SKAO REGIONAL CENTRE NETWORK

Enabling SKA science in the global SKA Regional Centre Network

Rosie Bolton
rosie.bolton@skao.int

Interim SRCNet Project Lead

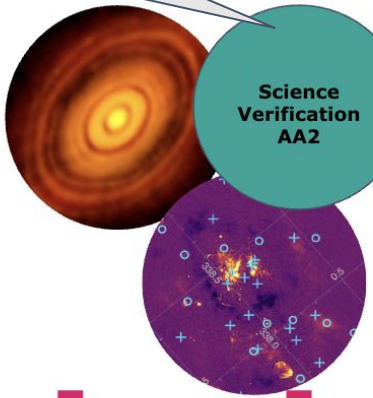
2/09/2024

SKA Observatory Time line to Science

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

Construction



Science Verification AA2

Science Verification AA*

Formal end of construction

Call for cycle 0 proposals

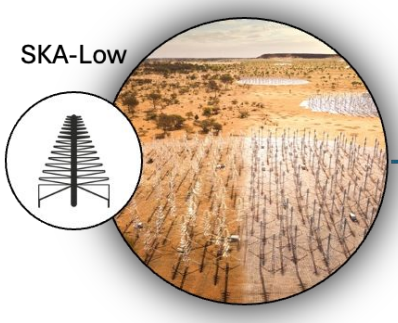
2024 2025 2026 2027 2028 2029

Commissioning

*CURRENT ESTIMATES, subject to change

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

CSP

8.9 Tb/s

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

A thousand times faster than "good" home internet speed

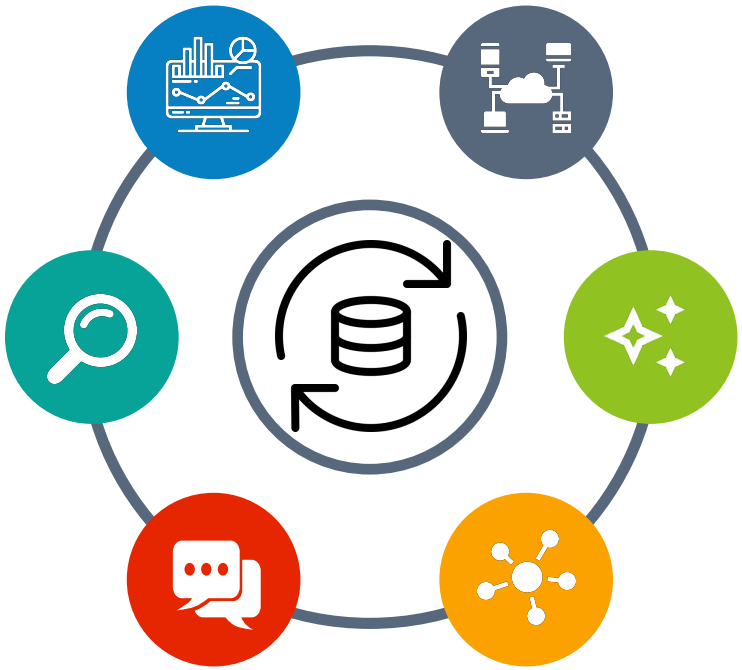
What are the SKA Regional Centres???

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.

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

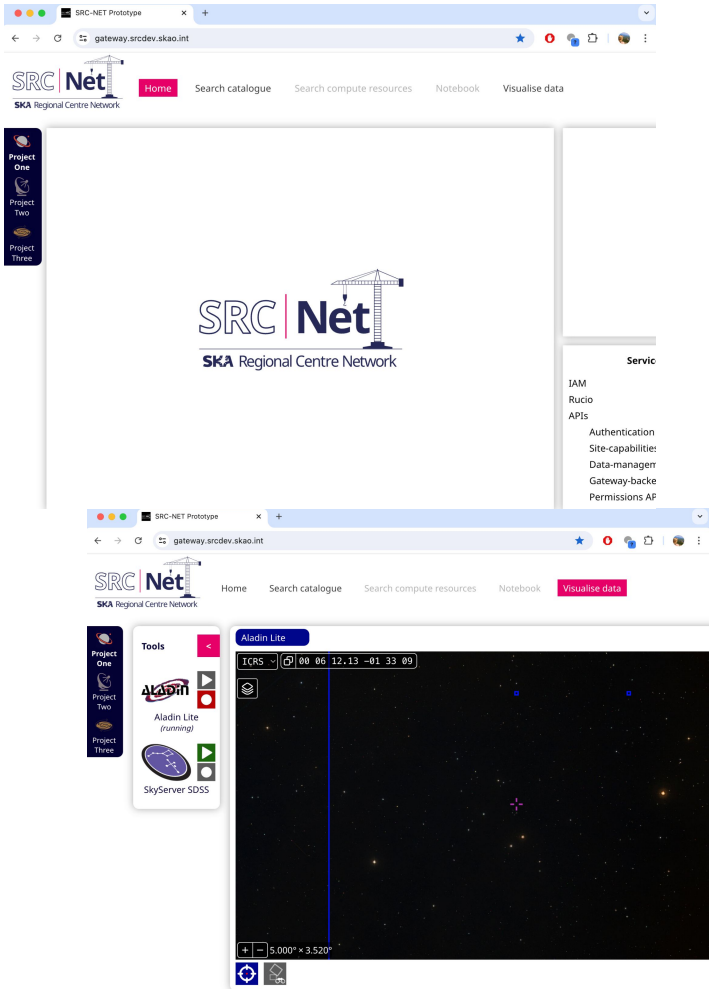
Intended user perspective

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

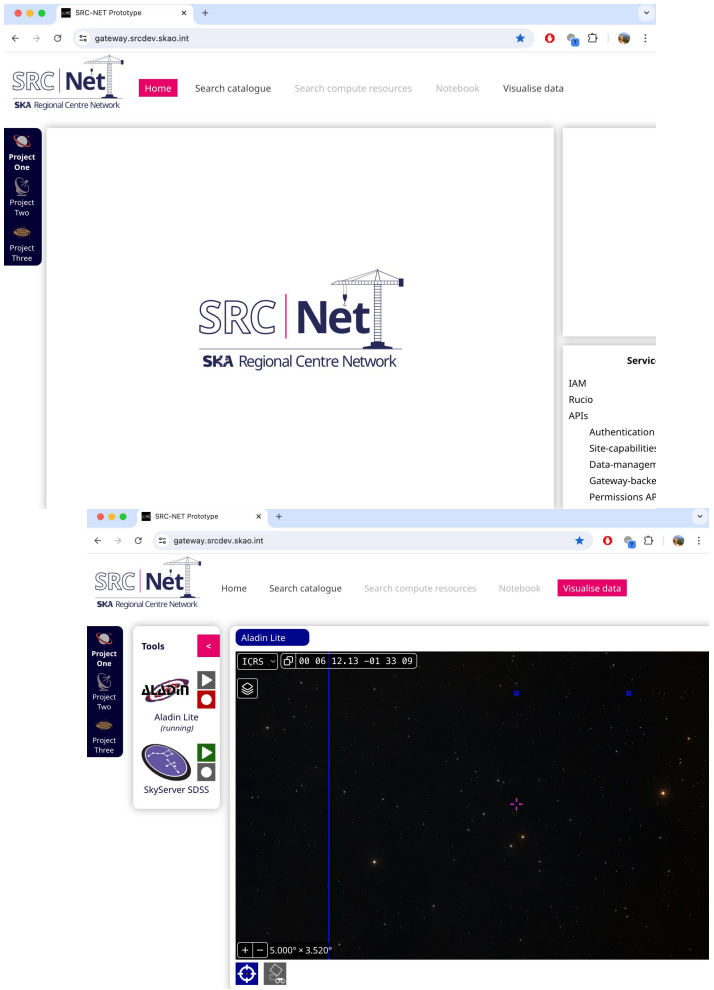


Intended user perspective

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

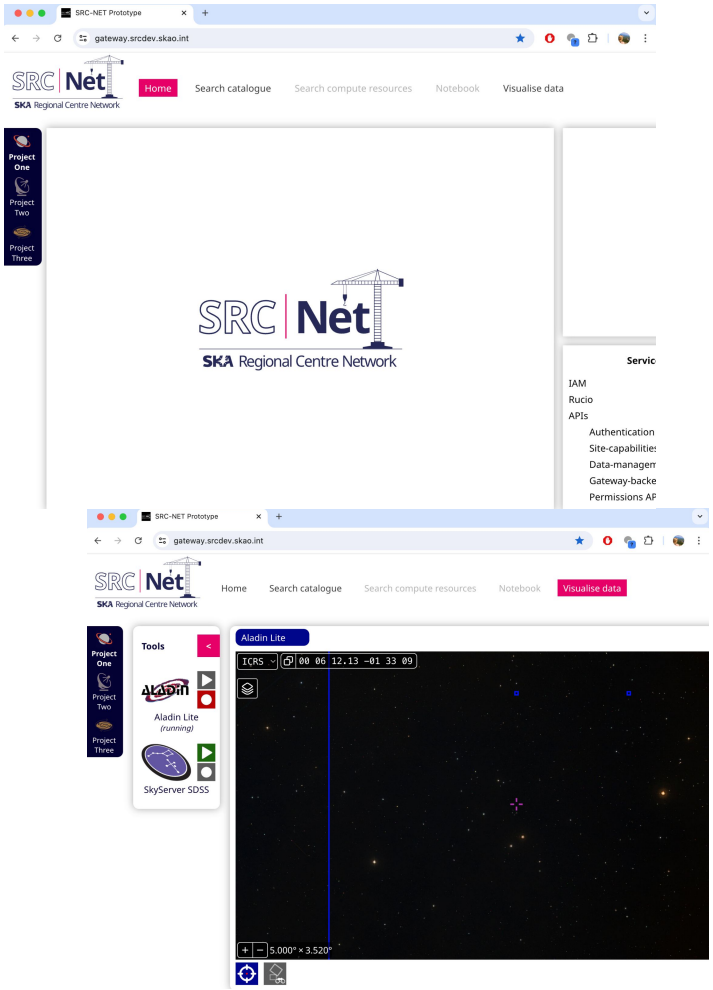


Intended user perspective

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

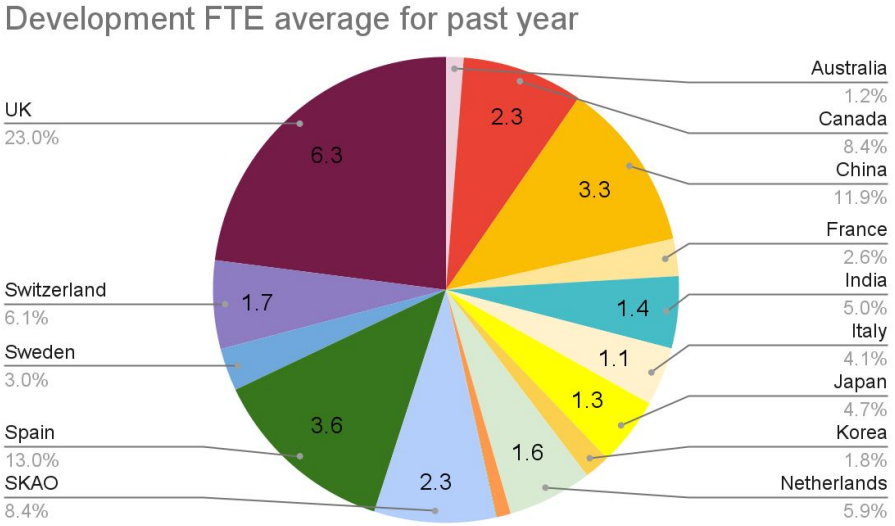


The SRCNet Project

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

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

SRCNet0.3 Version

Science Verification Use

Feb 2026

June 2022

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

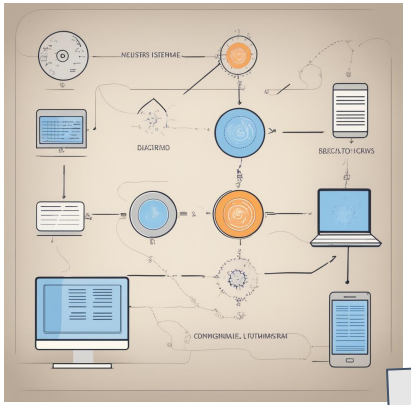
SRCNet0.2 version

Enhanced components

Further work preparing for AA2 and Science Verification

*may change

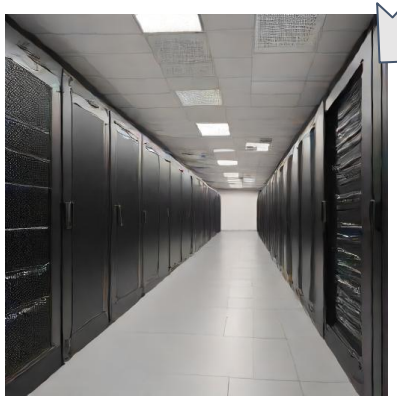
SRCNet composition



Software & services

SW Development and Service operations

SRC Net!



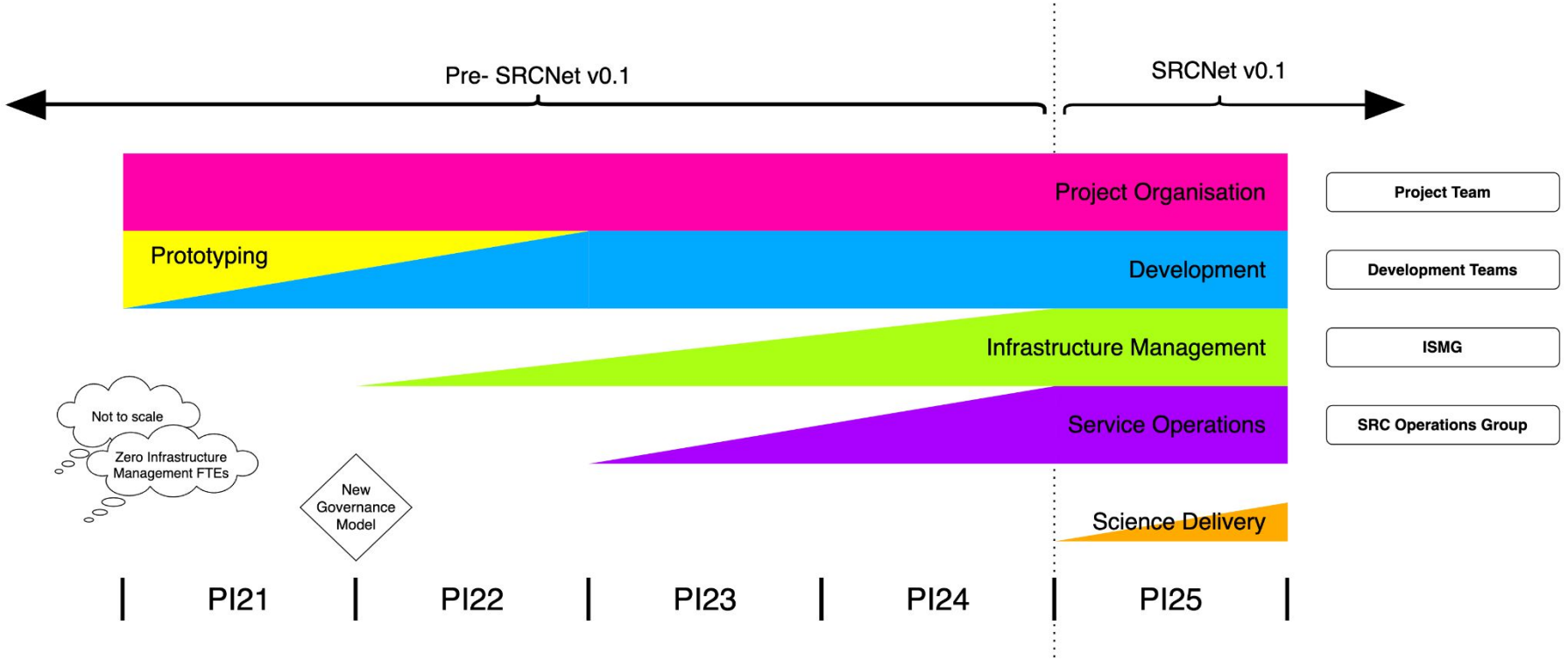
Hardware

Science Delivery

Science Users

DeepAI images

Value Stream Development



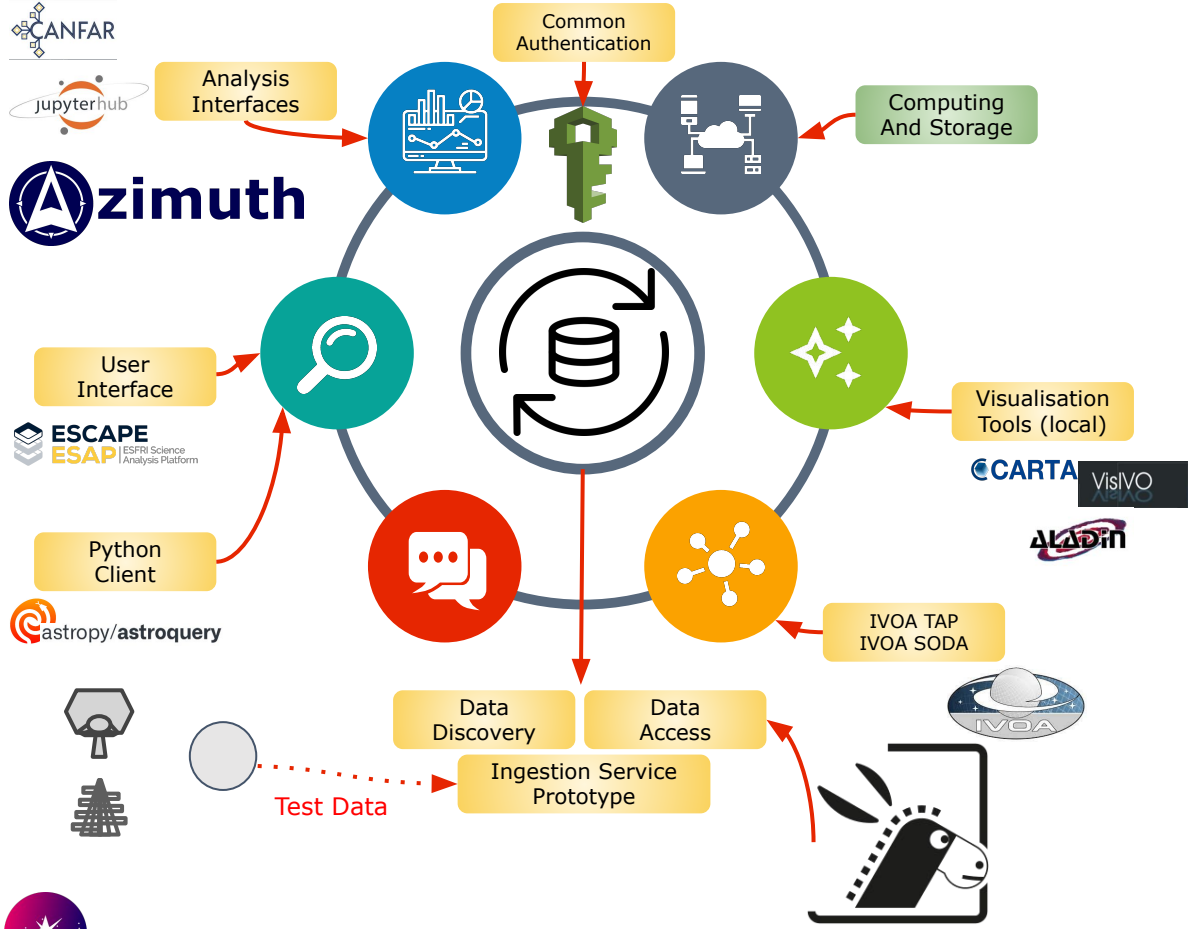
SRCNet 0.1 is our first big milestone!!

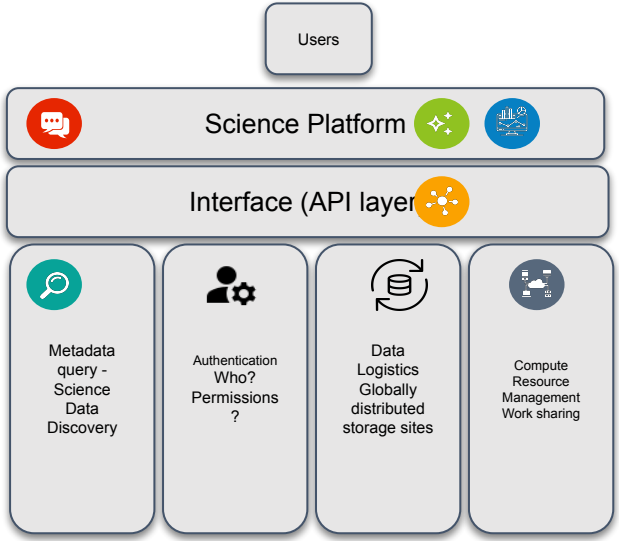
At least 4 sites running full set of compulsory local services; global services also running to support this

First test of full SRCNet architecture

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

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)

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

SRCNet0.3 Version

Science Verification Use

Feb 2026

June 2022

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

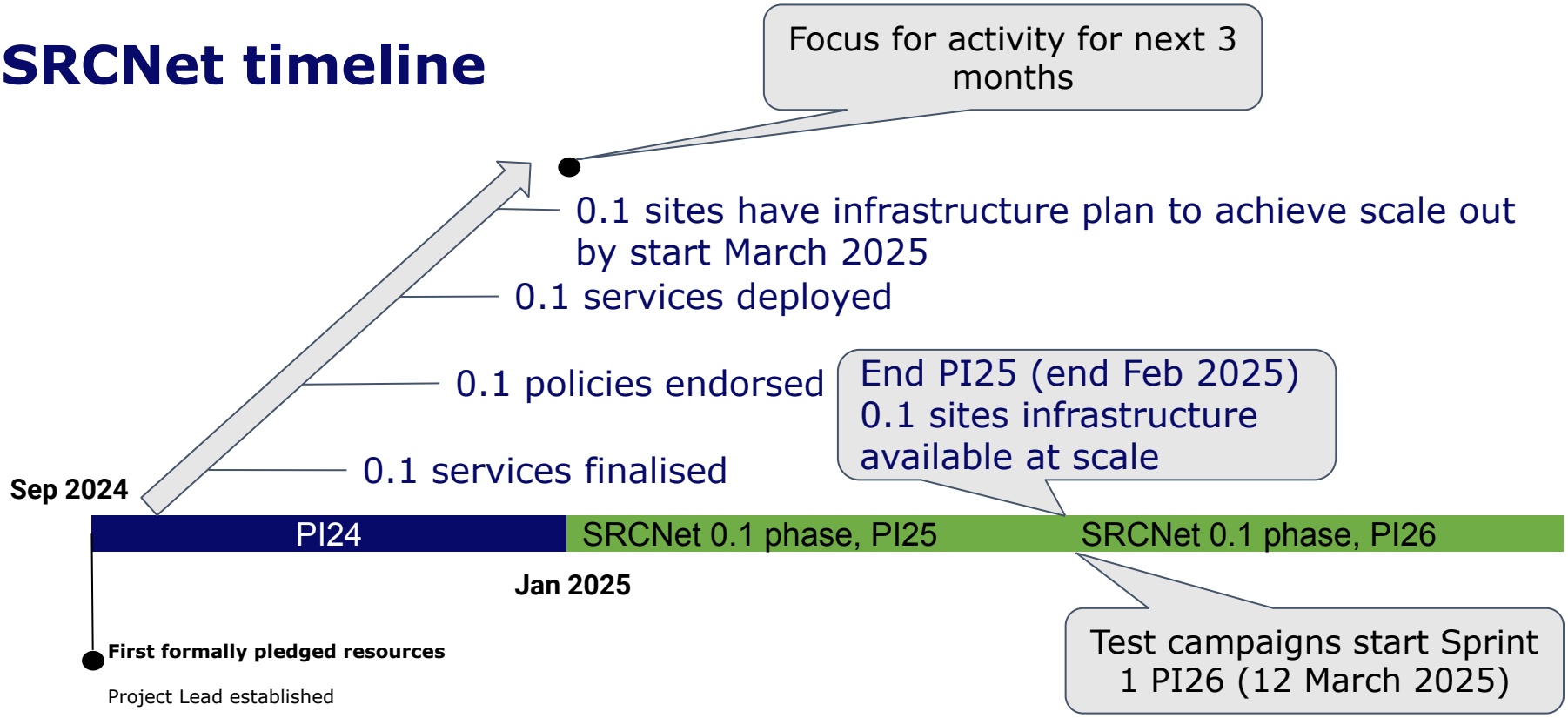
SRCNet0.2 version

Enhanced components

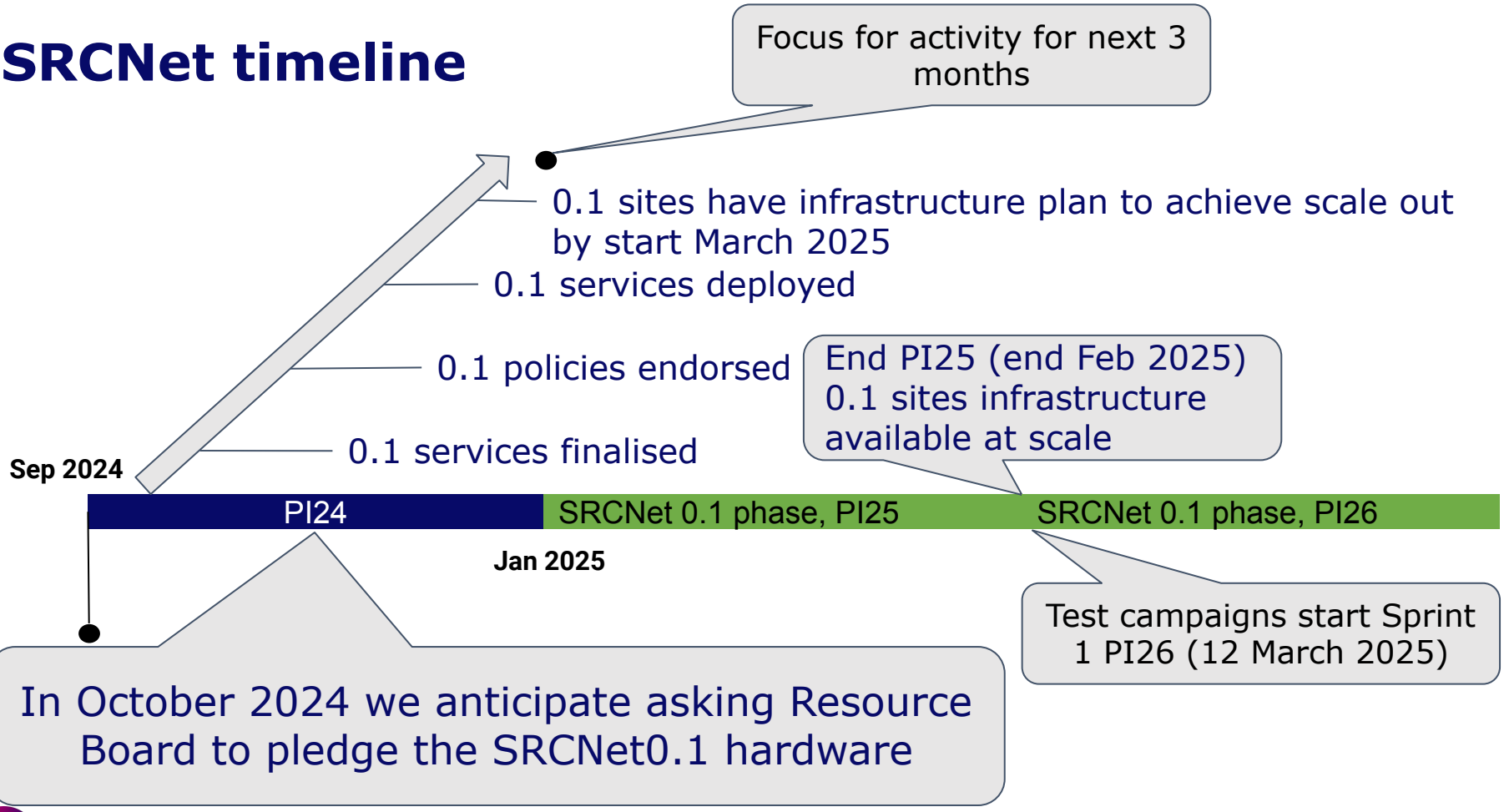
Further work preparing for AA2 and Science Verification

*may change

SRCNet timeline



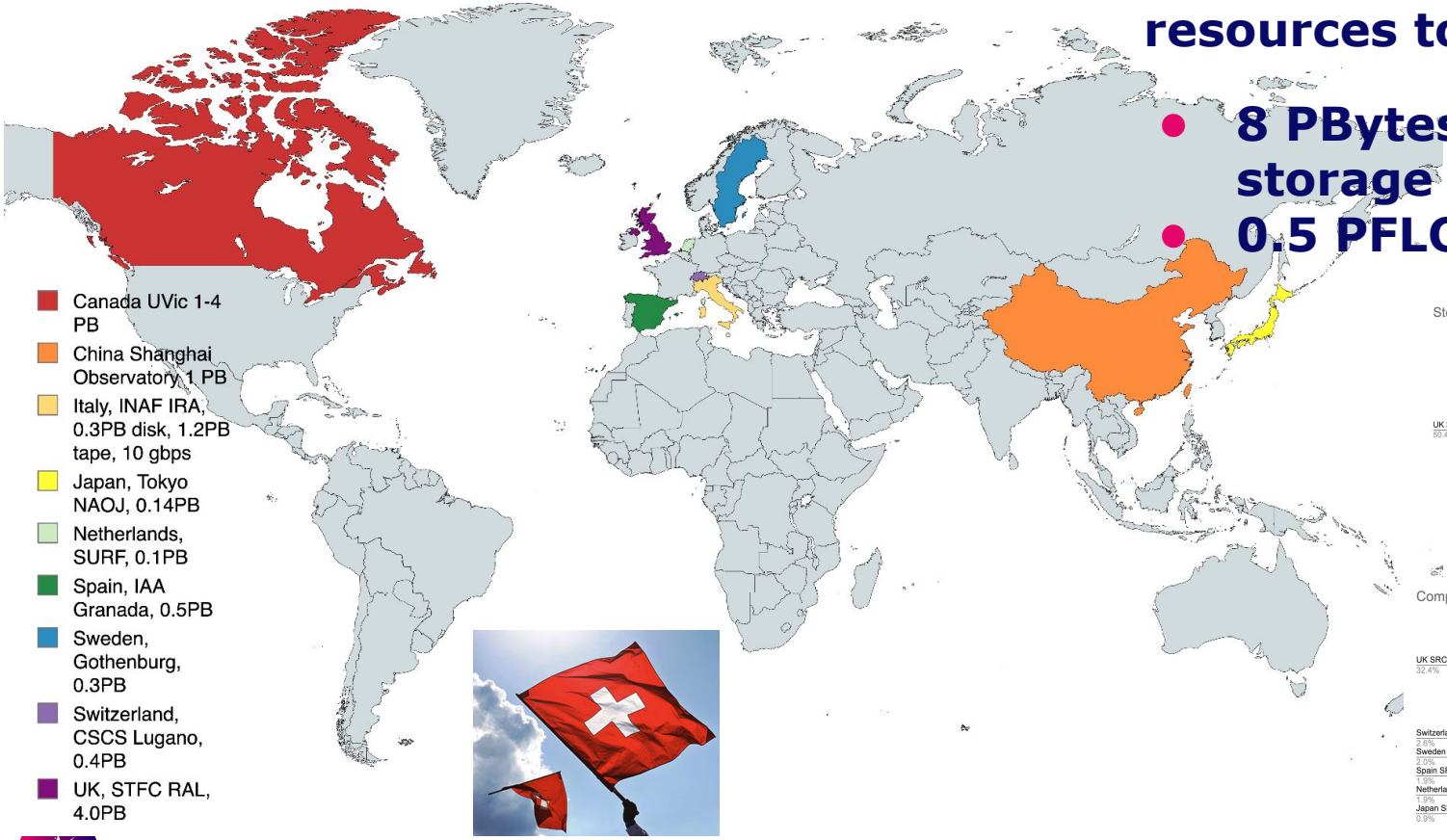
SRCNet timeline

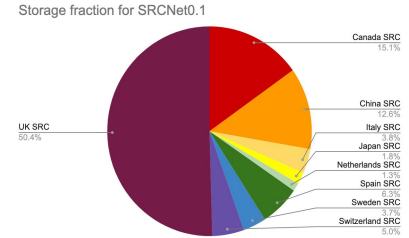


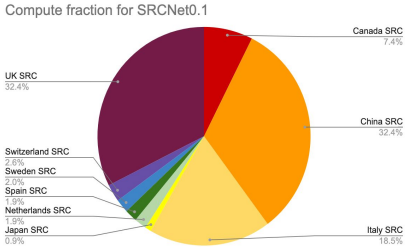
SRCNet0.1 planned sites

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

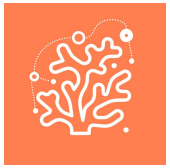
- 8 PBytes total storage
- 0.5 PFLOPS Compute

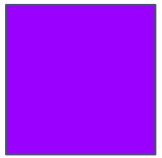




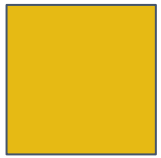


The *current SRCNet* teams

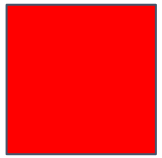

Coral


Purple


Lavender


Gold


Orange


Red


Magenta


Teal


Tangerine


Chocolate

DAAC

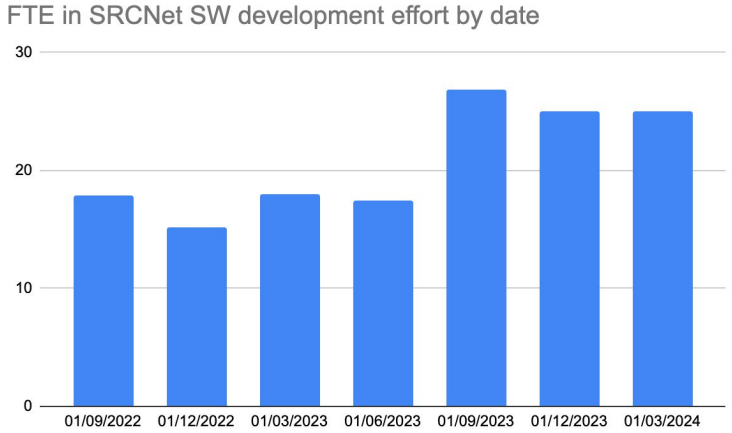
Indigo

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



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**

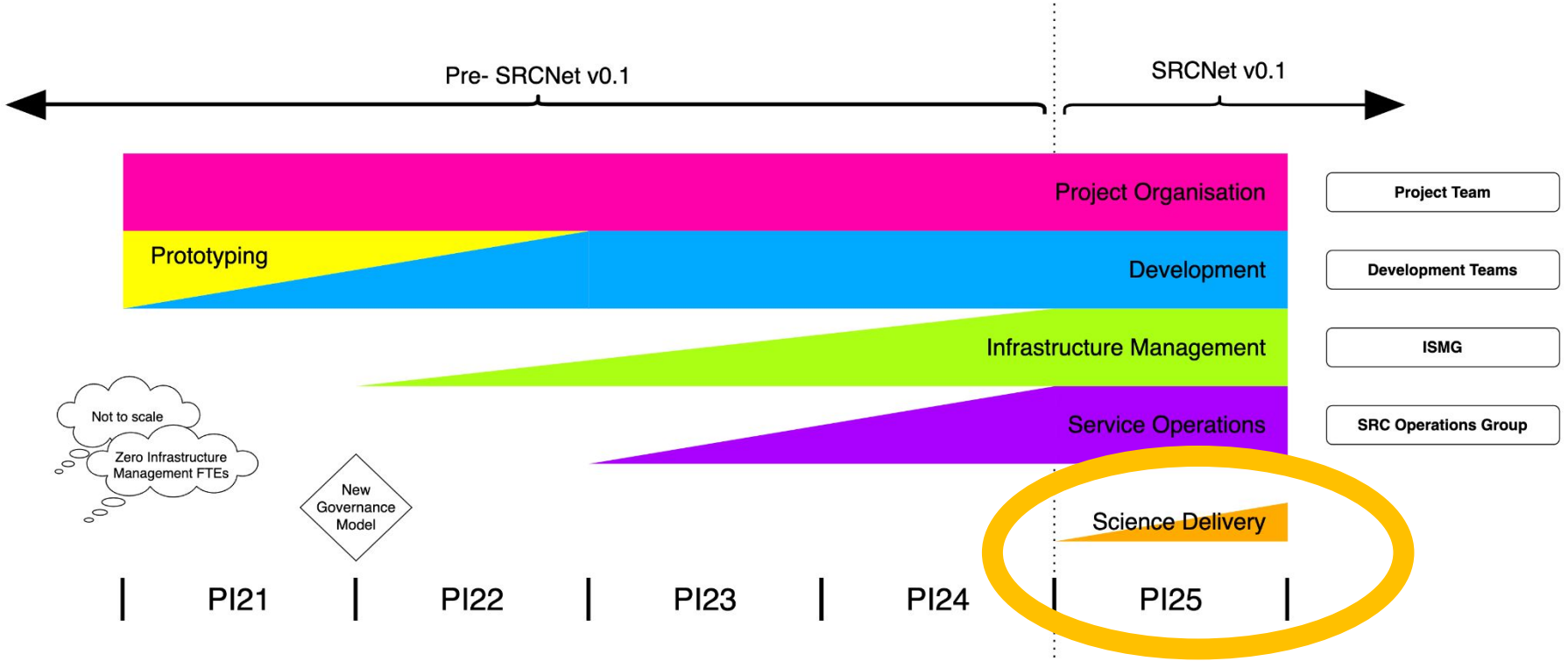
current PI numbers

PI24 planning this week!

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

Value Stream Development - Science Delivery



Current scientific work in SRCNet

We are building a suite of example workflows to support testing of our sites and to develop benchmarking tools

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

Chocolate team are using Karabo simulation software to make realistic data sets that could enable end-to-end SRCNet tests

Soon (by December) we will seek to strengthen our science work with additional FTEs and specific roles to help community engagement

End

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

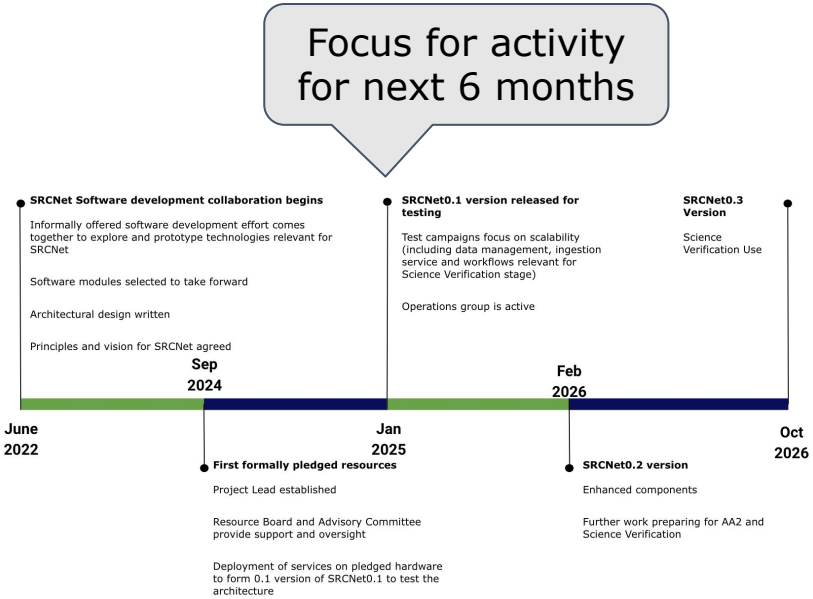
SKAO

www.skao.int

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



SRC Functionalities

Diverse set of components and activities to support users

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

SRC Network Principles (some of them!)

- *There will be a **common SKAO/SRC Network user account** that allows users access to SRC Network resources*
- *English will be the primary language of communication across the SRC Network*
- *There will be **one Helpdesk system** for the SRC Network and the SKAO.*
- *The SRC Network will **optimise its energy usage** whilst meeting the scientific goals of projects carried out in the SRC Network.*
- *Security of the SRC Network is the responsibility of the SRC Network.*
- *The SRC Network will **lead with principles of fairness, equity and inclusion** in all of its activities, and seek diversity of staff.*
- *The SRC Network will be committed to providing, and abiding to, **accessible and equitable tools, practices and processes**.*
- *The SRC Network will **provide workflow templates** to carry out basic and standard processing tasks.*
- *The SRC Network will **embrace FAIR and Open Science principles** whenever possible and appropriate.*
- *Resources pledged into the SRC Network will enter, and be allocated from, **a global federated pool**.*
- *The **allocation of resources will be per project.***
- *The **physical location of SKA data** products will be determined to **optimise access and minimise data redistribution** within the Network, as much as is feasibly possible.*
- *Data processed within the SRC Network will **automatically propagate all metadata and provenance information**.*

Highlighted SRC Network Principles

- *There will be a **common SKAO/SRC Network user account** that allows users access to SRC Network resources*

Single AAI System used by all SRCNet0.1 sites and services

Common policies for SRCNet sites (in addition to local policies)

- *Security of the SRC Network is the responsibility of the SRC Network.*

Ensure good user experience, for all users - Science Gateway UX

- *The SRC Network will be committed to providing, and abiding to, **accessible and equitable tools, practices and processes**.*

Implement IVOA standards and easy data and service discovery

- *The SRC Network will **embrace FAIR and Open Science principles** whenever possible and appropriate.*

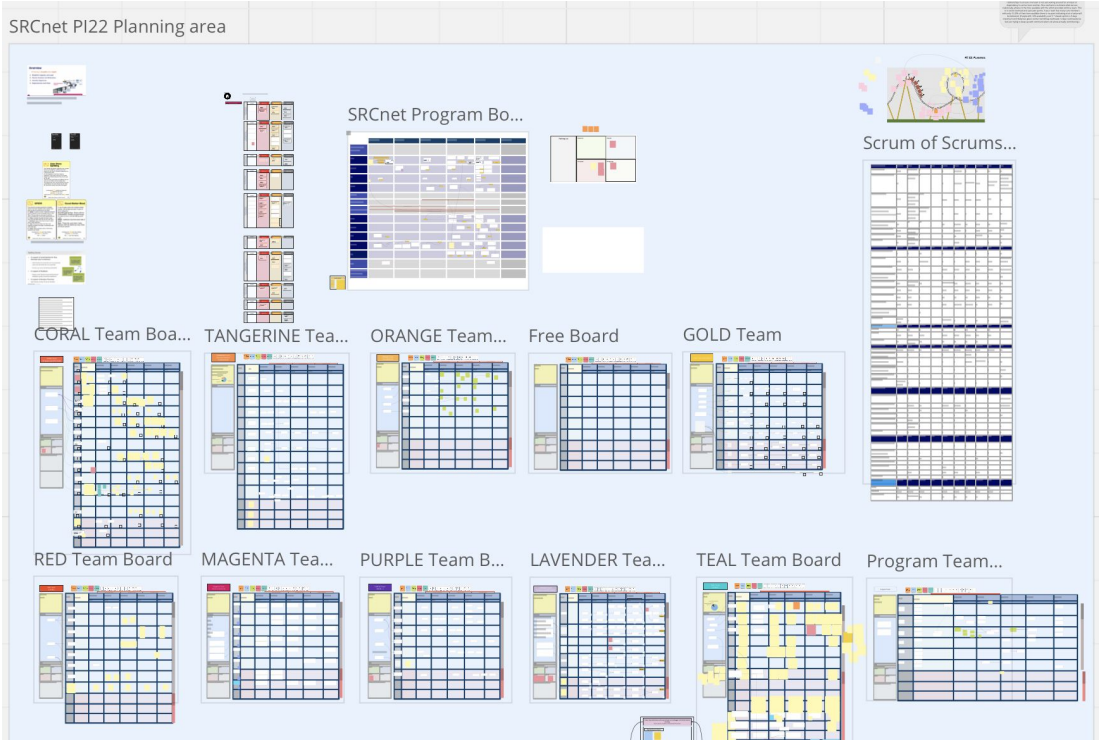
Single AAI System used by all SRCNet0.1 sites and services

- *Resources pledged into the SRC Network will enter, and be allocated from, **a global federated pool**.*

These two are related; users go to best location depending on data location and appropriate available services, replicas centrally planned / moved to optimise global access

- *The **physical location of SKA data** products will be determined to **optimise access and minimise data redistribution** within the Network, as much as is feasibly possible.*

Scaled Agile Framework for SW development



- **Teams Plan together (usually "distributed co-location") once per 3 months**

- **Regular demos (open)**
- **Several Communities of Practice (Identity management, science platform, HPC & Cloud)**
- **Advisory forums**
  - **inc. NREN forum**

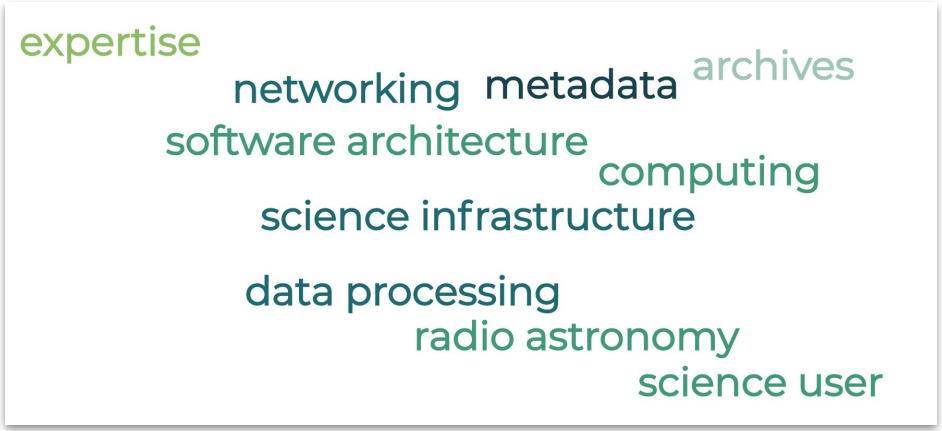
We are now updating management structures of the SRCNet project currently - Should hopefully give more stable resourcing and clearer understanding of roles.

SRCNet Advisory Committee

The primary role is to provide technical advice to the SRCNet Project Lead regarding technology and software choices including alignment with national SRC-related activities and relevant precursor and pathfinder solutions.

A wide range of expertise is needed:

1. Radio Astronomy, Archival data use and IVOA standards
2. Data processing within existing Observatories
3. Data access provision for existing Observatories and Science infrastructures (e.g. SKA Precursors and Pathfinders, ALMA, VLA, LSST, CTAO, CERN/WLCG)
4. Data centre infrastructure concerns: security, policy making, storage, high performance and cloud computing and long-haul networking



Candidate suggestions due by mid August

AC should be up and running by early Q4 2024

Architectural Principles

The main objective of the SRCNet is to maximise the science produced by the community using SKA data

The SRCNet development is a global effort done by all the SRCs

The SRCNet architecture should be scalable

The SRCNet architecture should be extensible

The SRCNet architecture should provide data and Computing Resilience

The SRCNet architecture should follow FAIR principles

The SRCNet Architecture should be designed to minimise cost and environmental impact and maximising throughput

The SRCNet architecture should allow federated execution

The SRCNet architecture should allow reproducibility of the execution of analysis workflows

The SRCNet architecture should ensure Data Integrity

The SRCNet architecture should be secure

The SRCNet architecture should minimise the environmental impact

SRCNet0.1 Implementation Plan

This plan details the following

- **Deployment Timeline:** A defined schedule outlining activities leading up to global deployment and subsequent network testing.
- **Shared Resources:** A breakdown of shared resources available per SRC, along with relevant local milestones.
- **Services and Topology:** A description of mandatory and optional services, along with the expected service deployment topology at the SRC level. That also includes the software stack to be used for every service.
- **Validation Tests:** Procedures for local validation testing across the diverse infrastructure to ensure overall homogeneity.
- **Operational Procedures:** Outline of operational activities including monitoring, services deployment, and coordination activities.
- **Testing Campaigns:** A plan for executing test campaigns on the deployed SRCNet v0.1.

SRCNet Resource Board

Resource Board representatives have been confirmed for all countries (thank you)

First meeting 4th July

Agenda:

- ToR and Election of Chair
- SRCNet Project progress update
- Resources (Current levels FTE and anticipated SRCNet0.1 hardware)
- Pledging mechanisms
- Immediate pledging needs for September-December 2024
- Long term roadmap

Resource Board Representatives

As of 14th June 2024		
Country	**Representative Details**	
AUSTRALIA	Dr Karen Lee-Waddell	Director AusSRC
CANADA	Dr Michael Rupen	Director DRAO/SKA Program Lead
CHINA	Prof Shen Zhiqiang	Director SHAO
FRANCE	Mr Arnauld Leservot	Deputy Head Digital Infrastructures and Services, MESR
GERMANY	Prof Wolfgang E. Nagel	Director of CIDS, ZIH and ScaDS.AI, TUD Dresden University
INDIA	Mr Sunil Ganju	Council Representative/DAE
ITALY	Dr Filippo Zerbi	Council Advisor/INAF
JAPAN	Dr Takuya Akahori	Council Observer/NAOJ
NETHERLANDS	Dr Jessica Dempsey	Council Representative/Director Astron
PORTUGAL	Mr Tiago Roque Peres	Portuguese Space Agency
SOUTH AFRICA	Mr Imraan Patel	Council Advisor/Deputy D-G, Department of Science & Innovation
SOUTH KOREA	Dr Hyunwoo Kang	Council Observer/KASI
SPAIN	Dr Francisco Colomer	Secretary General for Research, Ministry of Science & Innovation
SWEDEN	Prof John Conway	Director Onsala Space Observatory
SWITZERLAND	Ms Carolyn Crichton	Board Chair SKACH
UK	Mr George Madden	Head of SKAO Programme, STFC
SKAO	Dr Lewis Ball	Director of Operations, SKAO

SRCNet "Project Team"

This is the management group for the SRCNet project, including project management functions and software (agile release train) management

Rosie Bolton in post as Interim Project Lead

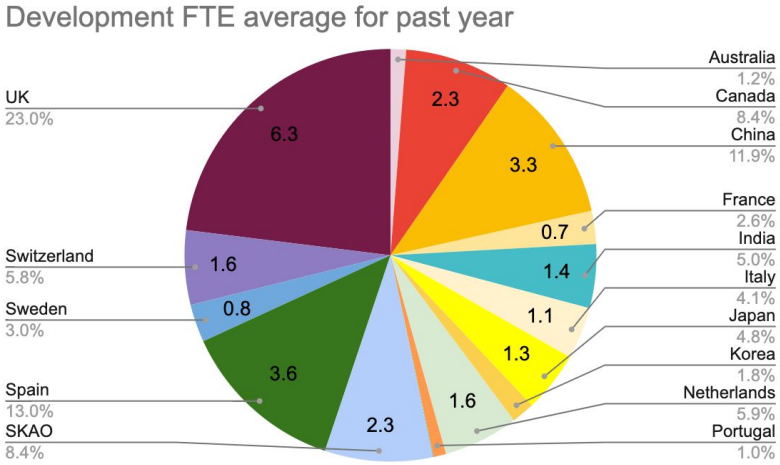
New Product Manager (Debashis Mitra) joined June 2024

Contributions to the Project Team come from Australia, Netherlands, SKAO and UK

Role	Country	Name	FTE
Project Lead	SKAO	Rosie Bolton	1
Project Manager	Netherlands	Janneke de Boer	0.75
Project Administrator	SKAO	Debra Turley	0.2
Release Train Engineer	UK	Jeremy Coles	0.65
Product Manager	SKAO	Robert Perry	1
Product Manager	Australia	Debashis Mitra	1
SRCNet Architect	SKAO	Jesús Salgado	1

Development effort in the SRCNet ART

Table shows development effort per quarter for the past year.



Metrics to help identify inefficiencies and make improvements will be provided to the Resource Board.

	Sep 2023	Dec 2023	March 2024	June 2024
	PI20	PI21	PI22	PI23
Australia	0.4	0.0	0.4	0.6
Canada	2.3	2.3	2.3	2.4
China	0.8	4.6	4.5	3.1
France	0.5	0.8	0.9	0.6
Germany	0.0	0.0	0.0	0.0
India	2.2	0.0	0.0	3.3
Italy	1.6	1.1	0.8	1.0
Japan	1.2	1.4	1.4	1.2
Korea	0.4	0.7	0.9	0.0
Netherlands	1.9	1.8	1.6	1.2
Portugal	0.5	0.6	0.0	0.0
South Africa	0.0	0.0	0.0	0.0
SKAO	2.1	2.0	2.4	2.8
Spain	4.4	3.6	3.8	2.6
Sweden	0.8	0.8	1.0	0.7
Switzerland	1.4	0.9	1.2	2.9
UK	4.5	3.6	5.7	11.5
Total	25.0	24.0	27.0	33.8