

# Enabling SKA science in the global SKA Regional Centre Network

Rosie Bolton  
rosie.bolton@skao.int

Version: 30/09/2024  
Document classification: Unrestricted



# SRCNet Project Introduction

These slides present a brief overview of the SRCNet project current status and upcoming goals.

If you have questions please reach out to  
[Rosie.Bolton@skao.int](mailto:Rosie.Bolton@skao.int)

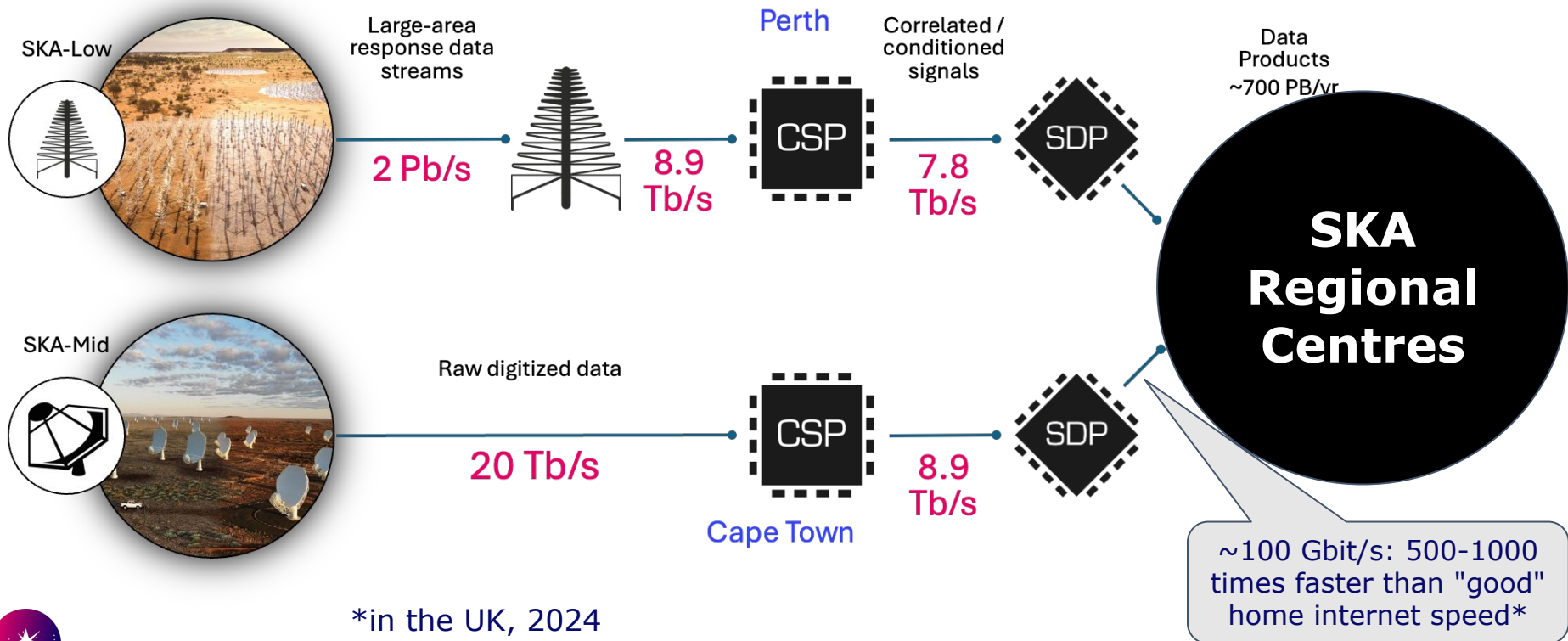
please ask permission before re-using any content.

Information correct at 30 September 2024



# What are the SKA Regional Centres???

Several stages of cool, amazing, cutting edge data processing within the observatory... but **NO USER ACCESS**



# What are the SKA Regional Centres???



## SKA Regional Centres

**Science Gateway**, giving access to **Science enabling tools and applications**

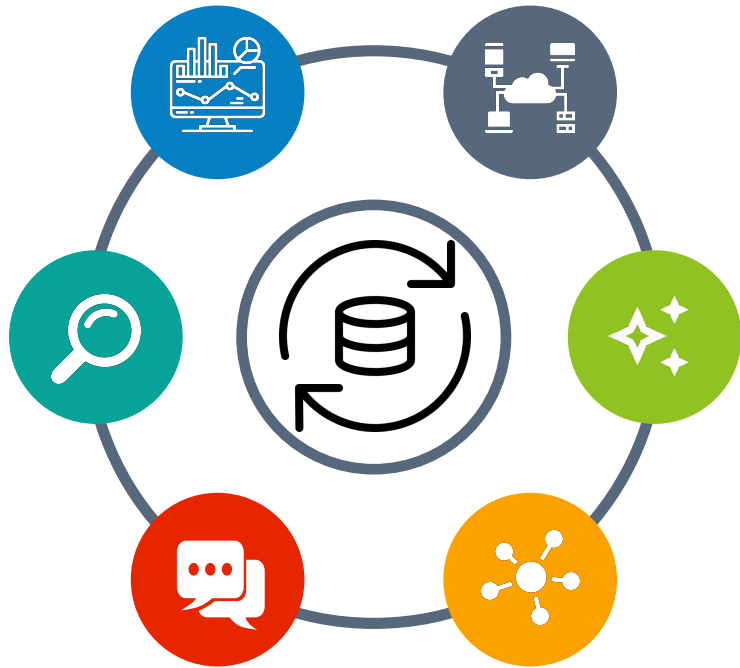
running on federated compute and storage

enabling users to discover data in the **global SKA archive**, develop workflows, perform analyses and collaborate

addresses the "orders of magnitude" data problem



# What are the SKA Regional Centres???



**Science Gateway**, giving access to **Science enabling tools and applications**

running on federated compute and storage

enabling users to discover data in the **global SKA archive**, develop workflows, perform analyses and collaborate

addresses the "orders of magnitude" data problem



# SRC Network Vision

We will develop and deploy a collaborative and federated network of SKA Regional Centres, globally distributed across SKA partner countries, to host the SKA Science Archive.

## **The SRC Network will...**

make data storage, processing and collaboration spaces available, while supporting and training the community, to...

**maximise the scientific productivity and  
impact of the SKA.**



### Science Enabling Applications

Analysis Tools, Notebooks,  
Workflows execution  
Machine Learning, etc

### Distributed Data Processing

Computing capabilities provided  
by the SRCNet to allow data  
processing

### Visualization

Advanced visualizers for SKA  
data and data from other  
observatories

### Interoperability

Heterogeneous SKA data from  
different SRCs and other  
observatories

### Data Management

Dissemination of Data to SRCs  
and Distributed Data Storage

### Data Discovery

Discovery of SKA data from the  
SRCNet, local or remote,  
transparently to the user

### Support to Science Community

Support community on SKA data  
use, SRC services use, Training,  
Project Impact Dissemination



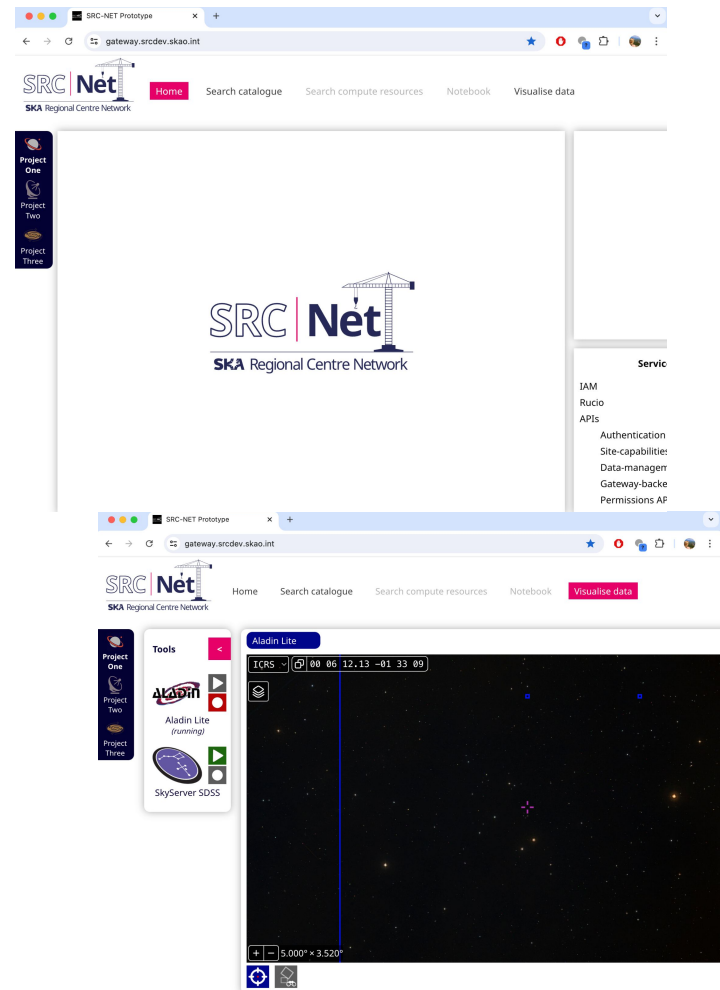
# Intended user perspective

As a user, you'll be a member of one or more groups with an *SRCNet* allocation

- SKAO User (with successful SKAO proposal)
- Archival data user

You will log in via the Gateway

You will be able to select a current project, or discover data sets to add to a project



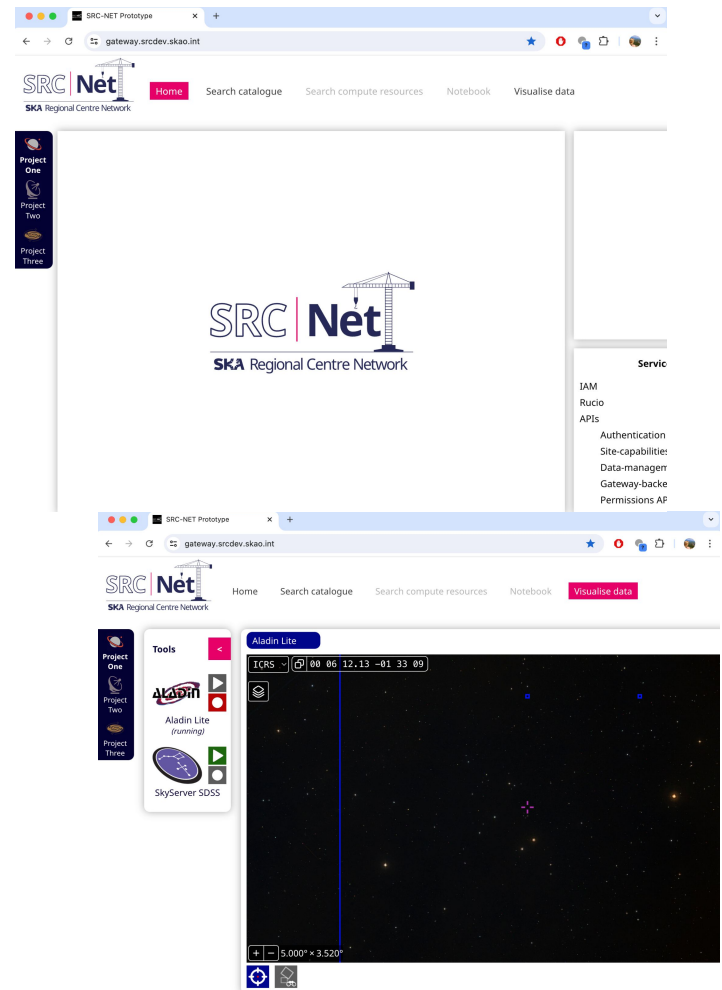


# Intended user perspective

Within a particular project, with some allocated resources, you'll be able to identify services available to support your analysis of the data products you need

Then you'll be able to launch those services and run analyses

You'll be able to save intermediate results locally on the SRC your analysis is running on, and upload final data products ("ADPs") into the archive

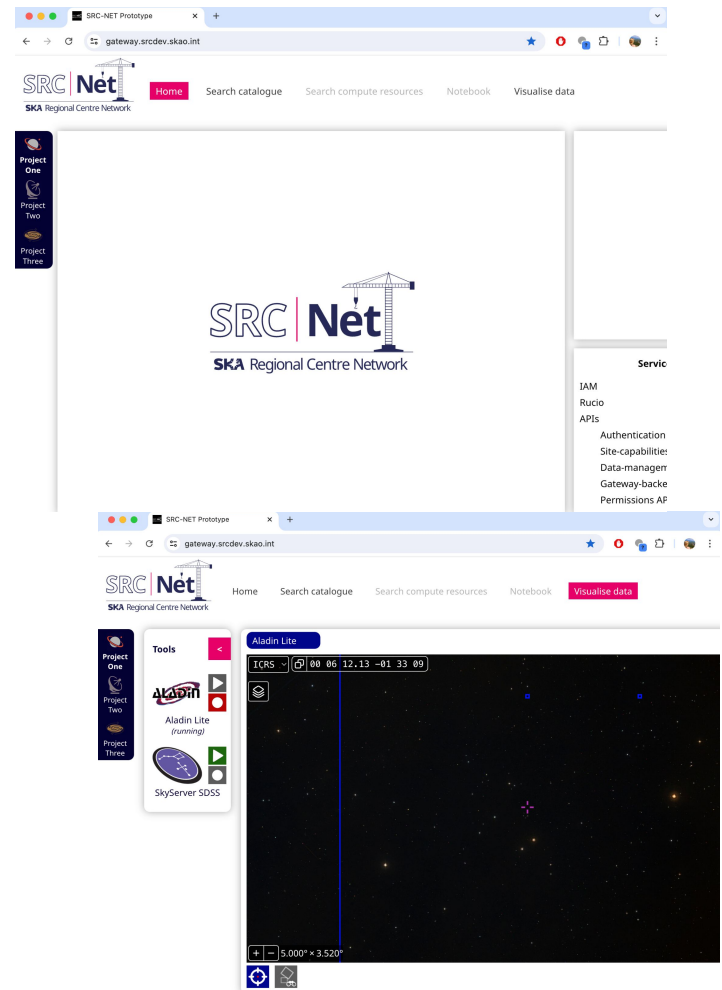


# Intended user perspective

You will be provided with some template workflows to speed up your analysis work

Break away from the **(doomed)** "download and analyse locally" paradigm

Great opportunity to foster reproducibility in workflows - I would love to see user workflows published alongside data in papers by default. Being forced to write software to run on SRCNet will make this final step easier



# Behind the scenes - all should be hidden from user

Several sites (around 10-20) spread globally

**Data replication must be efficient**, and minimised

"Move the user (or code) to the data" where possible



**The bulk SRCNet science archive will be centrally managed**

SRC Operations Group able to trigger replications

At least 2 copies on different SRCs, but also consider storage class (eg. disk faster but more expensive than tape) - data lifecycle support

Auto-recovery if one site fails

**Users shouldn't have to care which site is hosting them -**

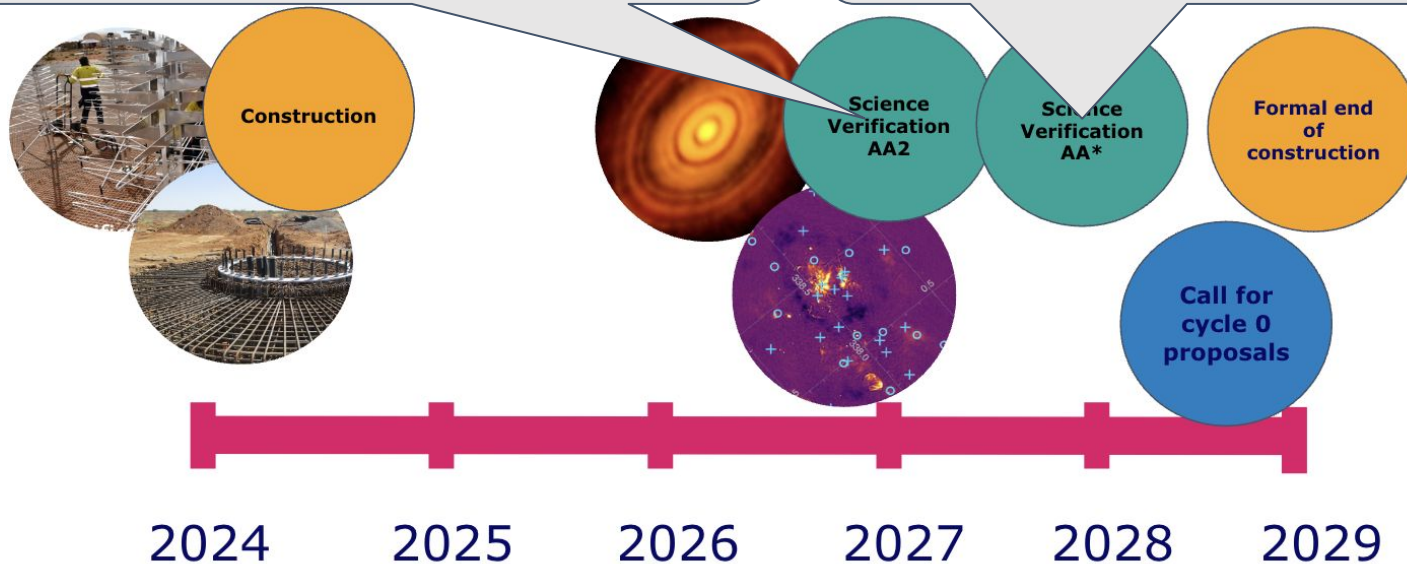
consistent experience across sites



# SKAO Science timeline

2026-2027 SV campaigns produce up to 3.5 PBytes\* of data each SV week

2027-2028 SV campaigns produce up to 14 PBytes\* of data each SV week



Call for cycle 0 proposals

Construction

Science Verification AA2

Science Verification AA\*

Formal end of construction

2024

2025

2026

2027

2028

2029

Commissioning

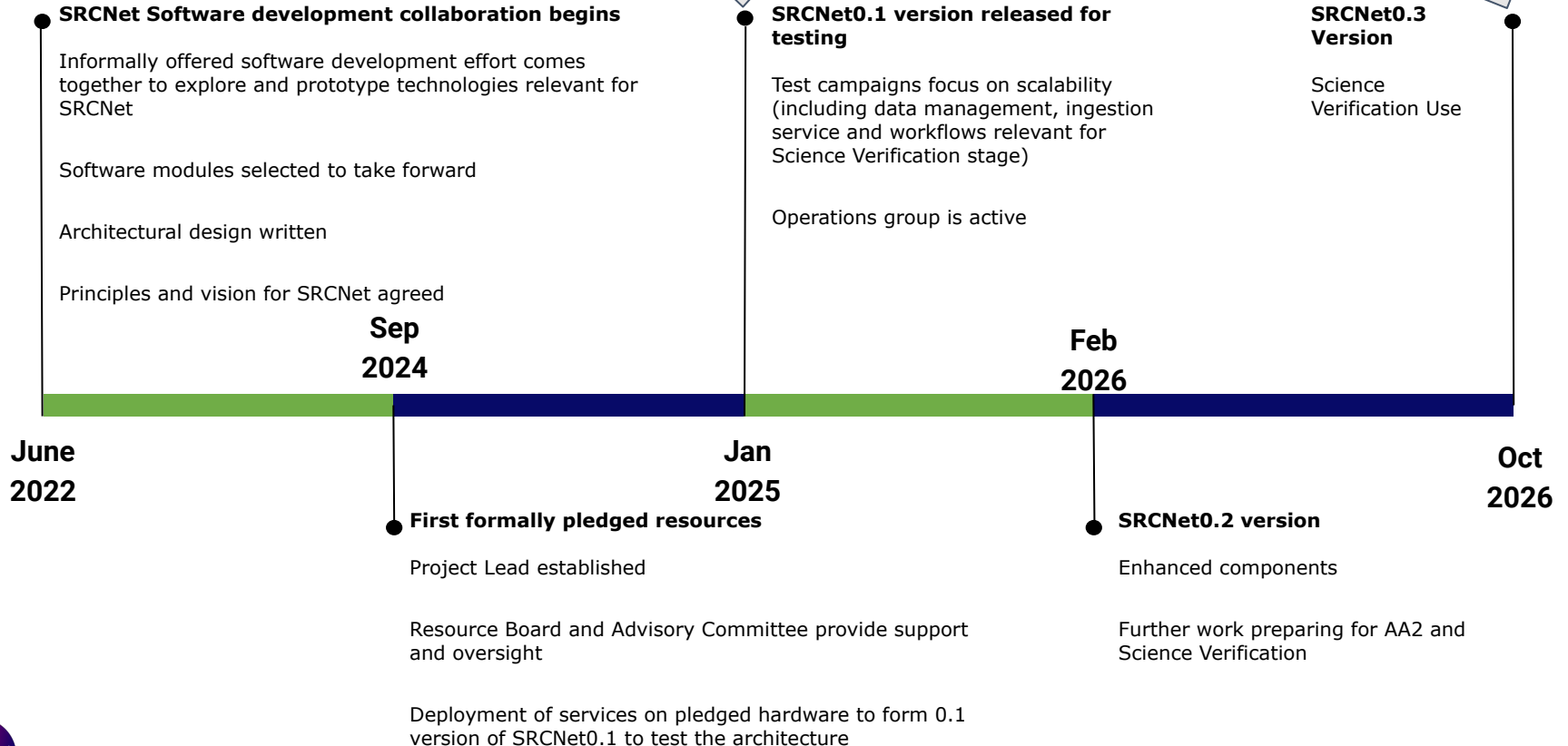
\*CURRENT ESTIMATES, subject to change



# SRCNet timeline

Focus for activity for next 6 months

Real scientists start to use SRCNet



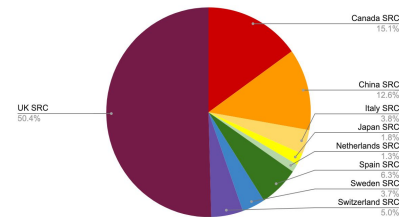
# SRCNet0.1 planned sites

9 sites will contribute  
compute and storage  
resources to SRCNet0.1

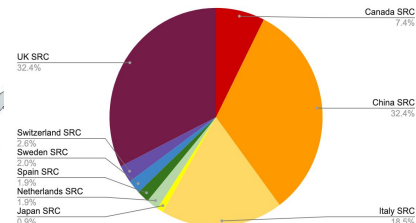
8 PBytes total  
storage  
0.5 PFLOPS Compute

- Canada UVic 1-4 PB
- China Shanghai Observatory 1 PB
- Italy, INAF IRA, 0.3PB disk, 1.2PB tape, 10 gbps
- Japan, Tokyo NAOJ, 0.14PB
- Netherlands, SURF, 0.1PB
- Spain, IAA Granada, 0.5PB
- Sweden, Gothenburg, 0.3PB
- Switzerland, CSCS Lugano, 0.4PB
- UK, STFC RAL, 4.0PB

Storage fraction for SRCNet0.1



Compute fraction for SRCNet0.1



# The SRCNet Project

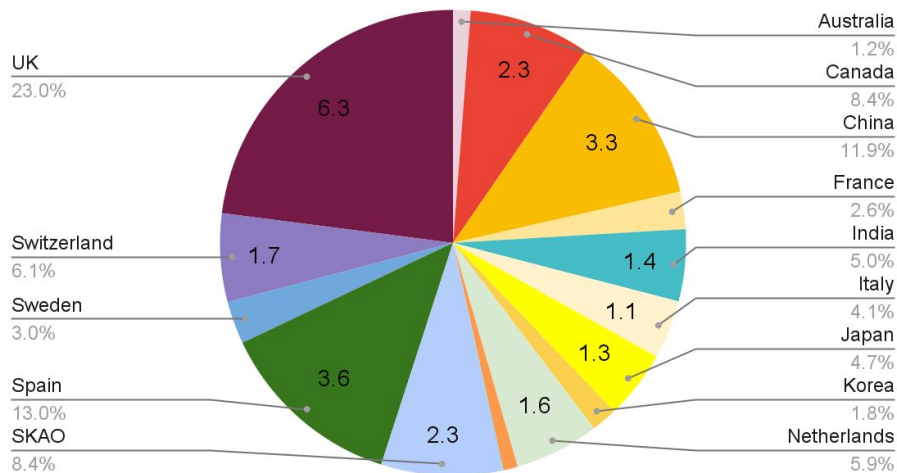
The SRCNet Project aims to deliver a working SRC Network in time for formal start of SKAO Operations, and for intermediate science verification stages

- End date July 2028

This is distinct\* from the long term "steady state" functioning of the fully-formed SKA Regional Centre Network

(\*This distinction is important because governance structures are expected to be reassessed for long term functioning)

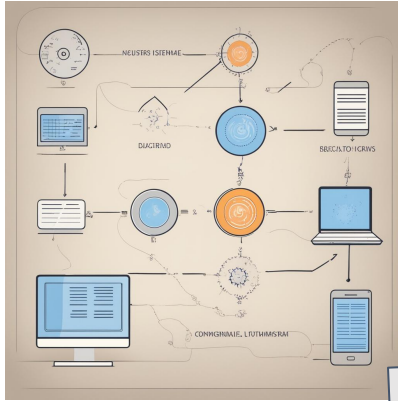
Development FTE average for past year



**Currently about 40 person-worth of effort from 13 countries plus SKAO**



# SRCNet composition



Software & services



SW Development and Service operations



Hardware



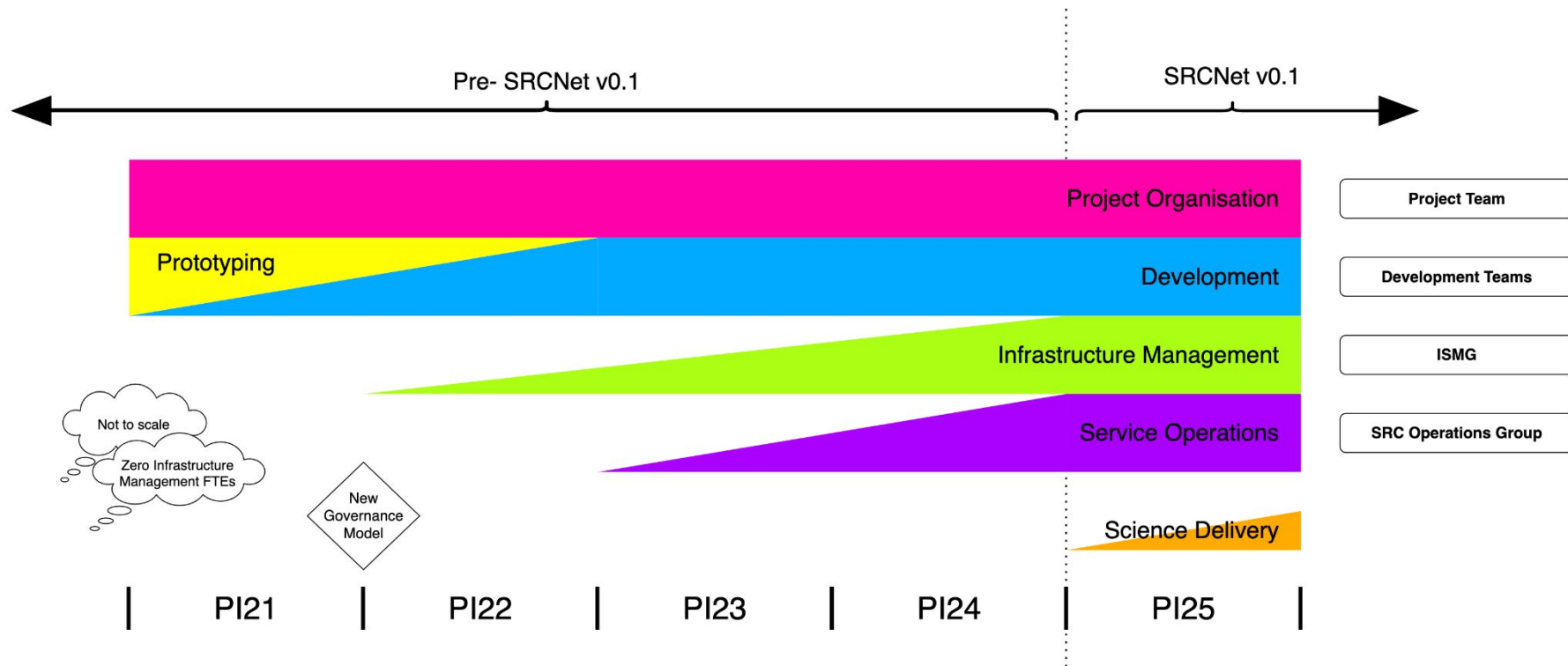
Science Delivery

Science Users

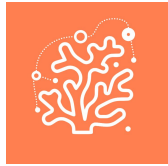




# Value Stream Development



# The current SRCNet teams



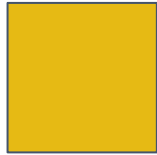
Coral



Purple



Lavender



Gold



Orange



Red



Magenta



Teal



Tangerine



Chocolate

DAAC

Indigo

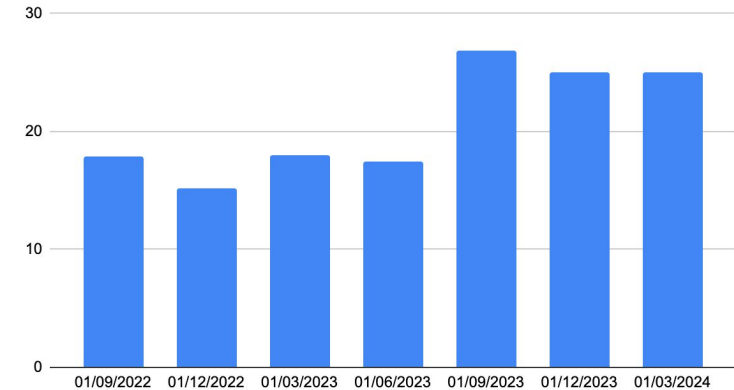
Since June 2022 we have been working as a team-of-teams

Engagement from across most SKA countries

Currently 40+ persons-worth of effort

70+ contributors

FTE in SRCNet SW development effort by date



# SRCNet timeline\*

Focus for activity for next 3 months

Real scientists start to use SRCNet

## SRCNet Software development collaboration begins

● Informally offered software development effort comes together to explore and prototype technologies relevant for SRCNet

Software modules selected to take forward

Architectural design written

Principles and vision for SRCNet agreed

Sep  
2024

## SRCNet0.1 version released for testing

Test campaigns focus on scalability (including data management, ingestion service and workflows relevant for Science Verification stage)

Operations group is active

Feb  
2026

## SRCNet0.3 Version

Science Verification Use

June  
2022

PI24

SRCNet 0.1 phase

SRCNet 0.2 phase

Jan  
2025

Oct  
2026

## ● First formally pledged resources

Project Lead established

Resource Board and Advisory Committee provide support and oversight

Deployment of services on pledged hardware to form 0.1 version of SRCNet0.1 to test the architecture

## ● SRCNet0.2 version

Enhanced components

Further work preparing for AA2 and Science Verification

# **SRCNet 0.1 is our first big milestone!!**

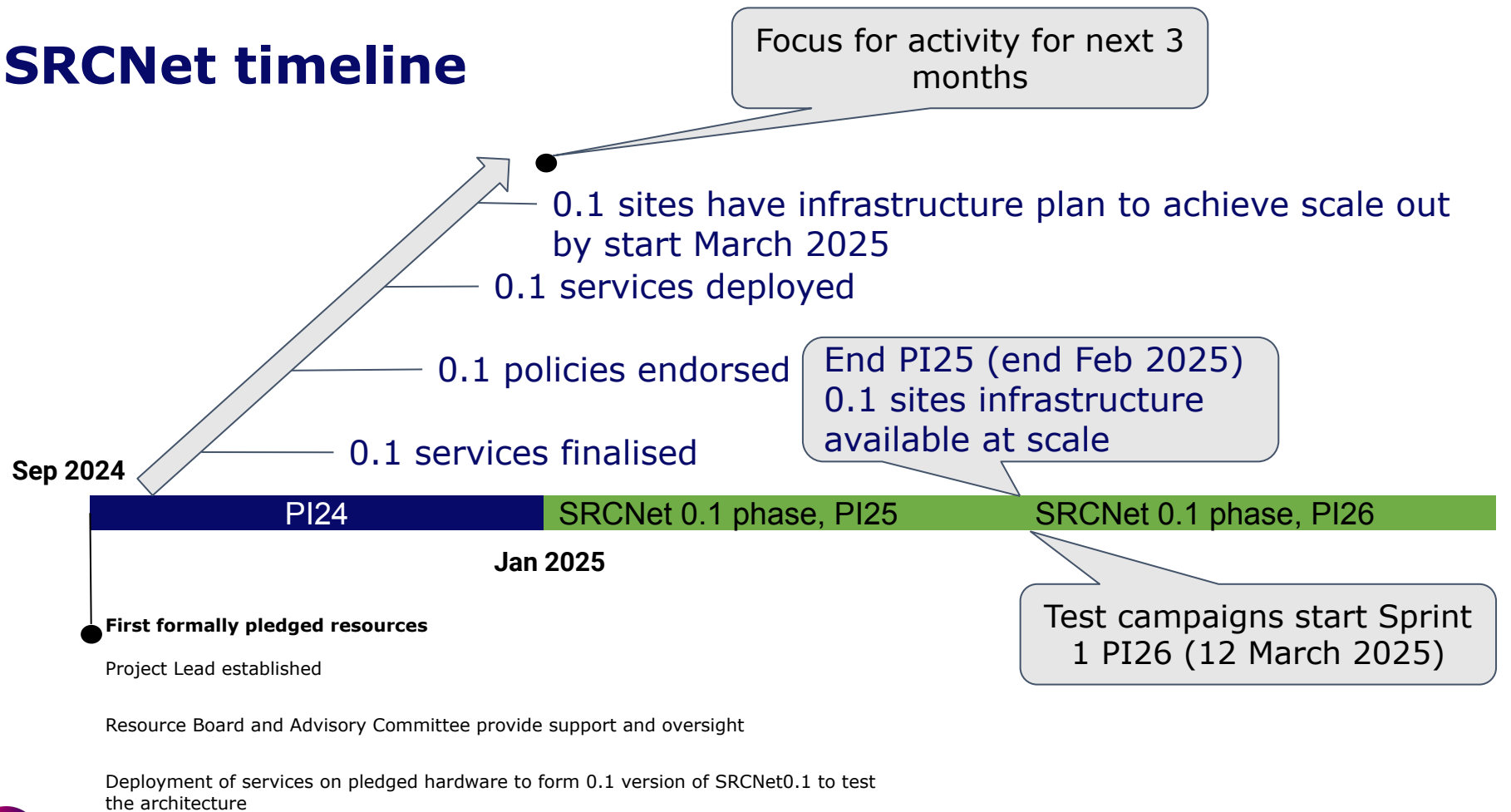
At least 4 sites running full set of compulsory local services; global services also running to support this

First test of full SRCNet architecture

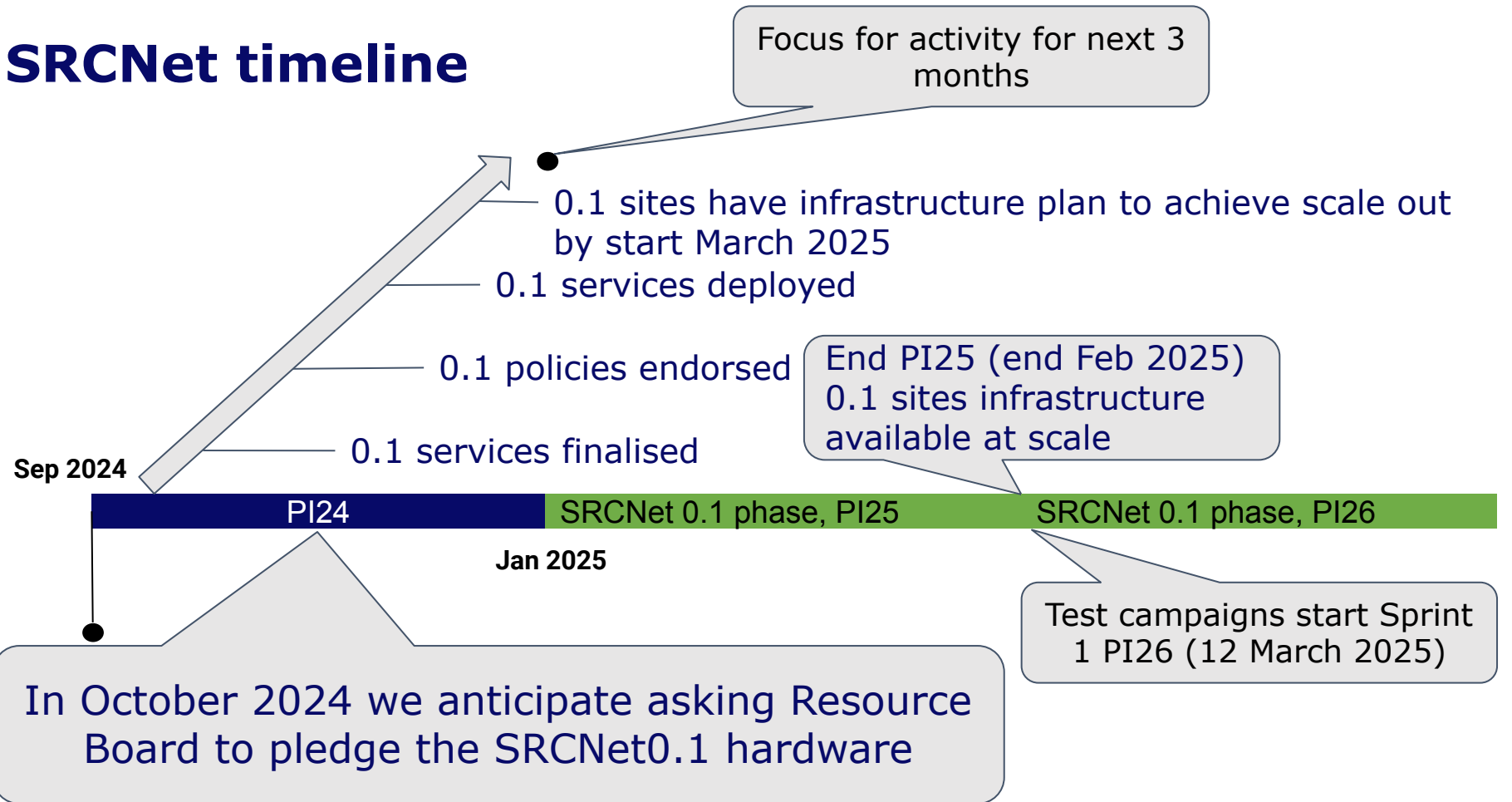
Due end of PI24; ie. 20th November 2024 (final work day of sprint 5)



# SRCNet timeline



# SRCNet timeline

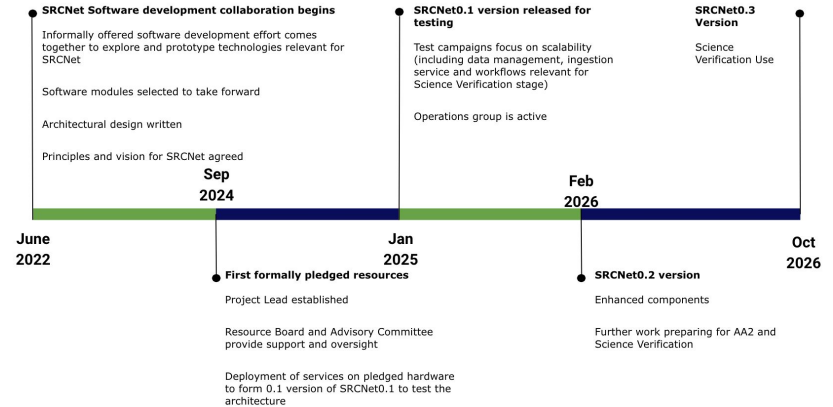


# SRCNet0.1

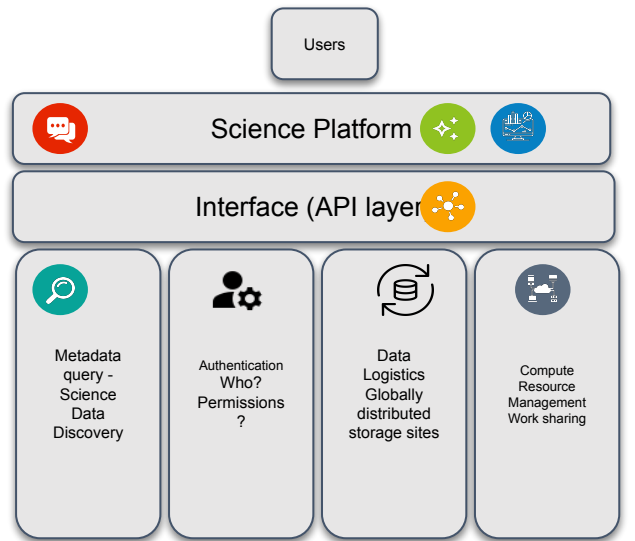
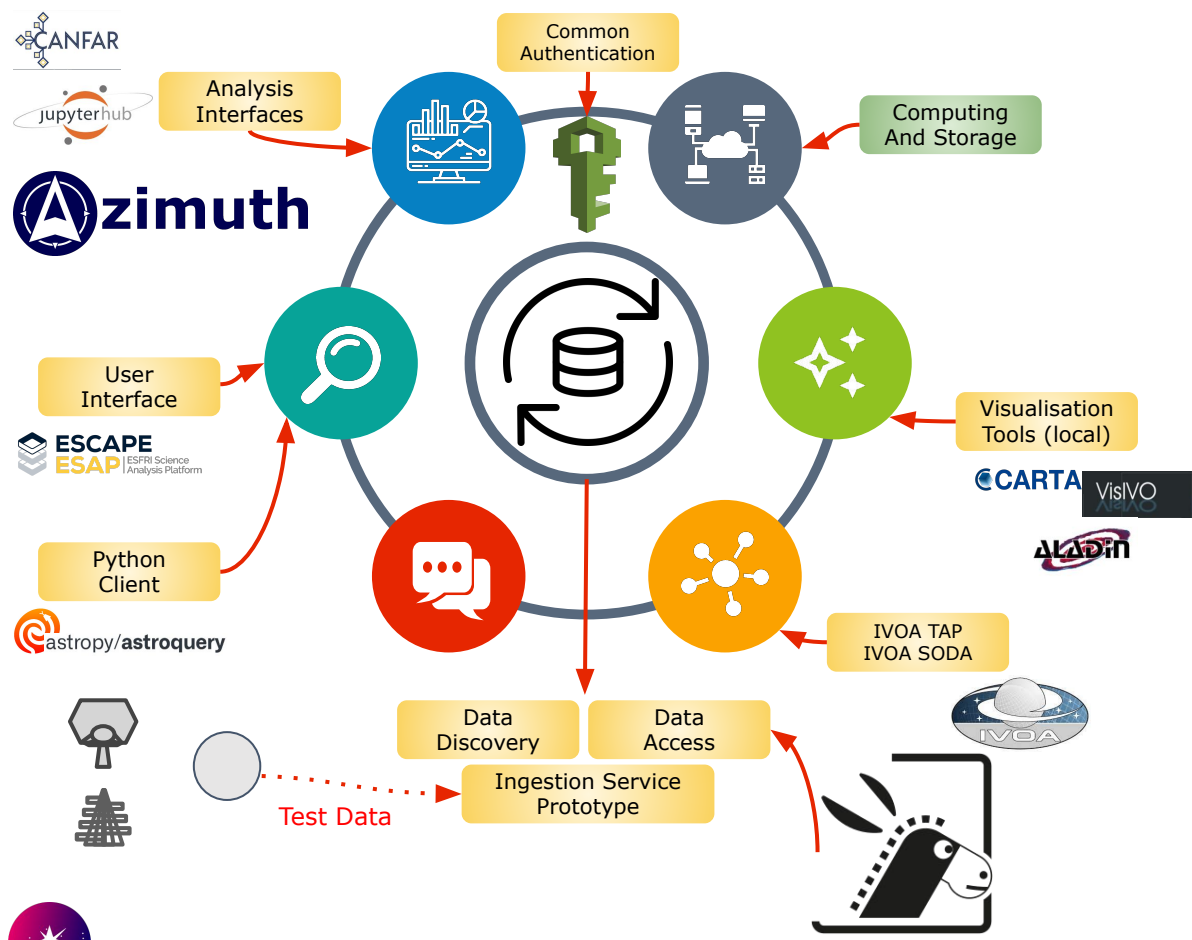
This is an "engineering" version

- **Built to show the architecture and test how it works**
- **Internal only** - no user-facing activities
- Exclusive storage to use in testing
- Compute to use during testing campaigns (may be backfilled when idle)
- Learn how to deploy and operate the services
- Set up of the SRC Operations Group, with limited scope

Focus for activity for next 6 months



# Basic Functionality Covered by v0.1

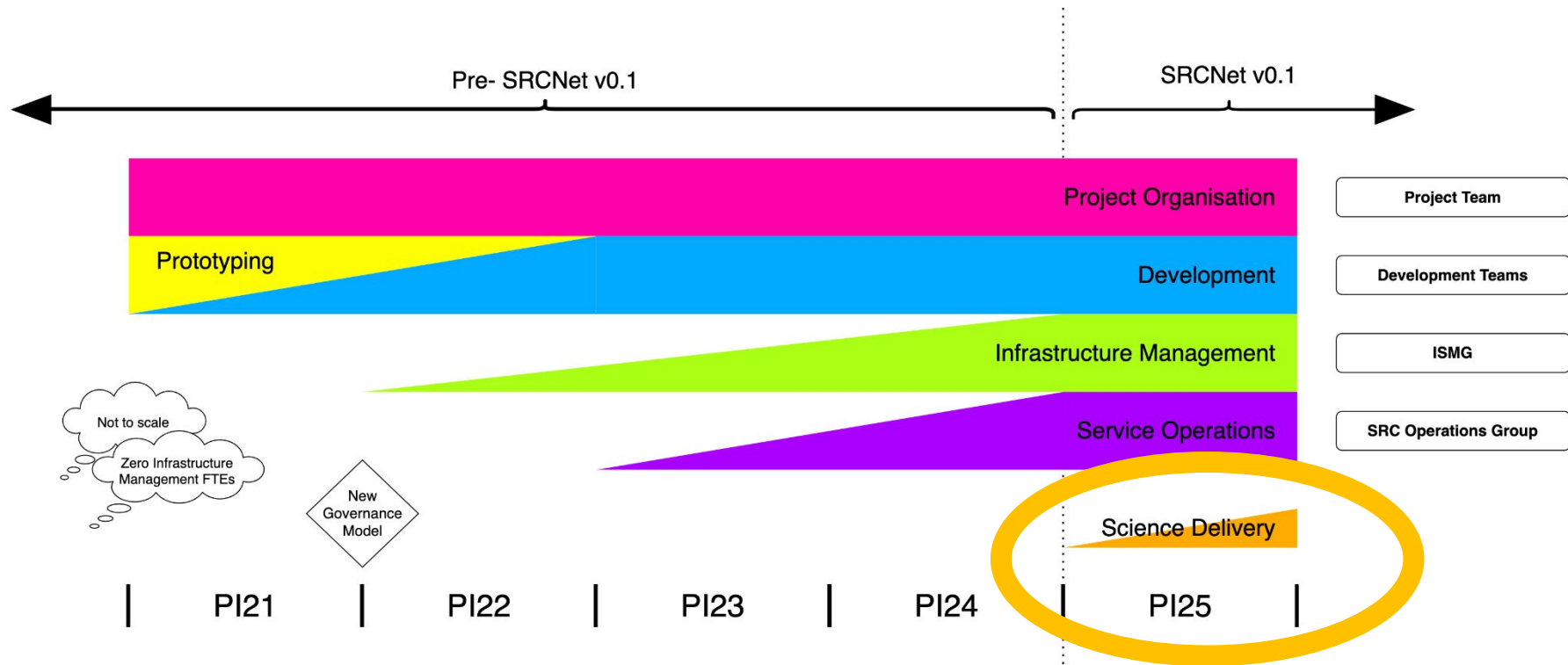


- Common Authentication: IAM
- Visualisation Tools (local)
- IVOA Protocols: TAP, SODA
- Data Discovery and Access from (Rucio)Data Lake
- Ingestion Service Prototype
- Python Client: Astroquery Module
- User Interface: Gateway  
<https://gateway.srcdev.skao.int/>
- Analysis Interfaces: JupyterHub (compulsory); CANFAR Science Platform, Azimuth (UK)





# Value Stream Development - Science Delivery



# Near term FTE resource needs for the SRCNet Project

PI	PI23	PI24	PI25
Start Date	12 June 2024	11 September 2024	11 December 2024
Value Stream	FTEs		
Organisation	6	6	6
Development	34	34	37
Service Operations	0	2	4
Science Delivery	0	0	13
Infrastructure Management	0	0	0
<b>TOTAL</b>	<b>40</b>	<b>42</b>	<b>60</b>

Stable resourcing for PI23 and PI24

PI25: Jump in the level of effort needed if we are to meet plans in the SRCNet Top Level Roadmap 40 to 60 FTE

current PI numbers

PI24 planning 2-6 September!



# Current scientific work in SRCNet

We are building a suite of example workflows to support testing of our sites and to develop benchmarking tools, including simulation software to make realistic data sets that could enable end-to-end SRCNet tests

These help demonstrate relevant analyses to our developer community and are now runnable as part of a testing suite with dashboard

Soon (by December) we will seek to strengthen our science work with additional FTEs and specific roles to help community engagement and translation of use cases into technical requirements



End

*We recognise and acknowledge the  
Indigenous peoples and cultures that have  
traditionally lived on the lands on which  
our facilities are located.*

**SKAO**

[www.skao.int](http://www.skao.int)