SKAO REGIONAL CENTRE NETWORK

Enabling SKA science in the global SKA Regional Centre Network

Rosie Bolton
rosie.bolton@skao.int

Interim SRCNet Project Lead

30/09/2024

SRCNet Project Introduction

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

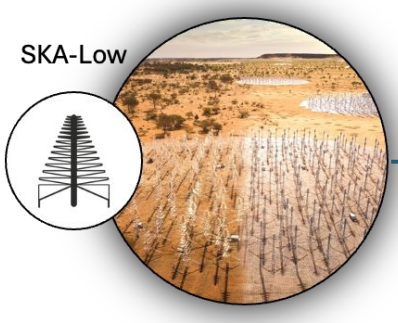
If you have questions please reach out to Rosie.Bolton@skao.int

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**



SKA-Low

Large-area response data streams

2 Pb/s

Perth

8.9 Tb/s

CSP

Correlated / conditioned signals

7.8 Tb/s

SDP

Data Products ~700 PB/yr

SKA-Mid

Raw digitized data

20 Tb/s

CSP

Cape Town

8.9 Tb/s

SDP

SKA Regional Centres

~100 Gbit/s: 500-1000 times faster than "good" home internet speed*

*in the UK, 2024

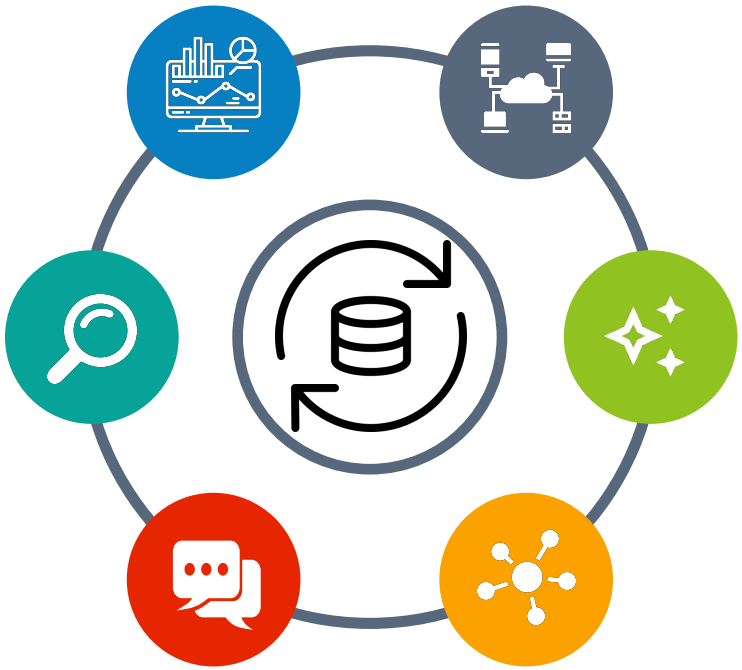
What are the SKA Regional Centres???

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

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

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

Visualization
Advanced visualizers for SKA data and data from other observatories

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

Interoperability
Heterogeneous SKA data from different SRCs and other observatories

Data Management
Dissemination of Data to SRCs and Distributed Data Storage

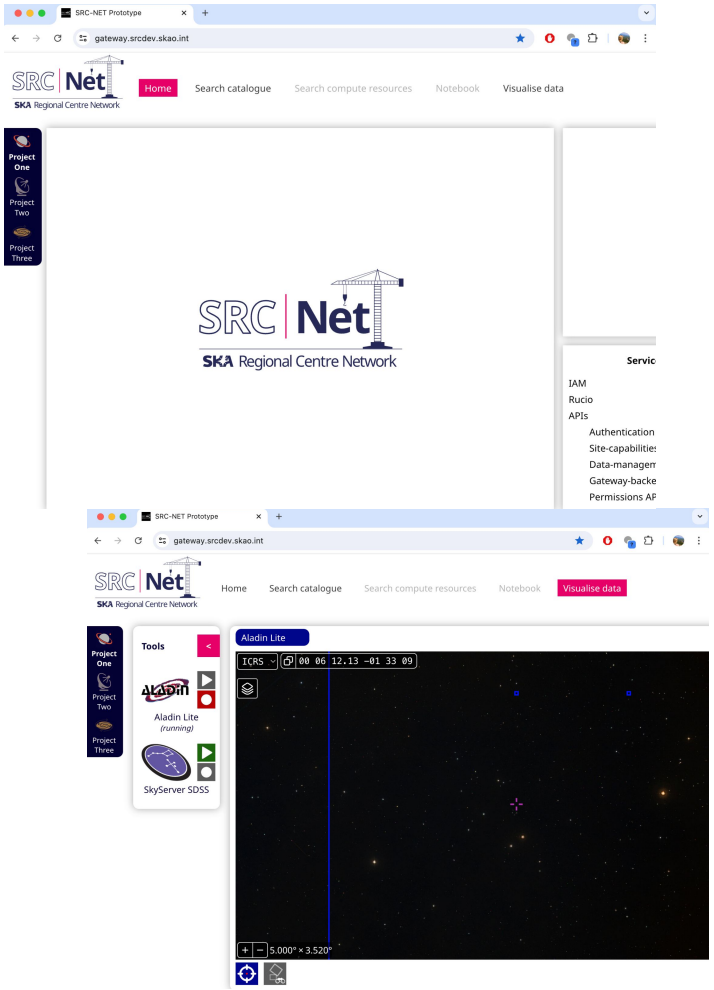
Intended user perspective

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

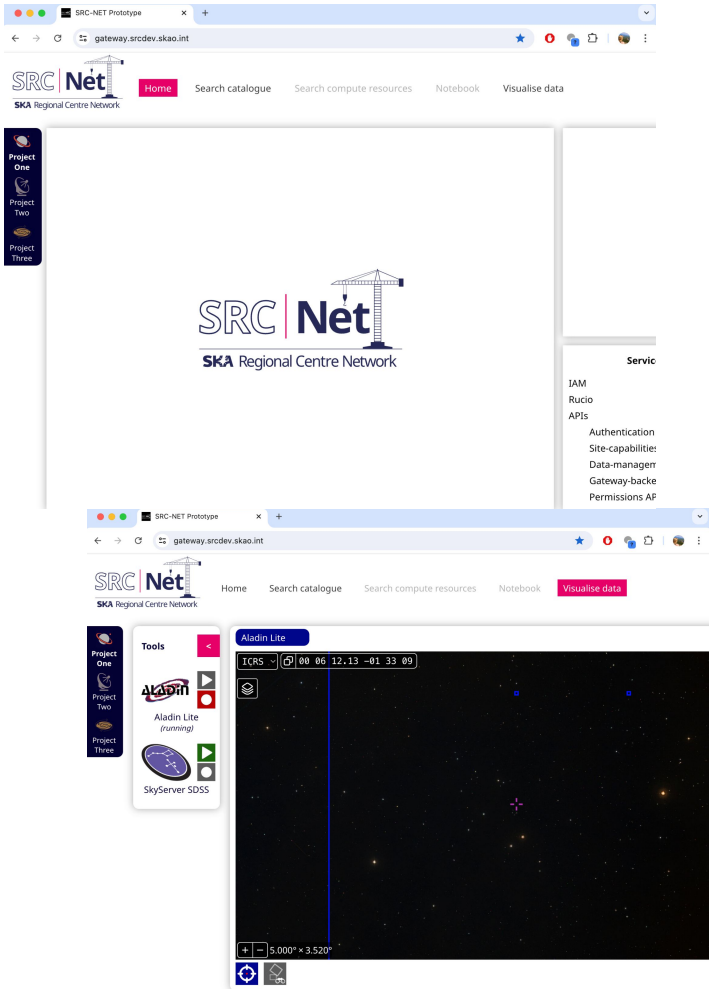


Intended user perspective

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

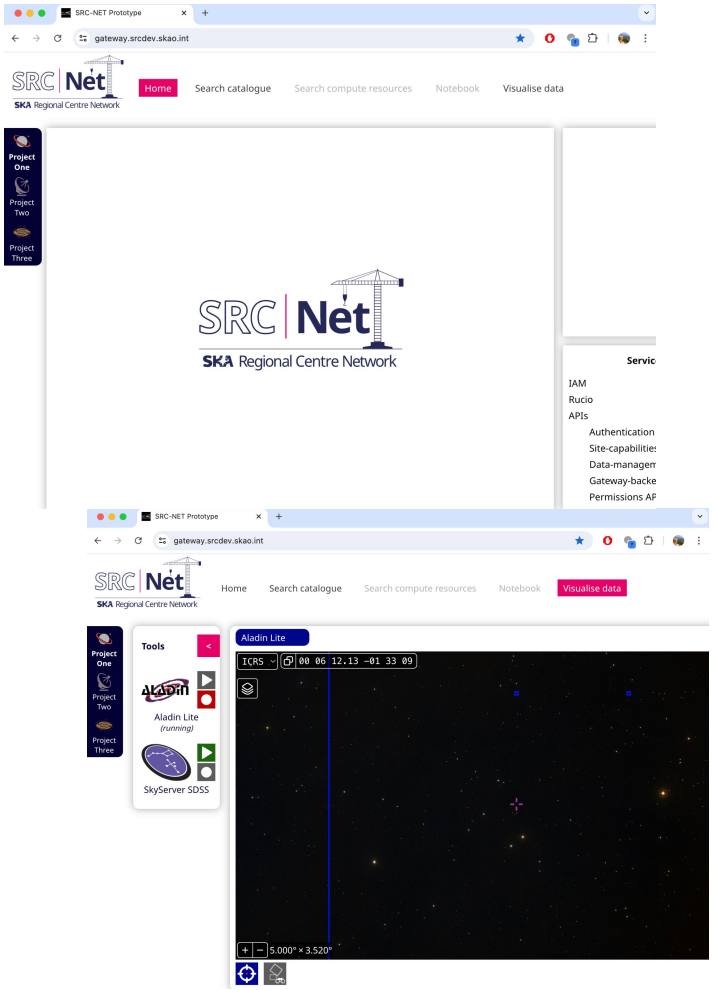


Intended user perspective

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) - date lifecycle support

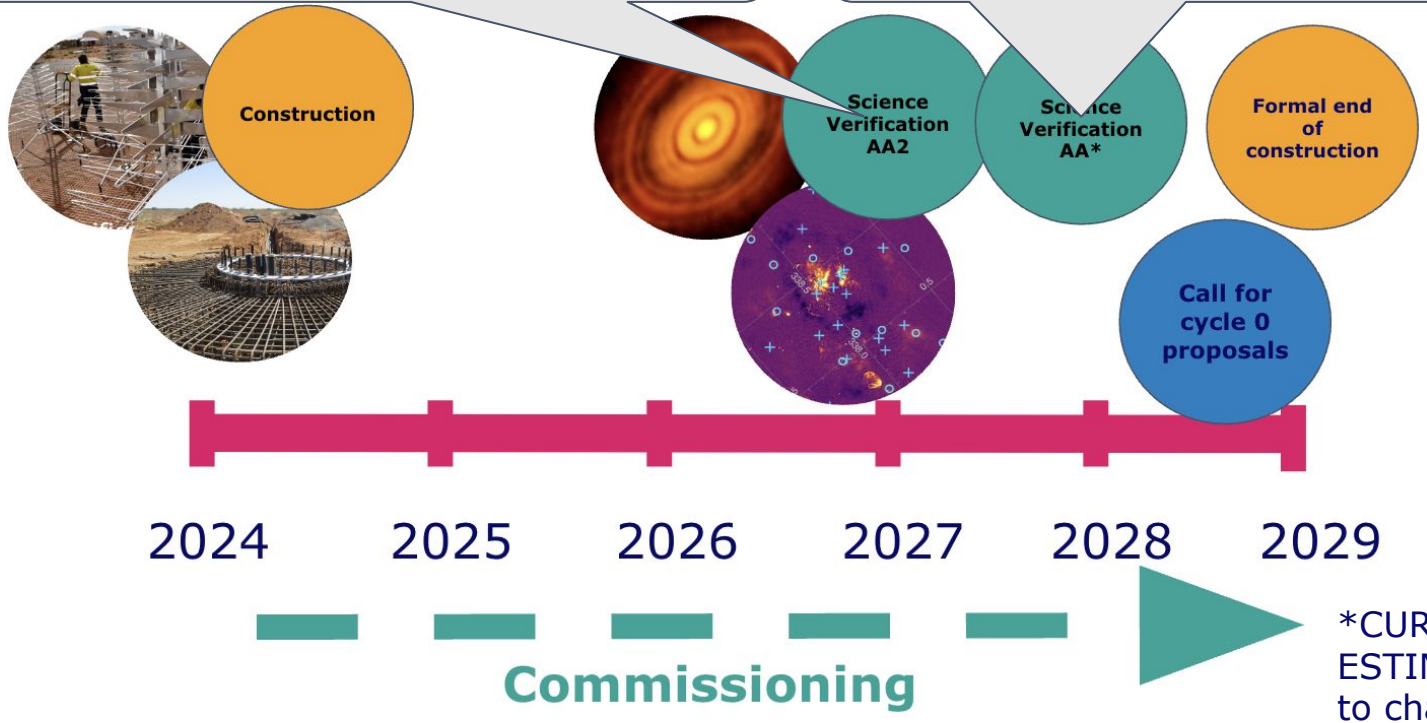
Auto-recovery if one site fails

Users shouldn't have to care which site is hosting them - consistent experience across sites

SKAO Science timeline

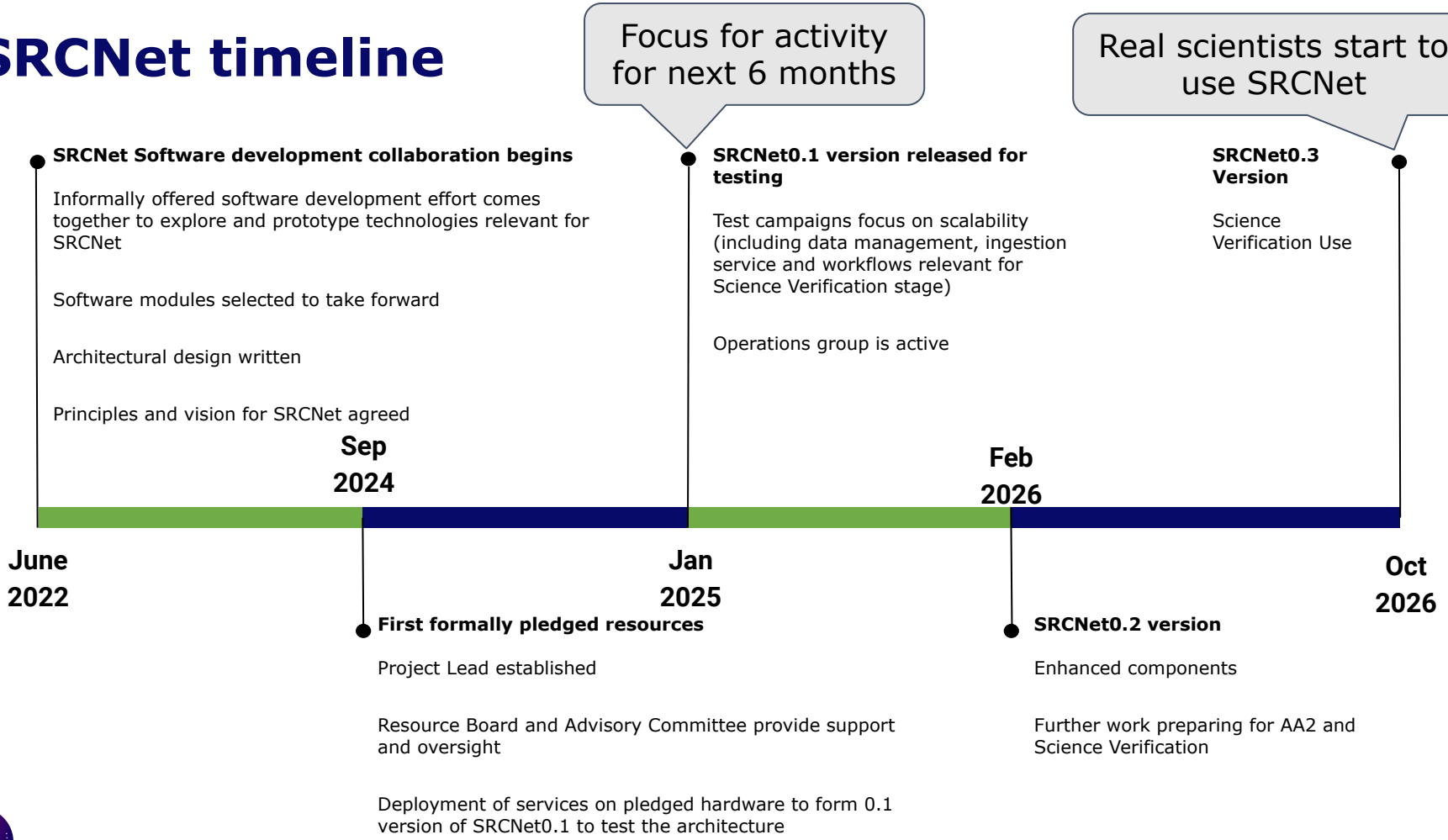
2026-2027 SV campaigns produce up to 3.5 PBytes* of data each SV week

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



*CURRENT ESTIMATES, subject to change

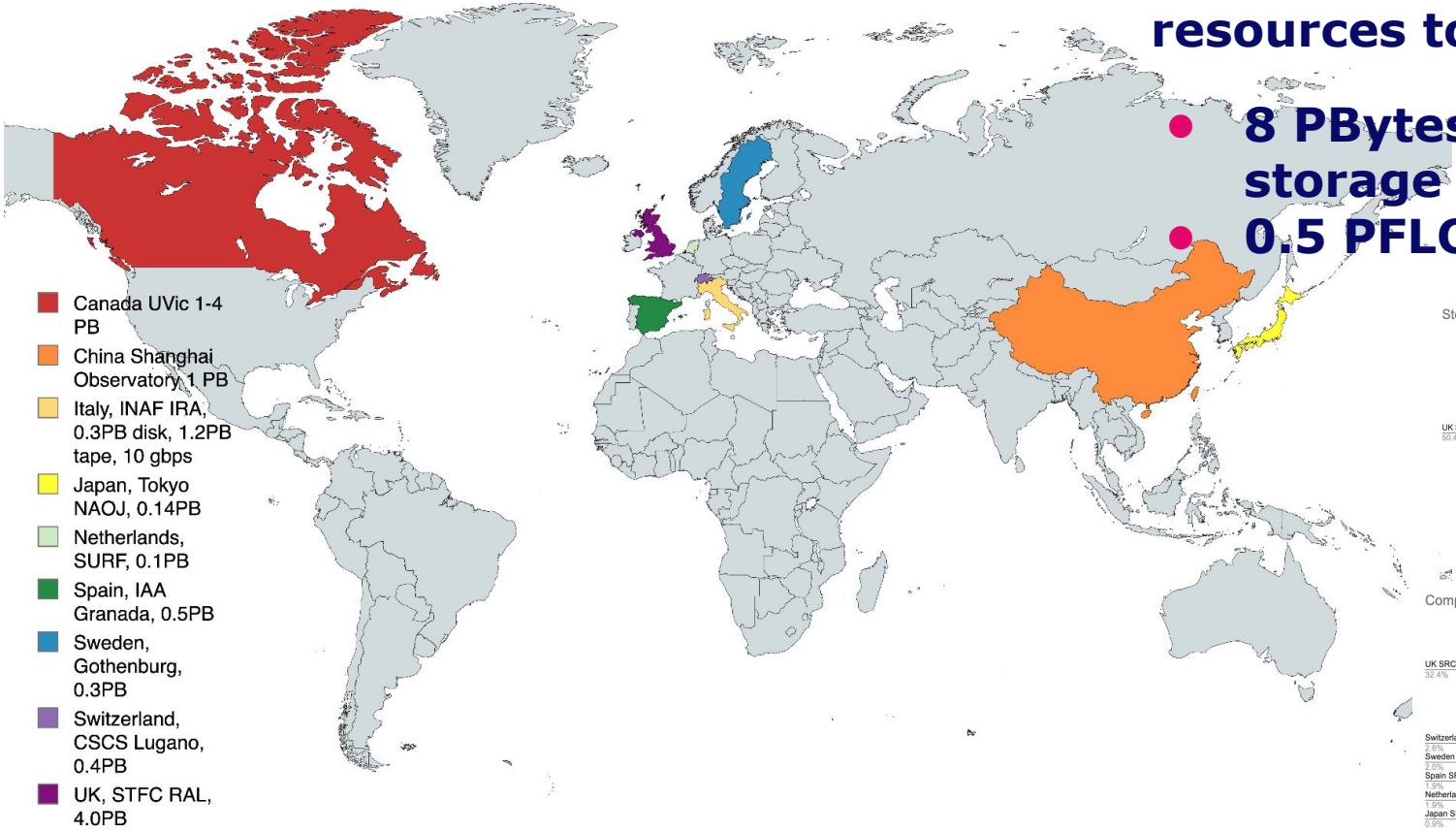
SRCNet timeline

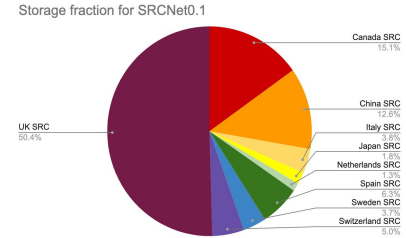


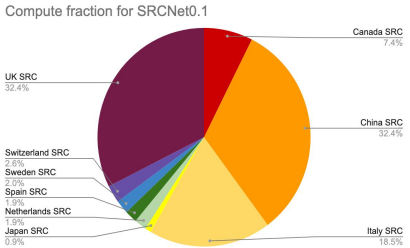
SRCNet0.1 planned sites

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

- 8 PBytes total storage
- 0.5 PFLOPS Compute





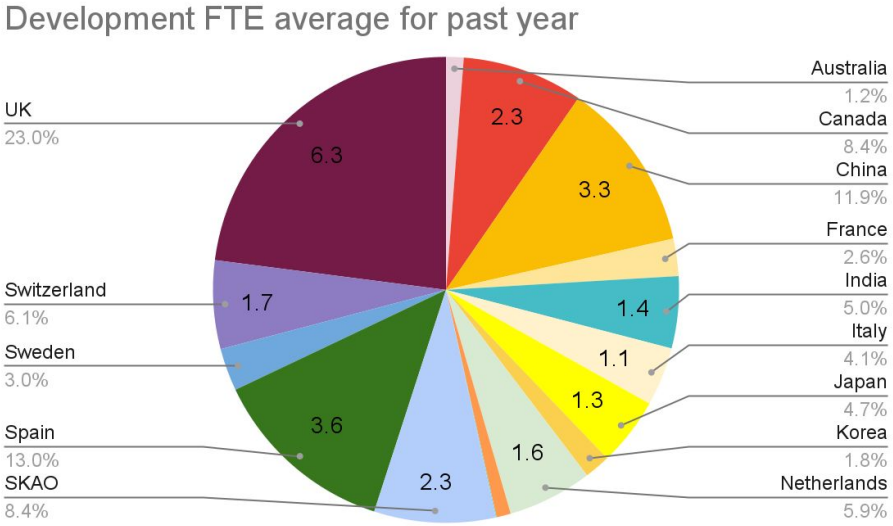


The SRCNet Project

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

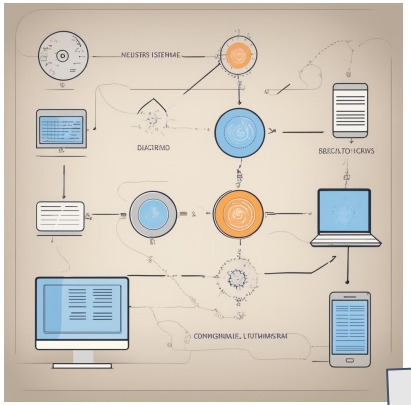
This is distinct* from the long term "steady state" functioning of the fully-formed SKA Regional Centre Network



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

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

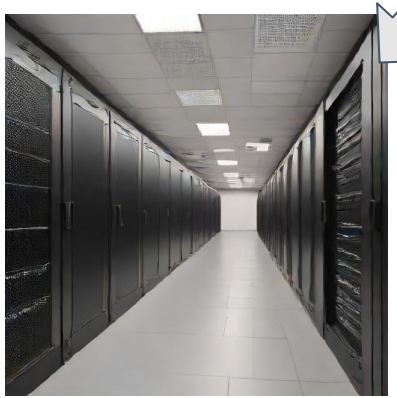
SRCNet composition



Software & services

SW Development and Service operations

SRC Net!



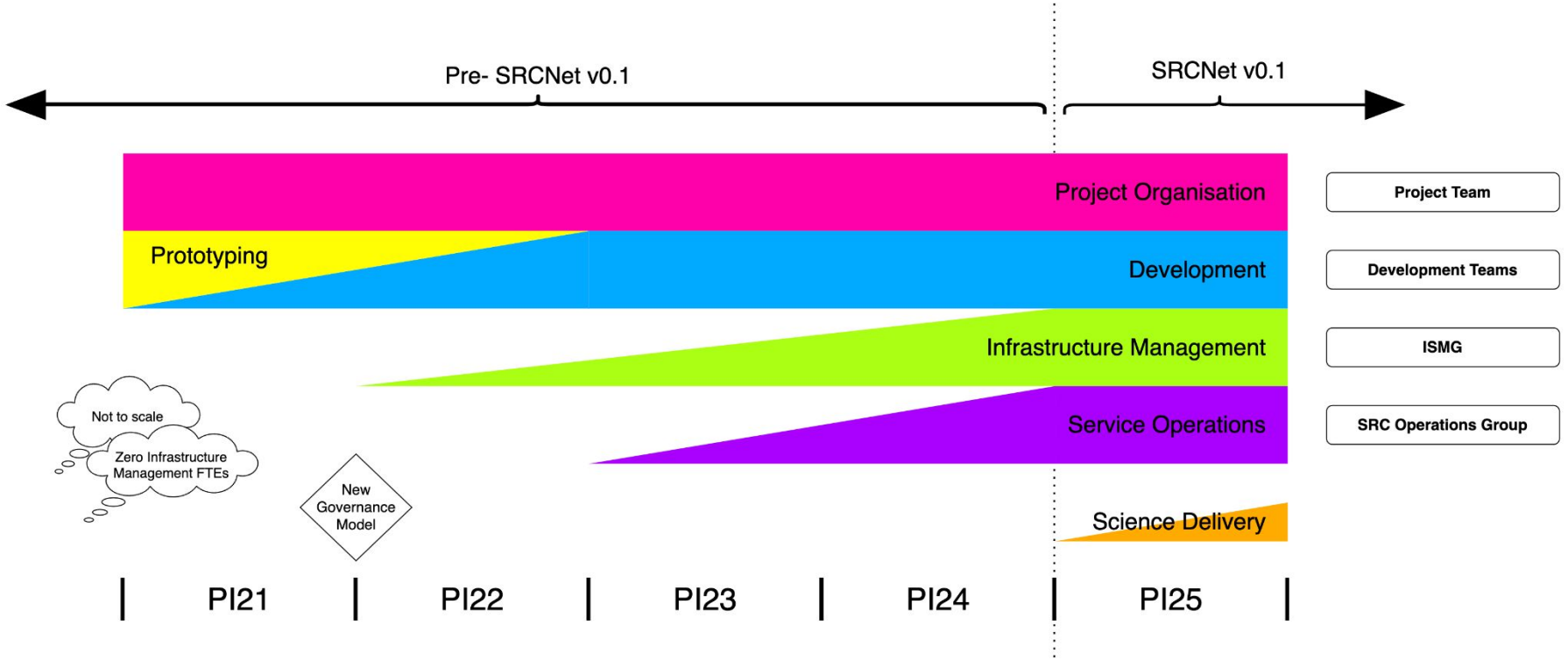
Hardware

Science Delivery

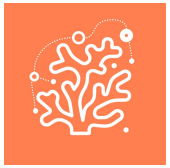
Science Users

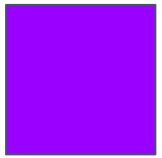
DeepAI images

Value Stream Development

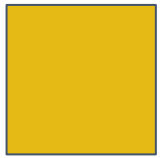


The *current SRCNet* teams

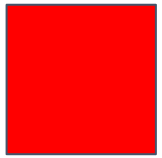

Coral


Purple


Lavender


Gold


Orange


Red


Magenta


Teal


Tangerine


Chocolate

DAAC

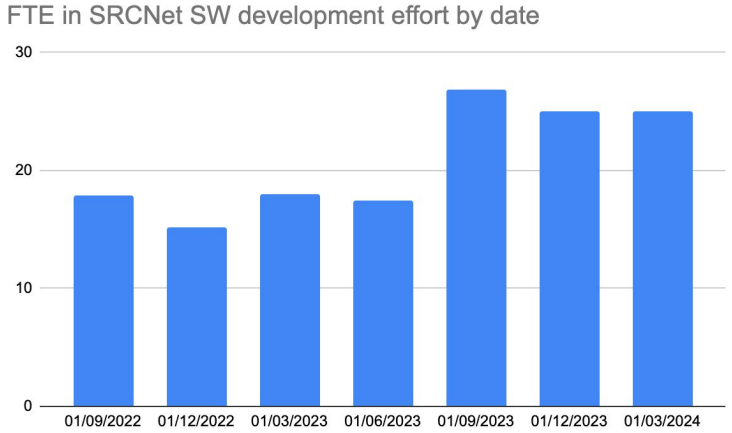
Indigo

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

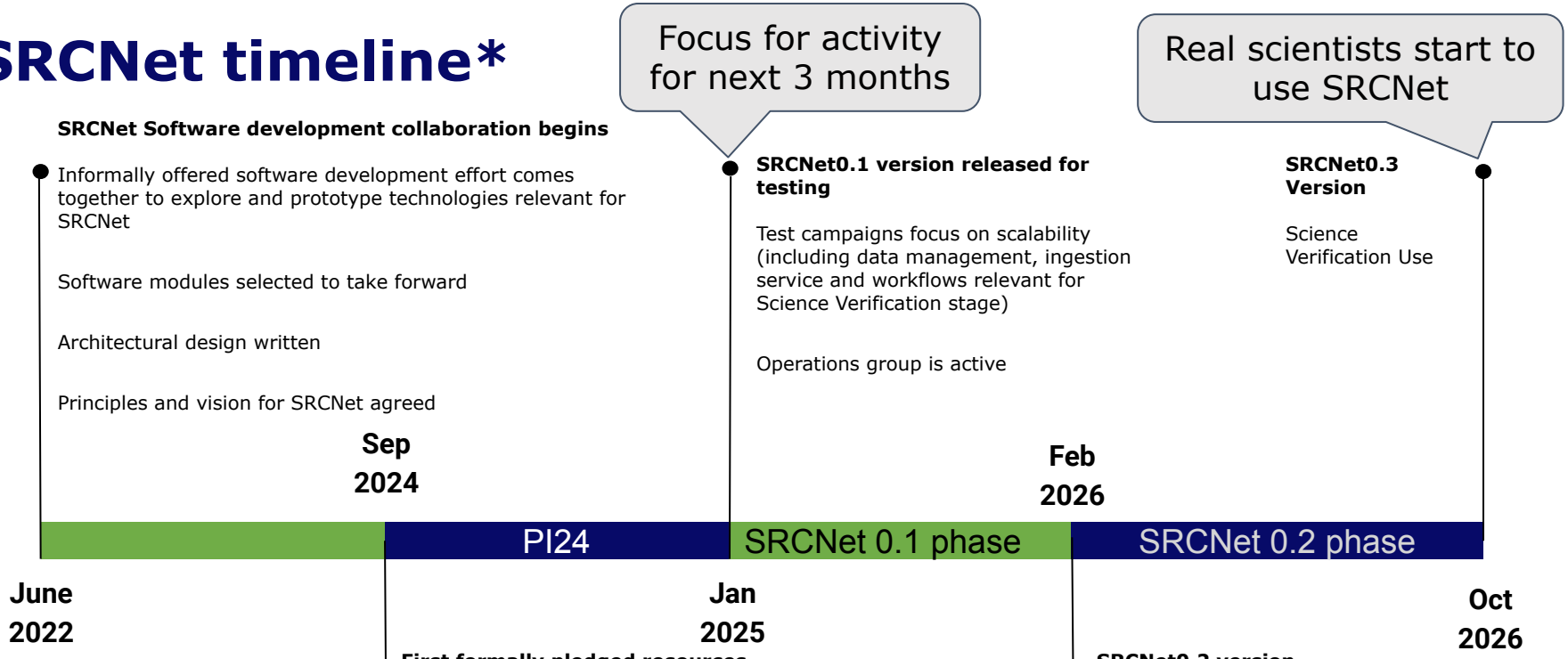
Engagement from across most SKA countries

Currently 40+ persons-worth of effort

70+ contributors



SRCNet timeline*



PI: Program Increment (3 month development cycle)

*may change

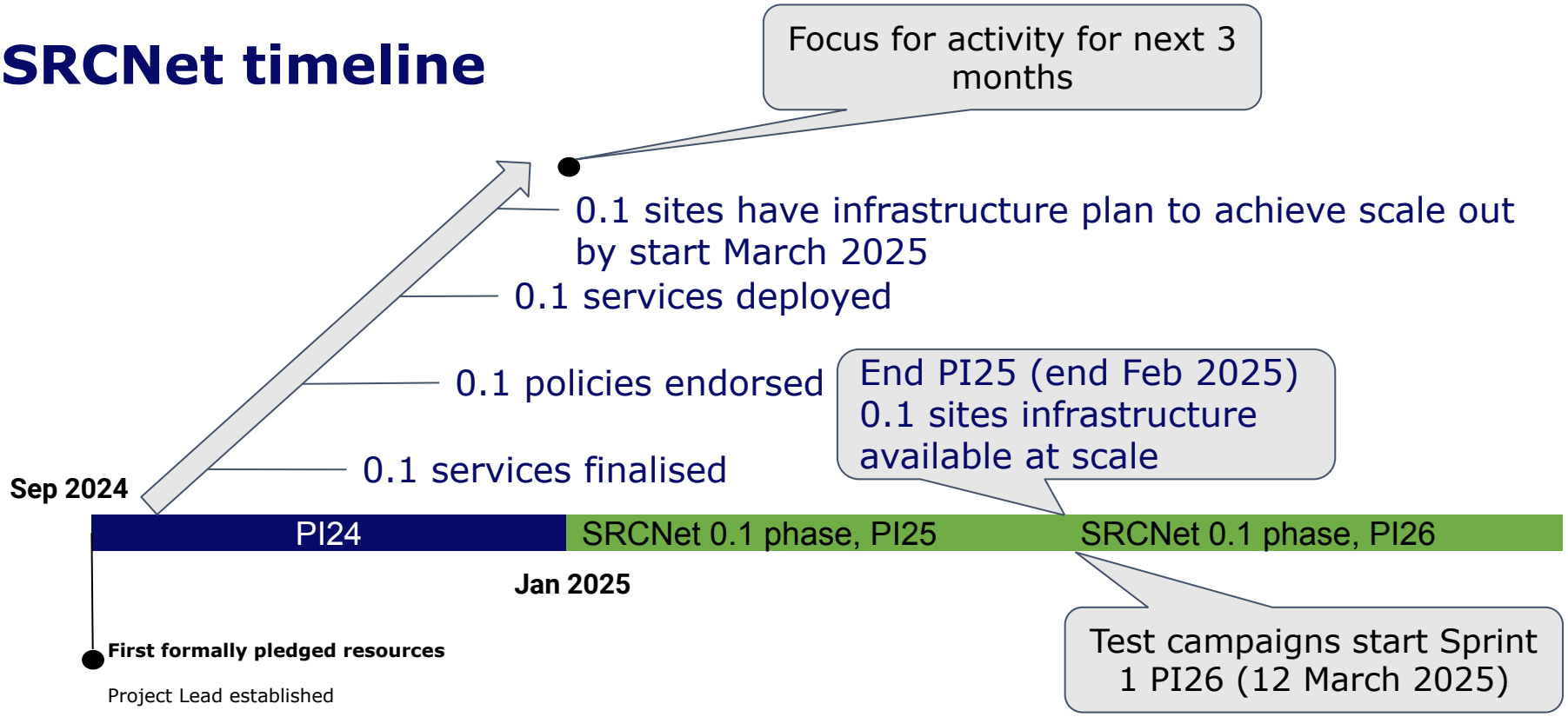
SRCNet 0.1 is our first big milestone!!

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

First test of full SRCNet architecture

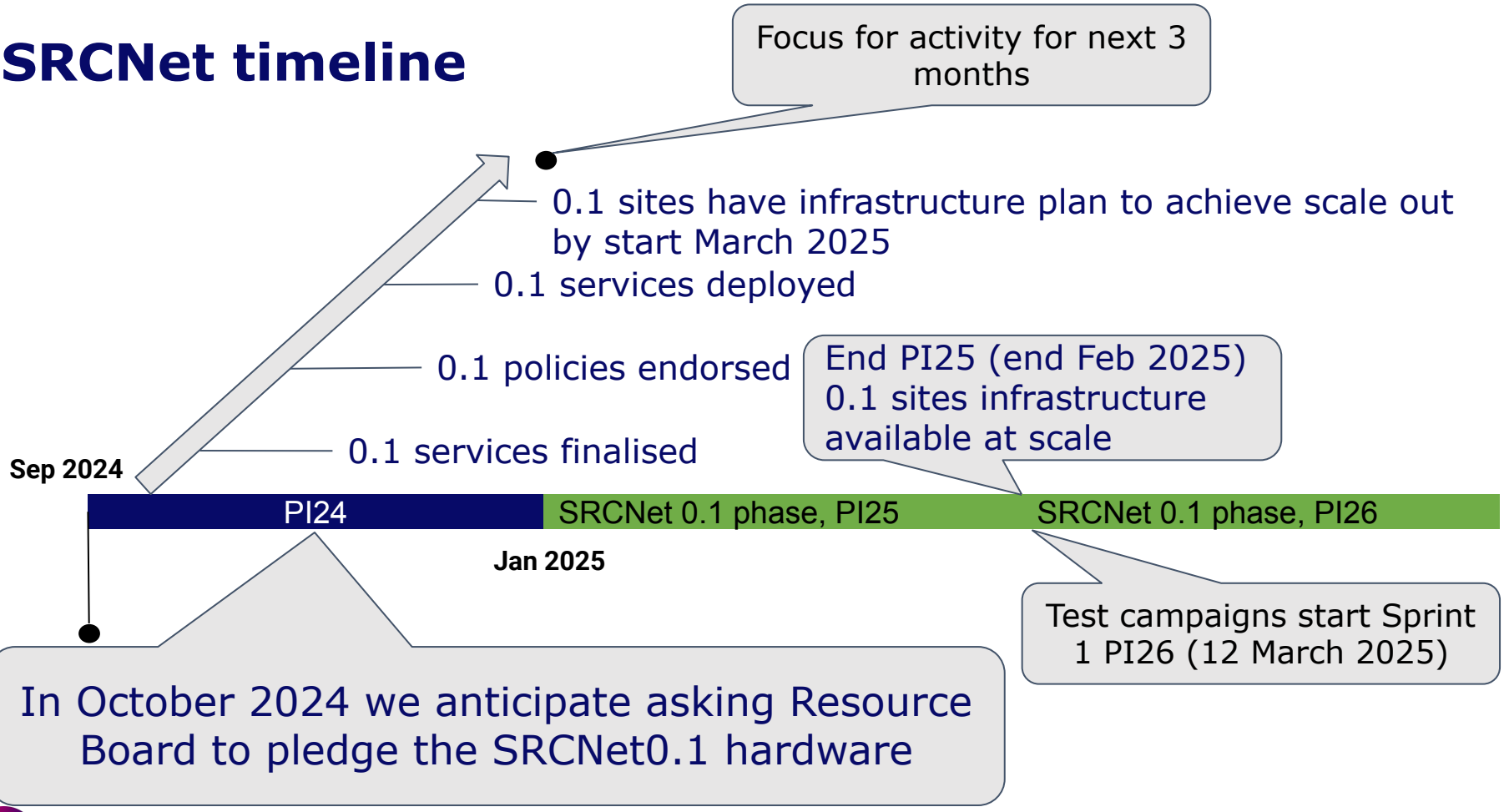
Due end of PI24; ie. 20th November 2024 (final work day of sprint 5)

SRCNet timeline



PI: Program Increment (3 month development cycle)

SRCNet timeline

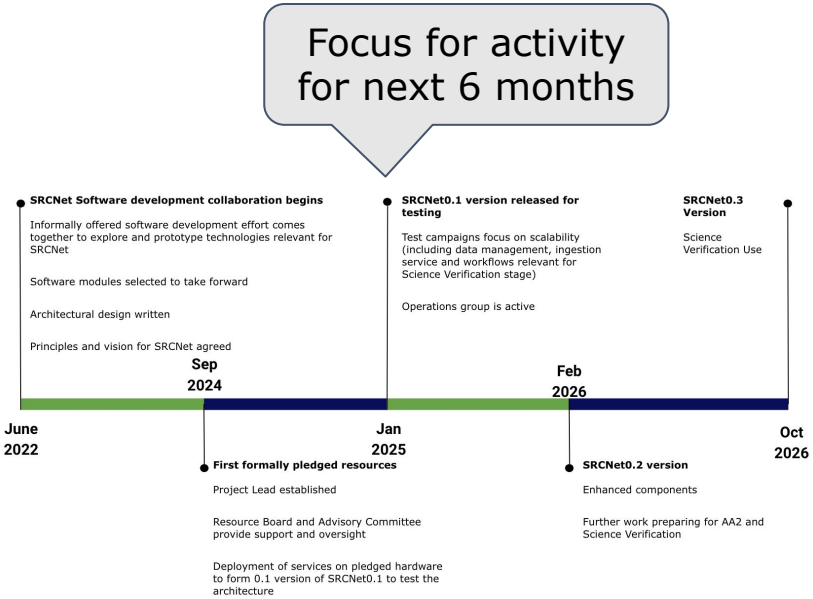


PI: Program Increment (3 month development cycle)

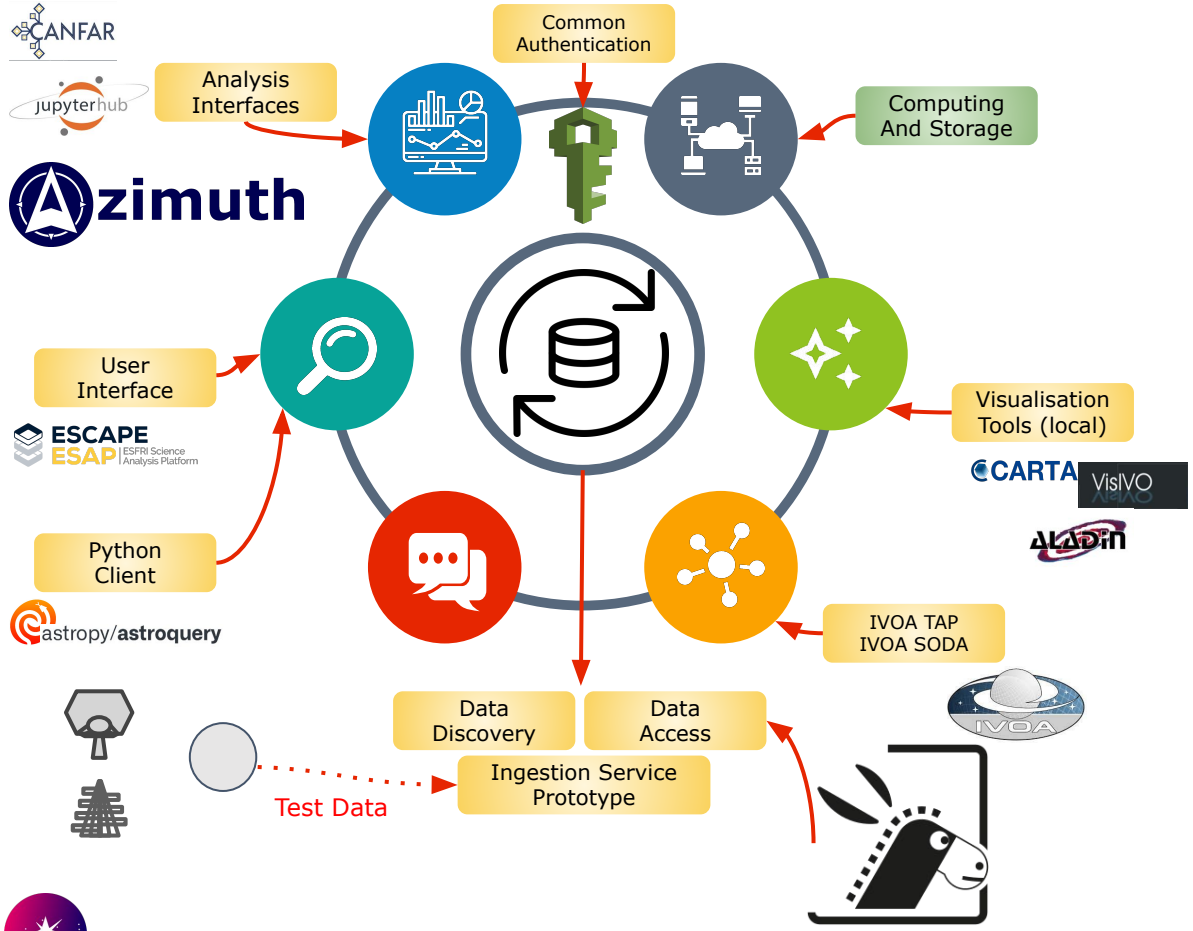
SRCNet0.1

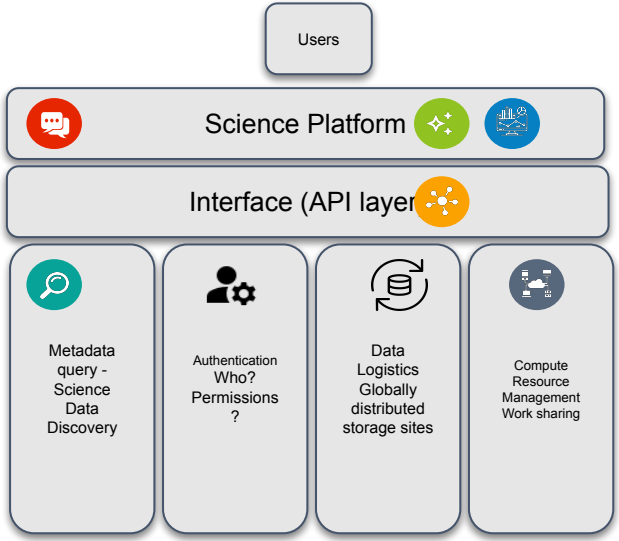
This is an "engineering" version

- **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



Basic Functionality Covered by v0.1



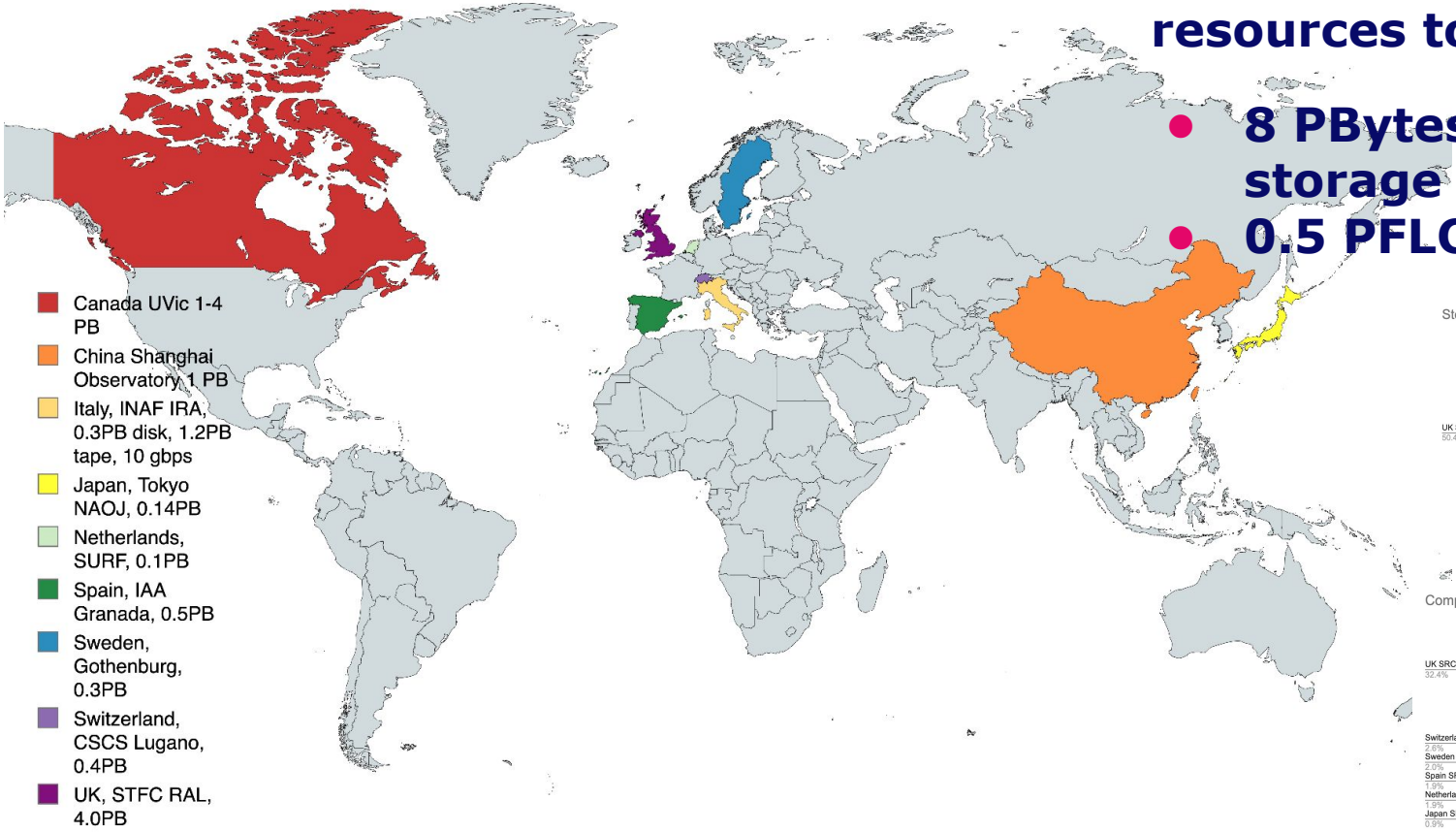


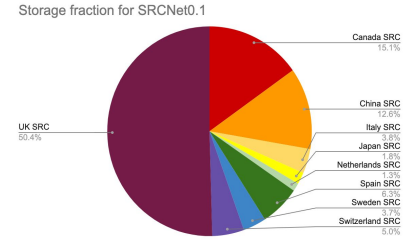
- Common Authentication: IAM
- Visualisation Tools (local)
- IVOA Protocols: TAP, SODA
- Data Discovery and Access from (Rucio)Data Lake
- Ingestion Service Prototype
- Python Client: Astroquery Module
- User Interface: Gateway **https://gateway.srcdev.skao.int/**
- Analysis Interfaces: JupyterHub (compulsory);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

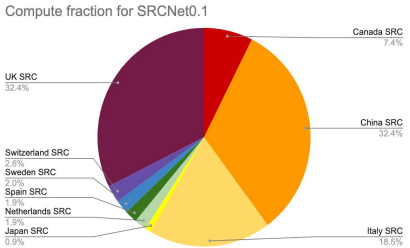
SRCNet0.1 planned sites

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

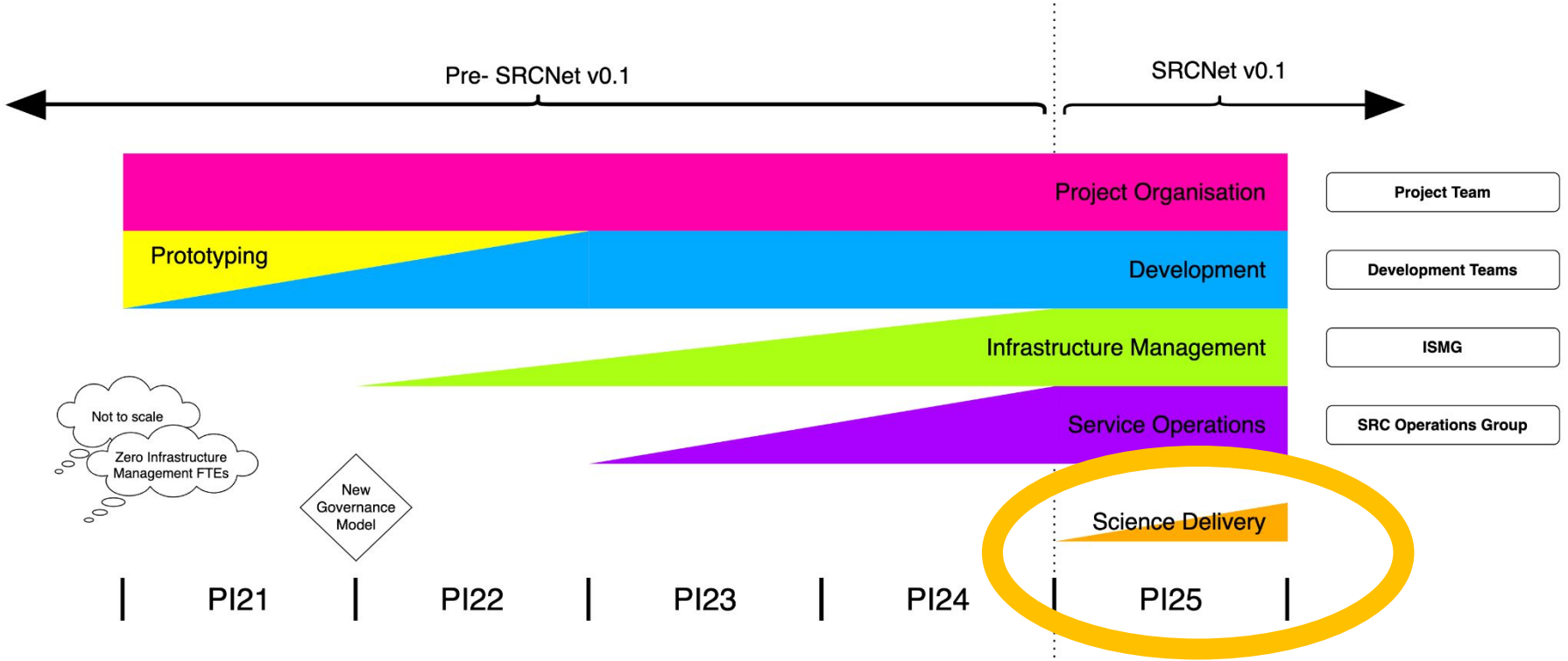
- 8 PBytes total storage
- 0.5 PFLOPS Compute







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
TOTAL	**40**	**42**	**60**

current PI numbers

PI24 planning 2-6 September!

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

PI: Program Increment (3 month development cycle)
FTE" Full-Time Equivalent, Personnel

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