim to enable use with other HPC resources when possible
 - Federated / common access (e.g IAM)
- RAL to provide TAPE storage (Archival/custodial and high-latency use)
 - The dominant set of bulk disk storage to remain at RAL
- Other sites with smaller (fast) cache / volatile datasets
- Confirm data flow from Telescopes (Ingress into RAL).
- With potentially O(PB) sizes datacubes, efficient movement of data between sites (when required) to be key (N.B) 100Gb/s => ~ 1PB/day
 - Aim to have integration for shared experience across sites
 - i.e. user workflows seamlessly/invisibly shared across the sites
- Rucio (and associated software) to be used for v0.1; data movement to be managed



Summary

- UK preparing for the 0.1 prototype deployment of SRCNet for 0.1
 - Leading contributions in Programme team, software development and benchmarking/profiling and testing
 - Will form a significant contribution to the final SRCNet resource allocation.
- Utilising best of the current approaches from cloud-native solutions,
 - providing science and self-services platforms, and
- Leveraging known solutions from WLCG (e.g. in Data management).
- Building up distributed and federated architecture for SRCNet, and astronomy UK science users
 - Precursor and Pathfinder projects to help development and inform choices
 - Participation in ‘Data Challenge scenarios’
 - Test campaigns for v0.1 in 2025
 - Science engagement and feedback from users critical
- Exciting set of technical and science-based challenges ahead!



www.uksrc.org



Joint mailing list: STFC UK SKA Observatory Science Committee
UKSKA-SCIENCECOMMUNITY@JISCMail.AC.UK



[uk-ska-regional-centre-uksrc](https://www.linkedin.com/company/uk-ska-regional-centre-uksrc)



[@UK SKARC](https://twitter.com/UK_SKARC)



THE UNIVERSITY
of EDINBURGH



Durham
University

University of
Hertfordshire **UH**



The University of Manchester

