SRCNet Project SWG Update

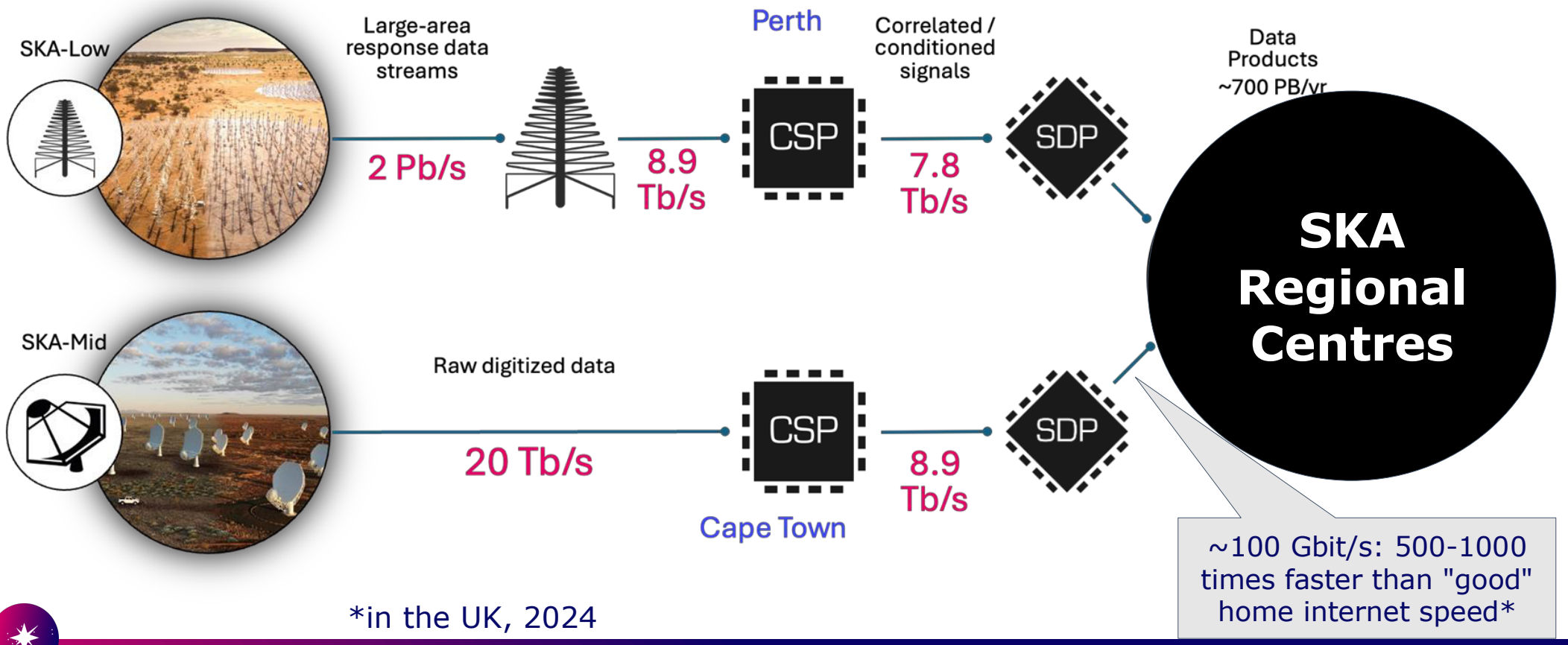
Rosie Bolton
rosie.bolton@skao.int

SWG Chairs meeting 19/11/2024

Document classification: Project Use Only

What are the SKA Regional Centres???

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



*in the UK, 2024

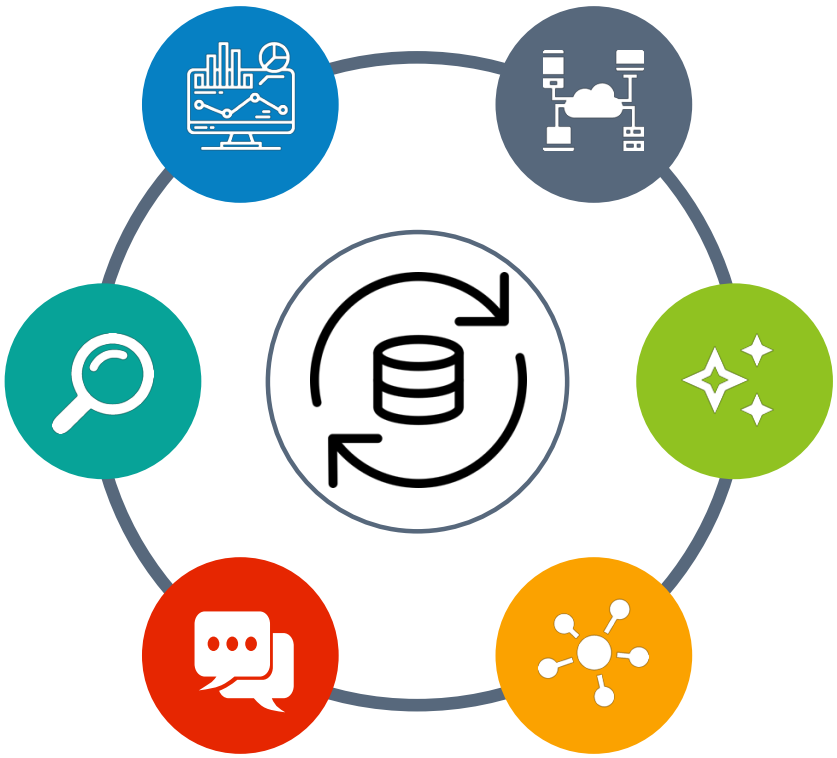
What are the SKA Regional Centres???

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.

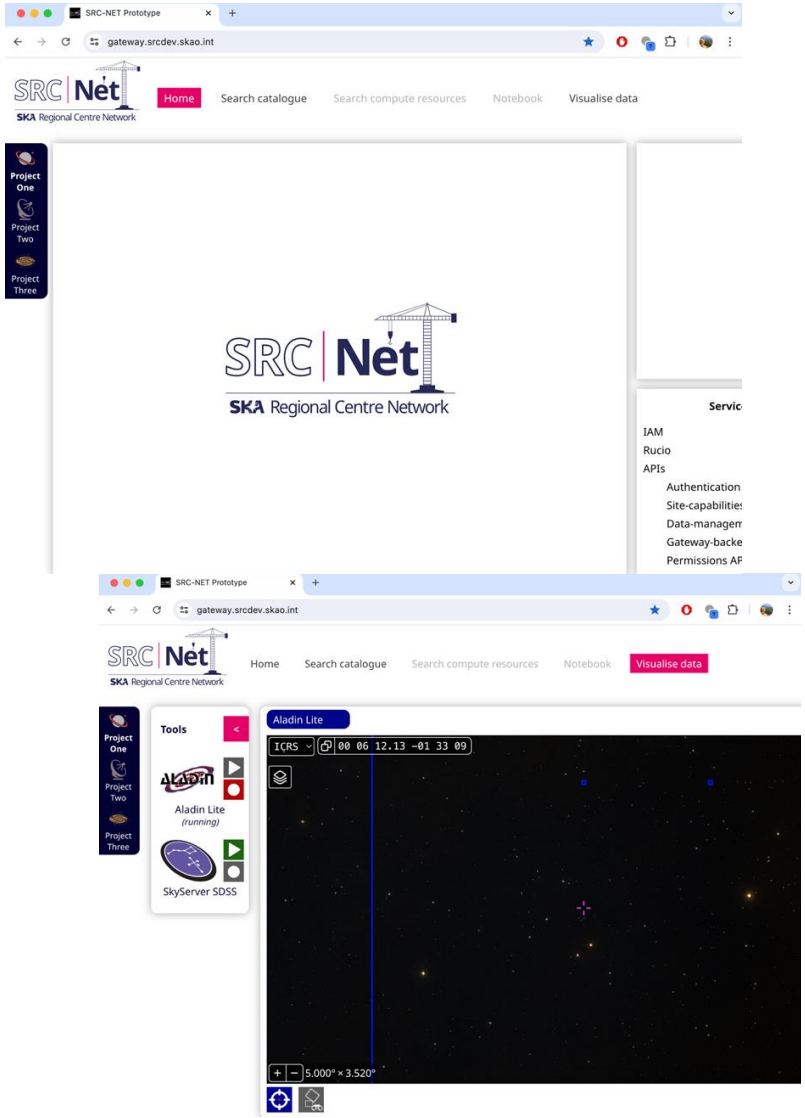
Intended user perspective

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

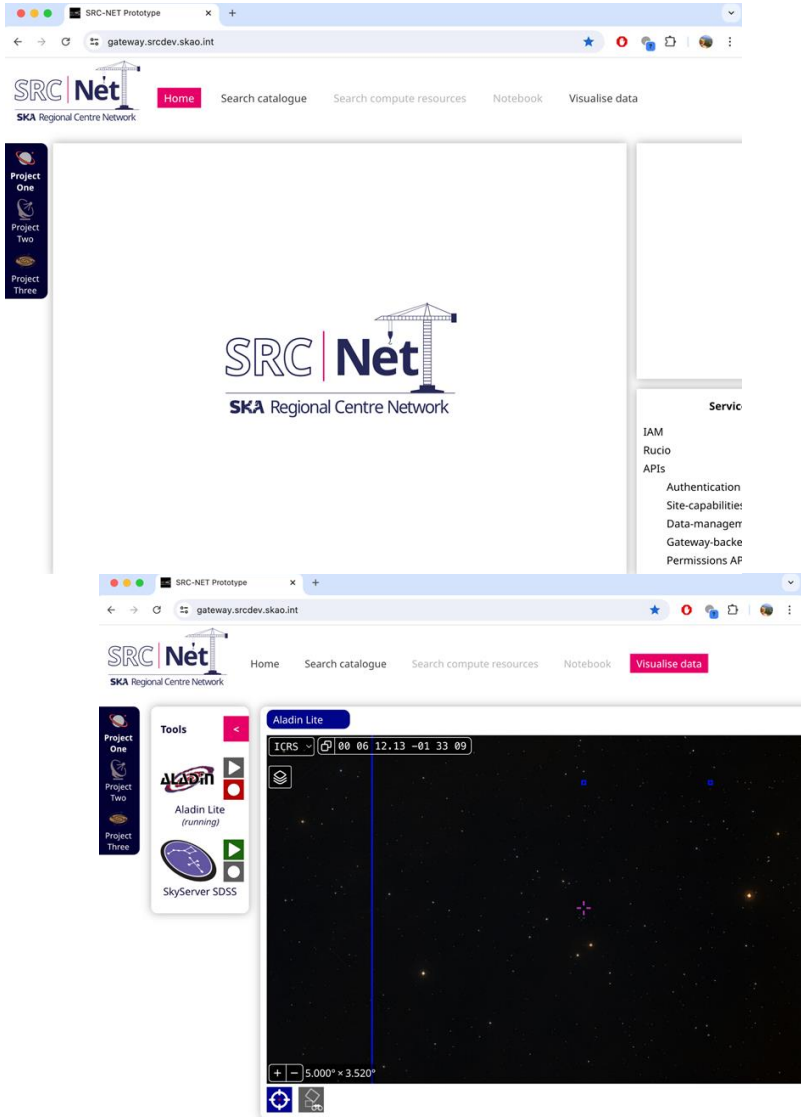


Intended user perspective

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

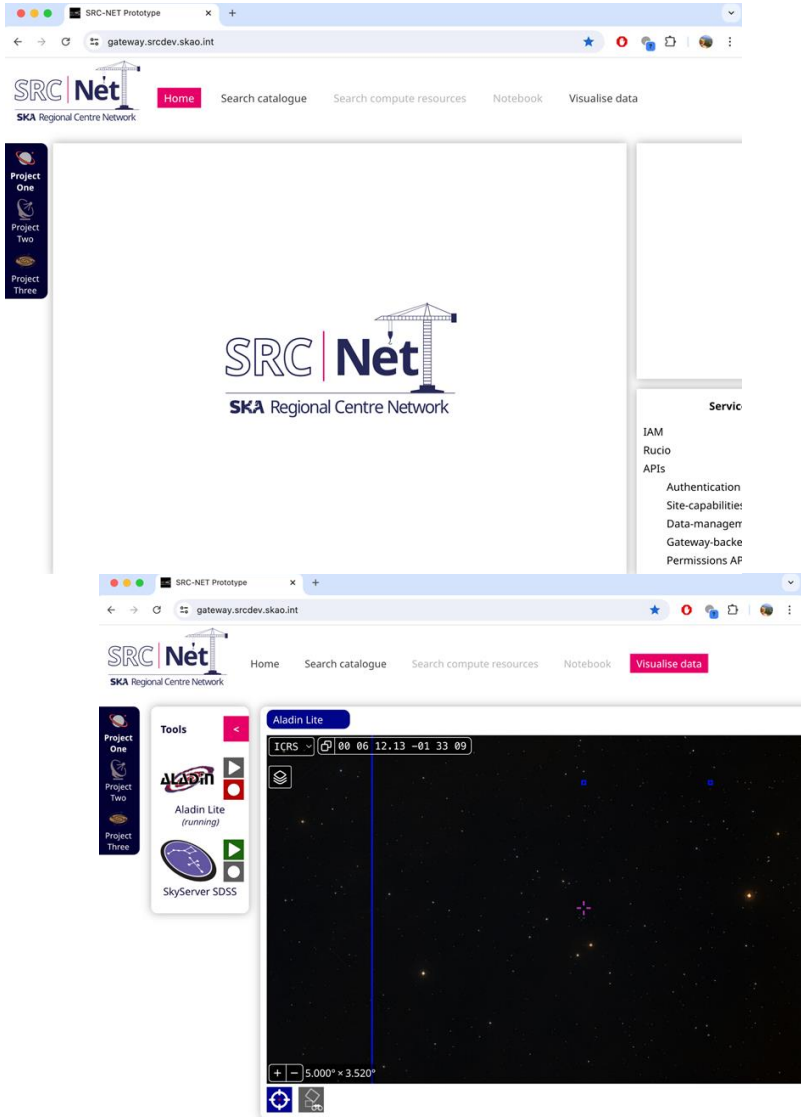


Intended user perspective

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

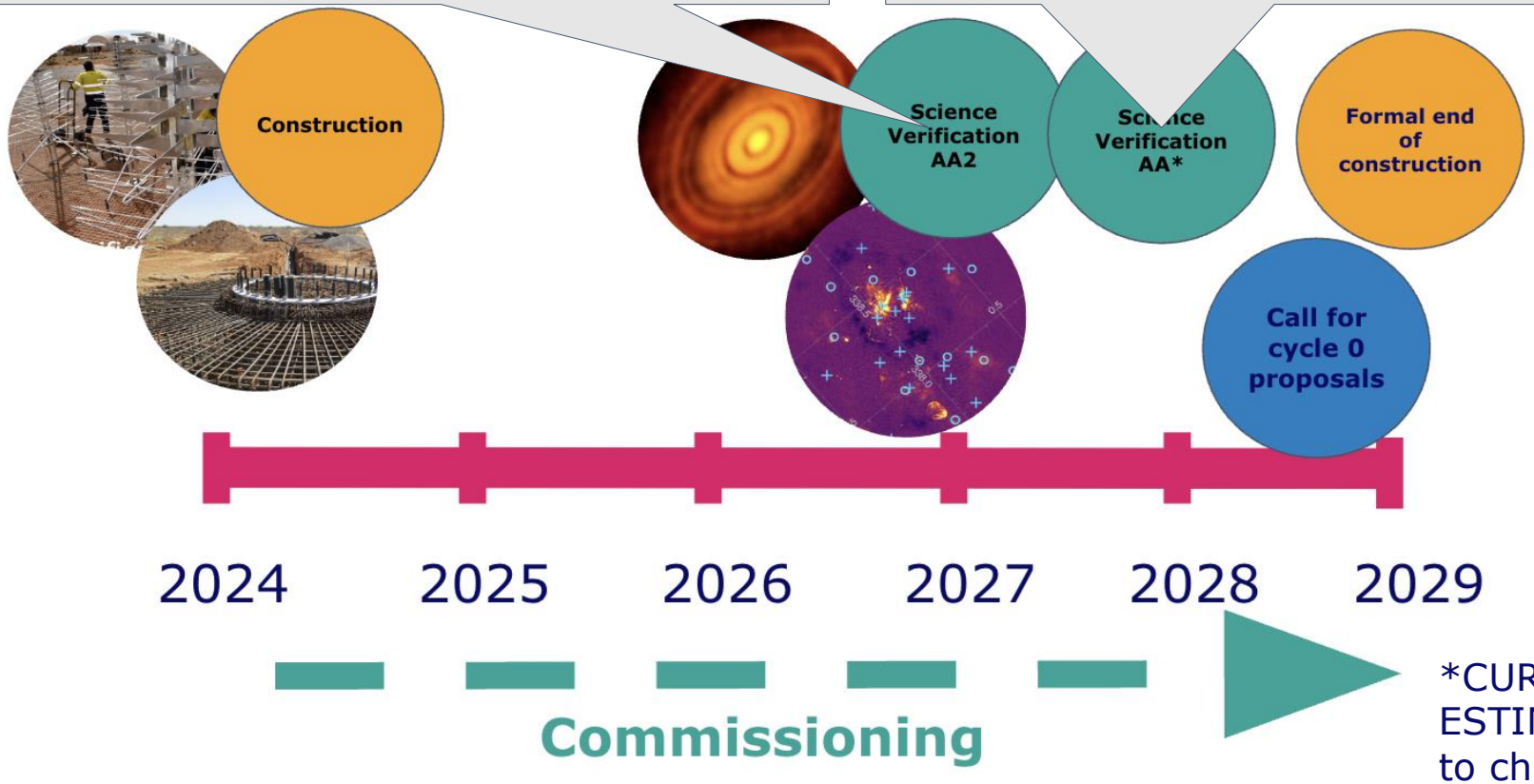
Auto-recovery if one site fails

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

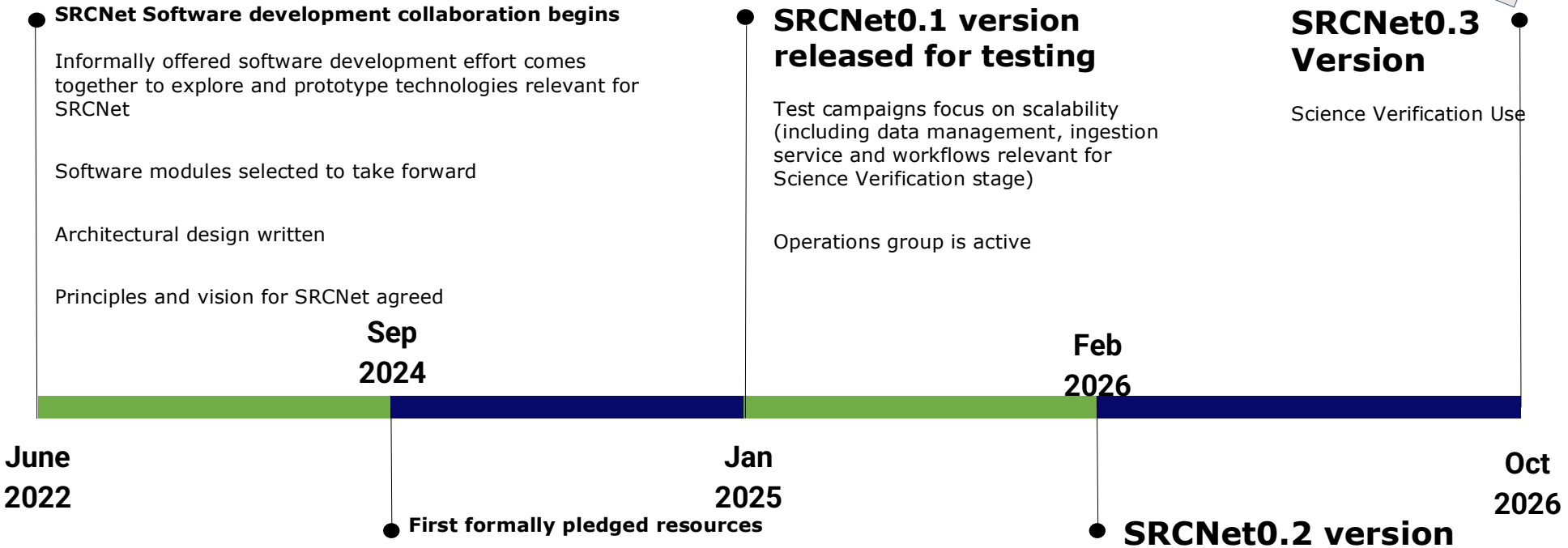
2027-2028 SV campaigns produce up to 14 PBytes* of data each SV week



*CURRENT ESTIMATES, subject to change

SRCNet timeline

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

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

SRCNet0.1 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

SRCNet0.2 version

Enhanced components

Further work preparing for AA2 and Science Verification

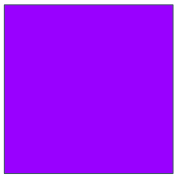
SRCNet0.3 Version

Science Verification Use

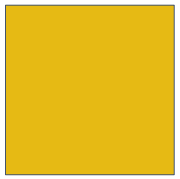
Top level roadmap: https://confluence.skatelescope.org/display/SNC/SRCNet+Documents?preview=/296945850/296945854/SRCNet%20Top%20Level%20Roadmap.pdf

The *current SRCNet* teams

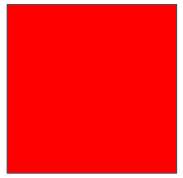

Coral


Purple


Lavender


Gold


Orange


Red


Magenta


Teal


Tangerine

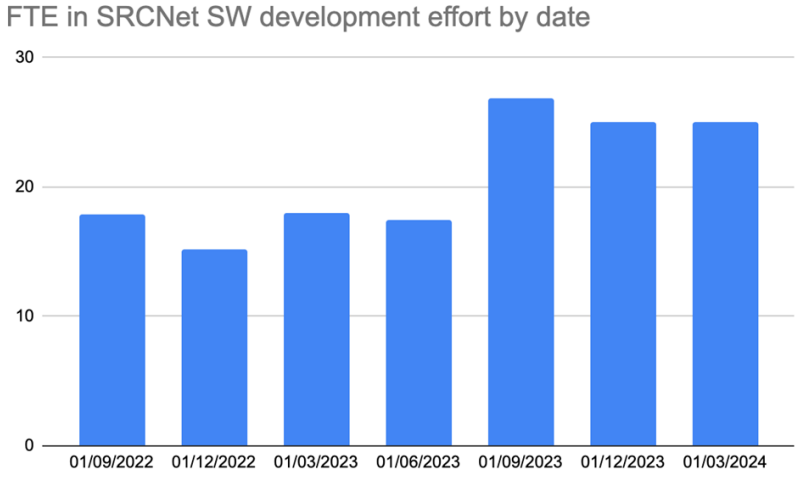

Chocolate

DAAC

Indigo

Since June 2022 we have been working as a team-of-teams
Engagement from across most SKA countries

SAFe framework Agile Release Train, same as for other SKAO SW development
Currently 46 FTE, from 13 countries plus SKAO, and 93 individuals

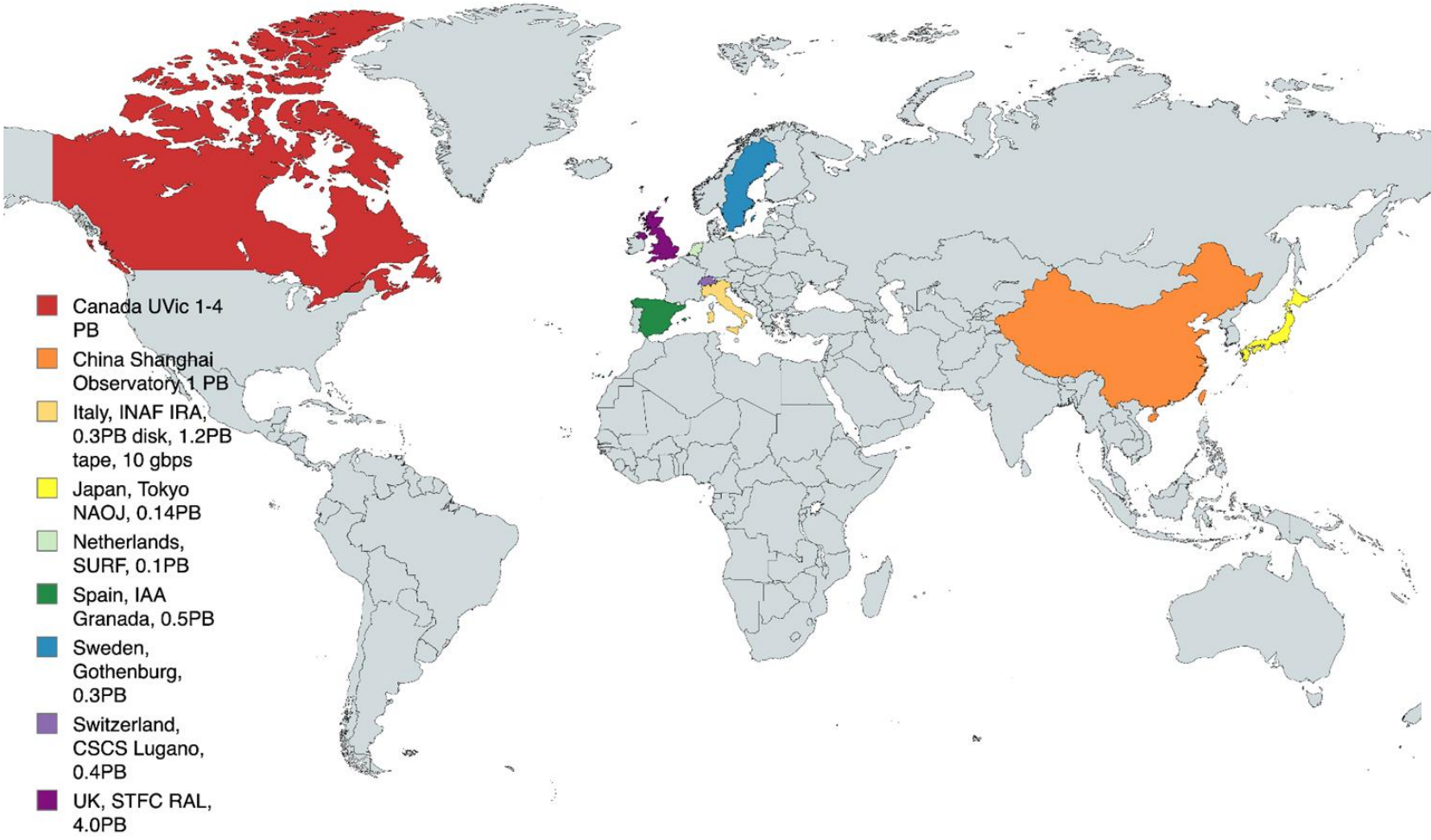


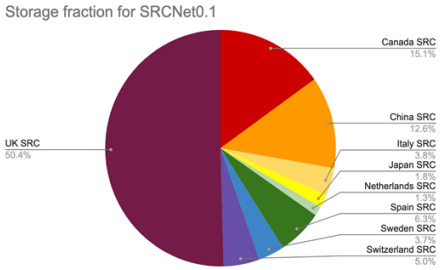
SRCNet0.1 planned sites

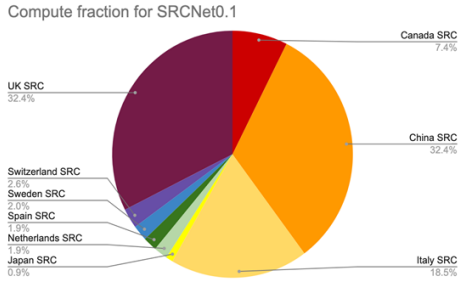
9 sites will contribute
d storage
o SRCNet0.1

total

PS Compute







SRCNet0.1

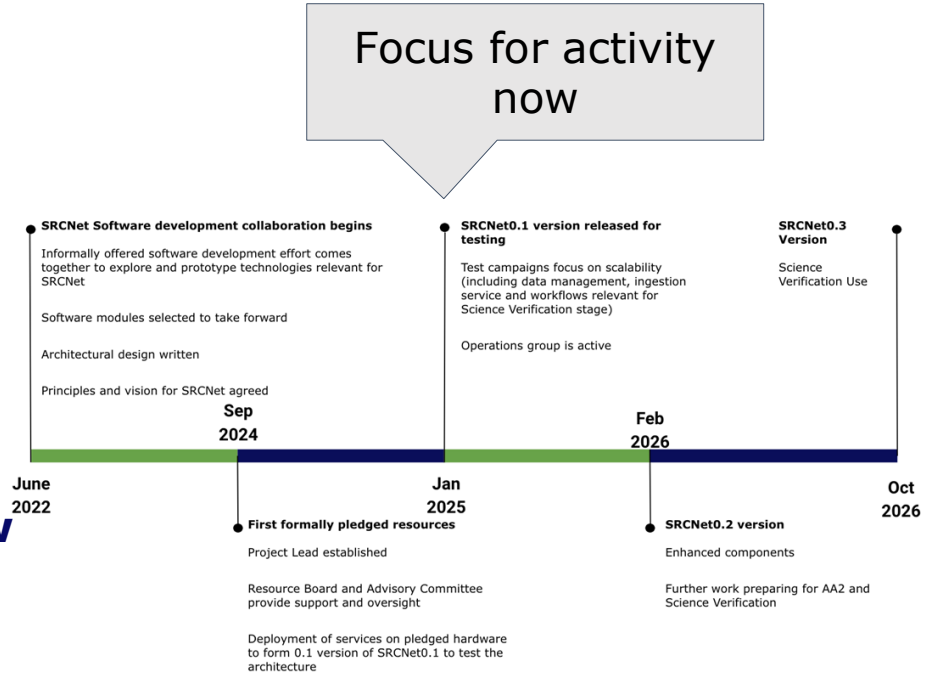
At least 4 sites running full set of compulsory local services; global services also running to support this

First test of full SRCNet architecture

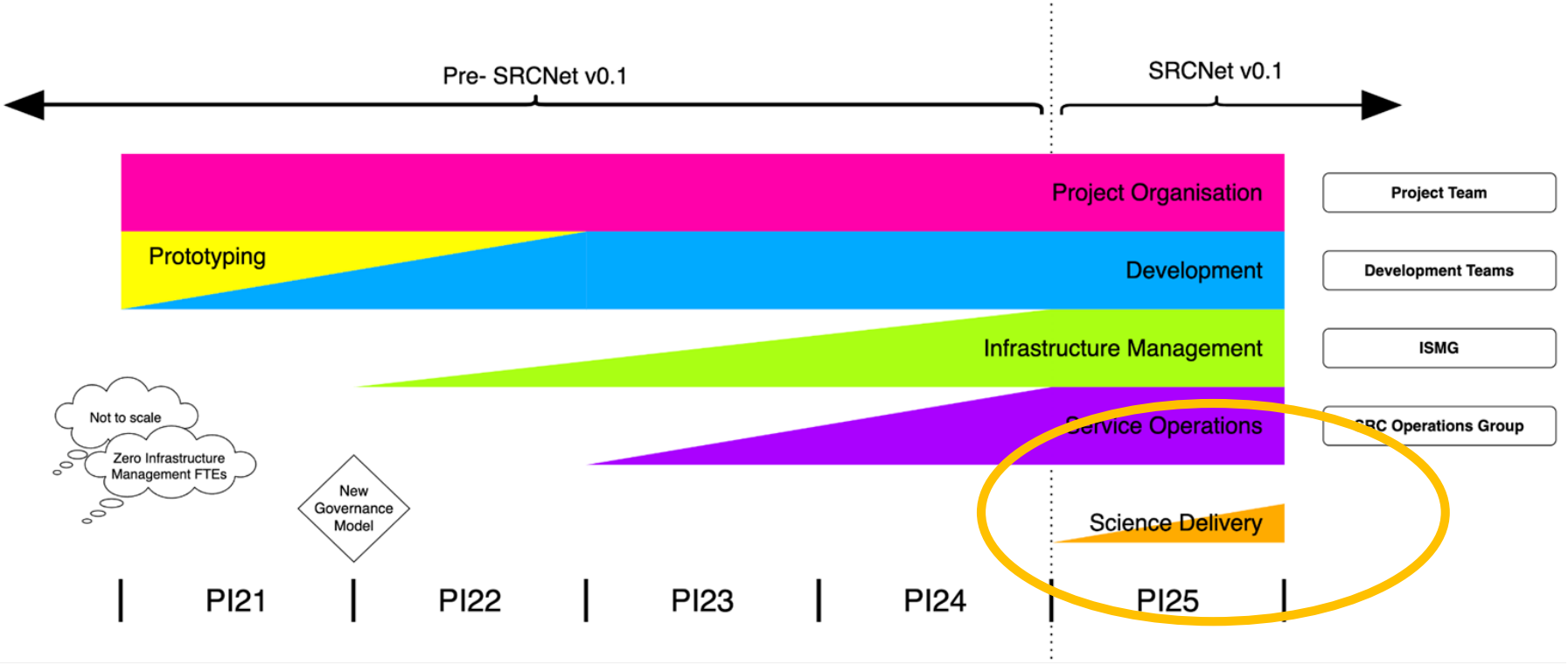
Due end of PI24; ie. 20th November 2024 (final work day of sprint 5)

This is an "engineering" version

- **Built to show the architecture and test how it works**
- **Internal only** - no user-facing activities
- **Learn how to deploy and operate the services**
- Exclusive storage to use in testing
- Compute to use during testing campaigns (may be backfilled when idle)



Value Stream Development - Science Delivery



Seeking Project Scientists to represent different science areas: Science Use Cases, Requirements.

Project Scientist Roles

We have asked SRCNet resource board members to offer committed people at 0.5 FTE to take on SRCNet roject Scientist roles. Their mandate will be to represent the needs of the science community in the work needed within the Science Delivery and development value streams (and other areas if needed). They will work closely with the Science Delivery Business Owner (Shari Breen).

We anticipate covering the following SKA scientific areas with these roles initially:

- Extragalactic
- Galactic
- Spectral Line
- Epoch of Reionisation
- Pulsars
- Transients (including use of transient buffer data)
- Solar Astronomy

SRCNet Science Delivery

SRCNet Project Scientists will work alongside the SKAO Project Scientists and with the SKA Science Operations group.

Science operations schedule is very important for understanding early SRCNet priorities

(we will prioritise Science Verification and early science)

We will let you know the appointed people (anticiapte having names by Friday 22nd, need to look at expertise, availability etc)