Science and Technology Facilities Council

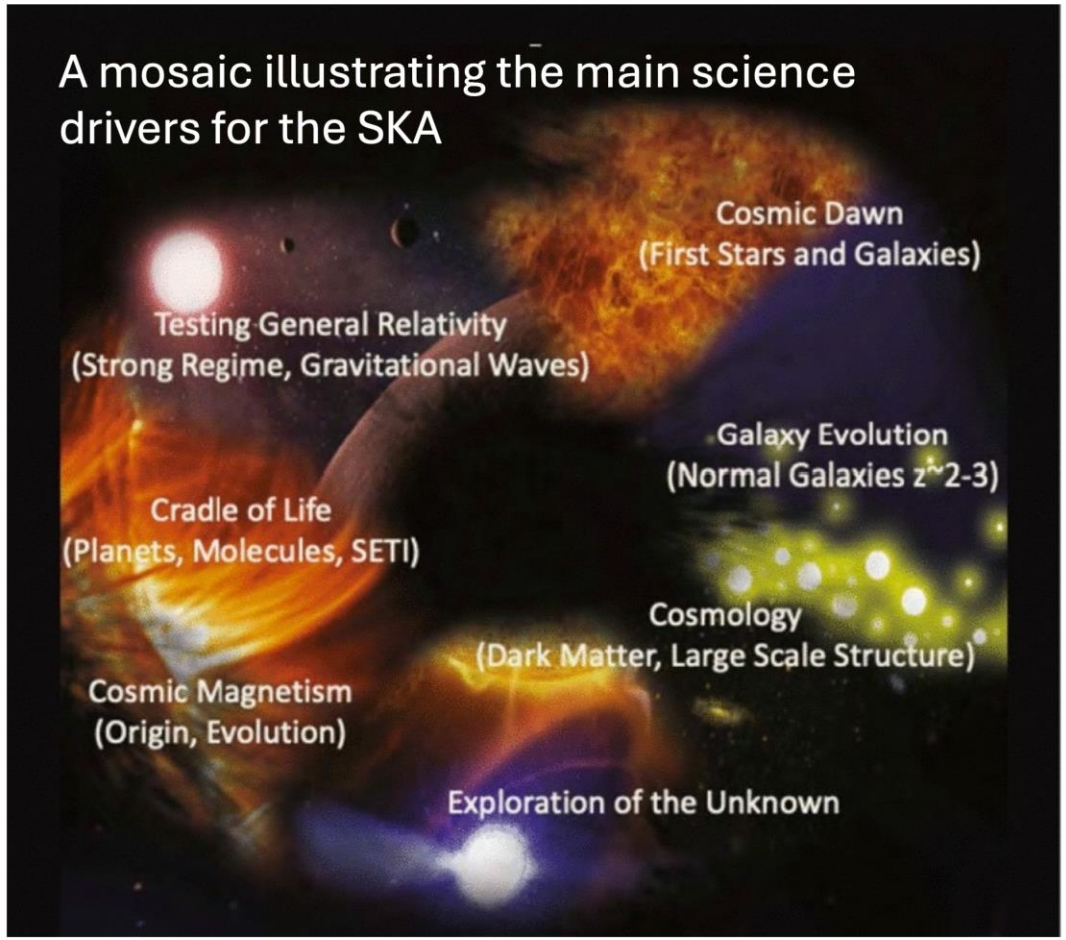
The path to exabyte astronomy: SRCNet v0.1 for the Square Kilometre Array

Ian Collier
James Walder
SRCNet members
24 October 2024, CHEP 2024, Krakow

Square Kilometer Array: Transforming radio astronomy

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.

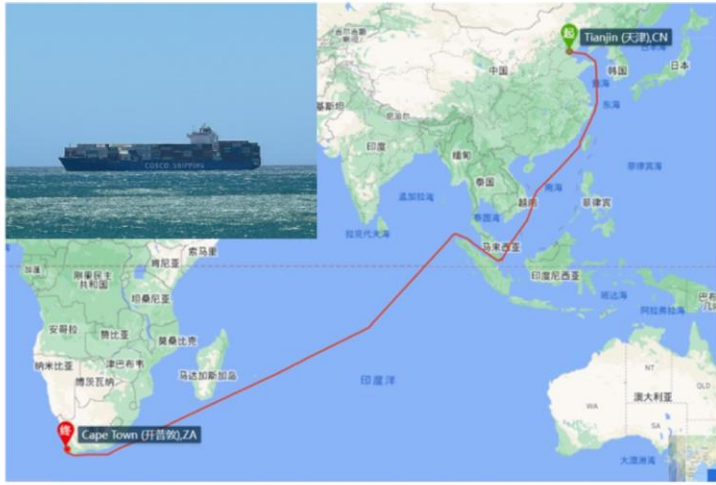
Composite image of the SKA telescopes, blending real hardware already on site with artist's impressions. credit: SKA Observatory

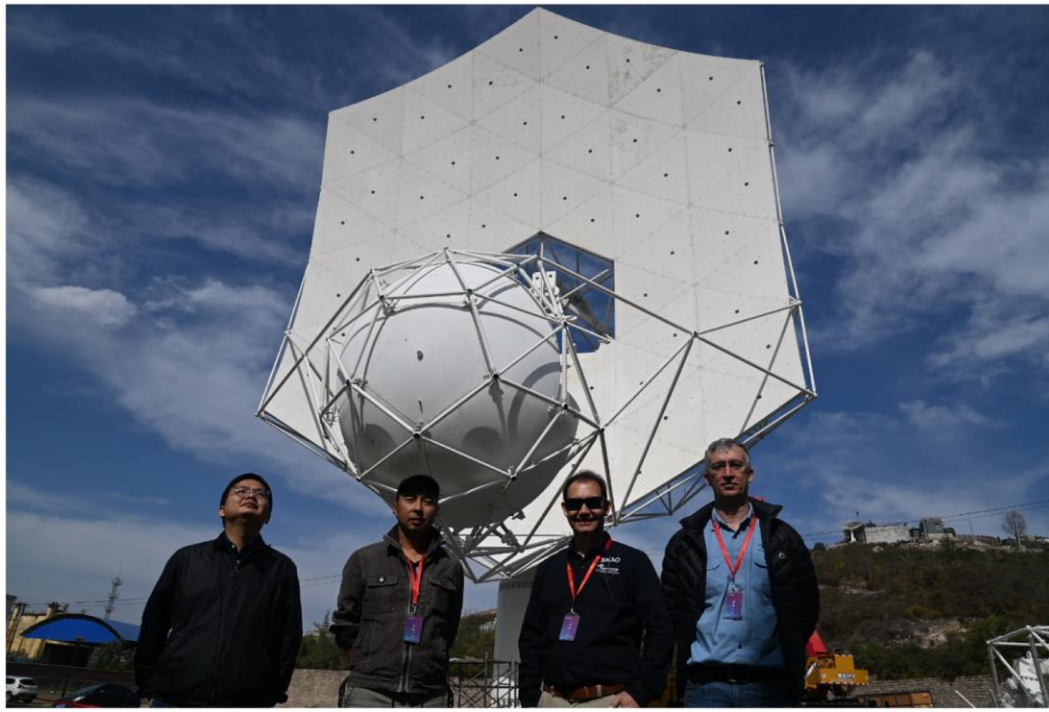


Credit: SKA Observatory

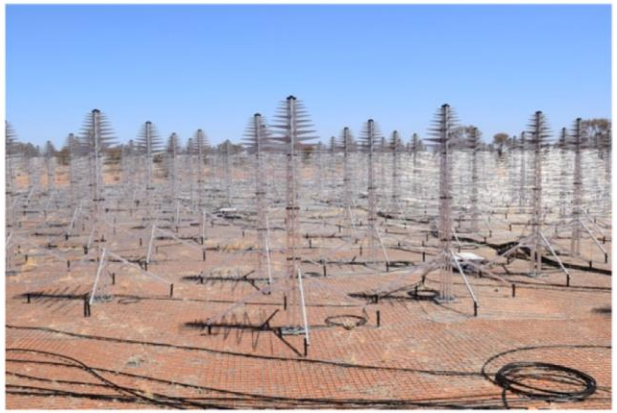
uk | SRC

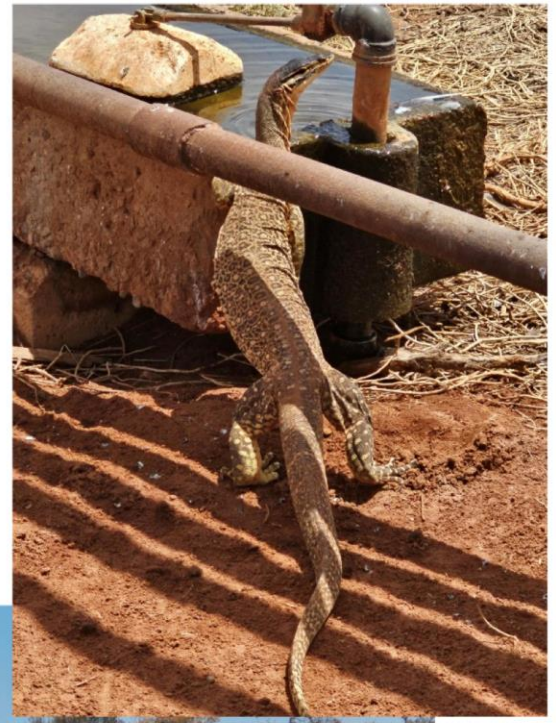
Construction steaming ahead! - Mid





Construction steaming ahead! - Low

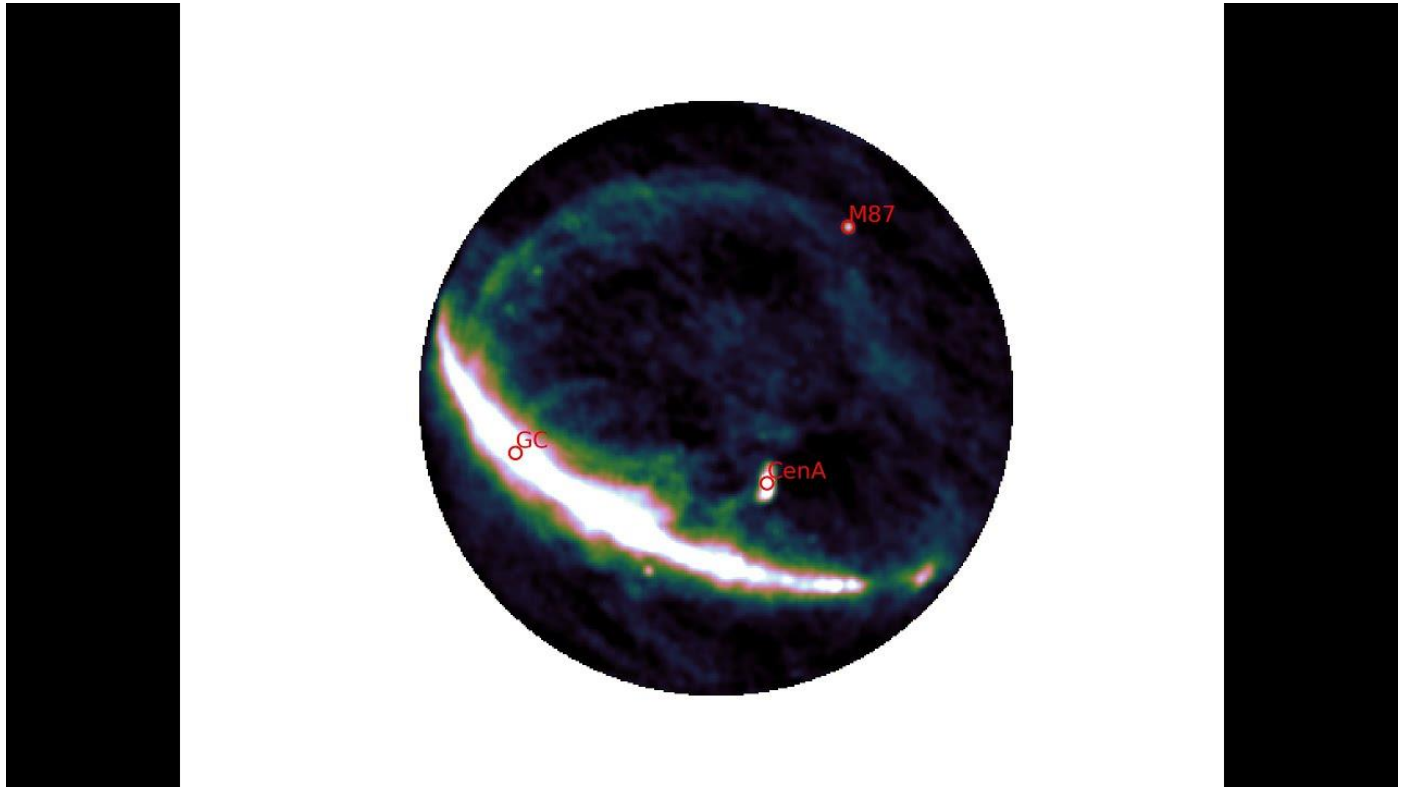




First image released from one SKA-Low station

- This is the first image and video from observations using **one complete SKA-Low station**, known as S-8, produced only **18 months after the start of construction** activities on site, and five months after the first antenna was installed.
- The completion of a station means not only assembling and installing the **256 antennas**, but also integrating them with all the computing systems behind them.
- The video shows a **24-hour observation**, with the Milky Way rising and passing overhead during the night time hours.

  Some other bright radio sources are marked, including the galaxies Centaurus A and M87, and the Sun is also visible during the day.



SKA Observatory
https://www.youtube.com/watch?v=zakuQ1-QrGg

Science and Technology Facilities Council

Construction Strategy

- **Target**: build the SKA Baseline Design (197 Mid dishes; 512 Low stations: AA4)
- Not all funding yet secured, therefore following Staged Delivery Plan (AA*)
- Develop the earliest possible working demonstration of the architecture and supply chain (AA0.5).
- Then maintain a continuously working and expanding facility that demonstrates the full performance capabilities of the SKA Design.

Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 Apr	2025 Nov
AA2	64 dishes 64 stations	2027 Mar	2026 Oct
AA*	144 dishes 307 stations	2027 Dec	2028 Jan
Operations Readiness Review		2028 Apr	2028 Apr
AA4	197 dishes 512 stations	TBD	TBD

First data release to the community expected in 2026/27 (for science verification)

What does this mean in terms of Operations?

Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 Apr	2025 Nov
AA2	64 dishes 64 stations	2027 Mar	2026 Oct
AA*	144 dishes 307 stations	2027 Dec	2028 Jan
Operations Readiness Review		2028 Apr	2028 Apr
AA4	197 dishes 512 stations	TBD	TBD

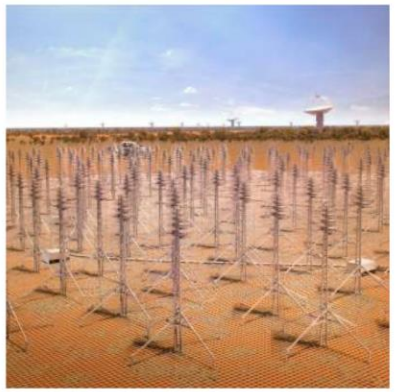
Pre science Verification
- Commissioning (+ Assembly, Integration and Verification) primary activity
- SRCs not needed to support AA0.5/AA1 commissioning
- Opportunity for testing (data, transfer, access, pipelines)!

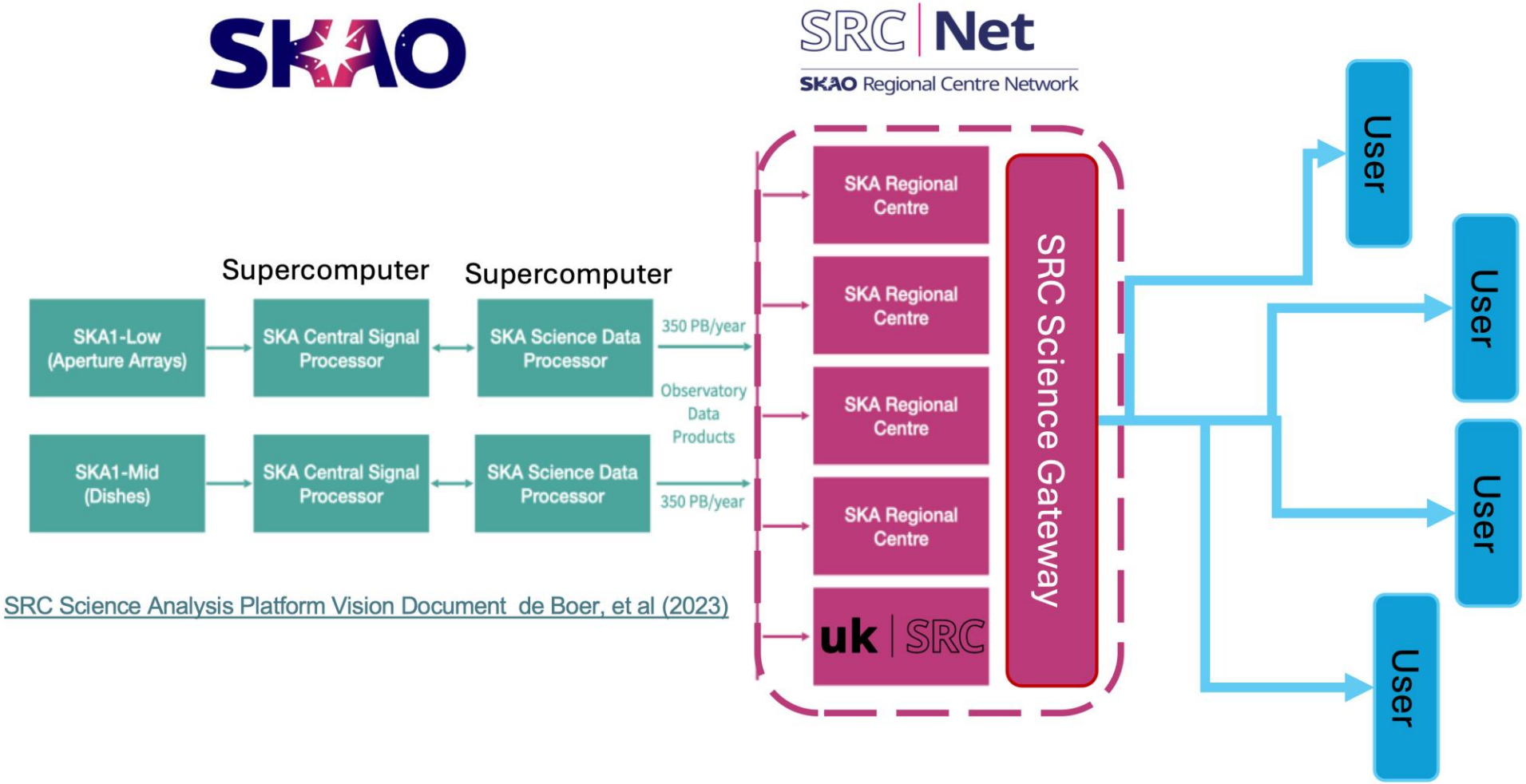
Science Verification
- Data immediately public
- Full dress rehearsal!
- Some SRCNet resources for analysis would be an advantage
- Observed as trickle but also in dedicated blocks
- (+ Commissioning etc ongoing)

Cycle 0
- "Proper" shared risk projects
- Teams, proprietary periods, visualisation, ADP creation etc

SKA Regional Centre Network

SRCNet will provide a portal for scientists to access SKA data – an exabyte data challenge!





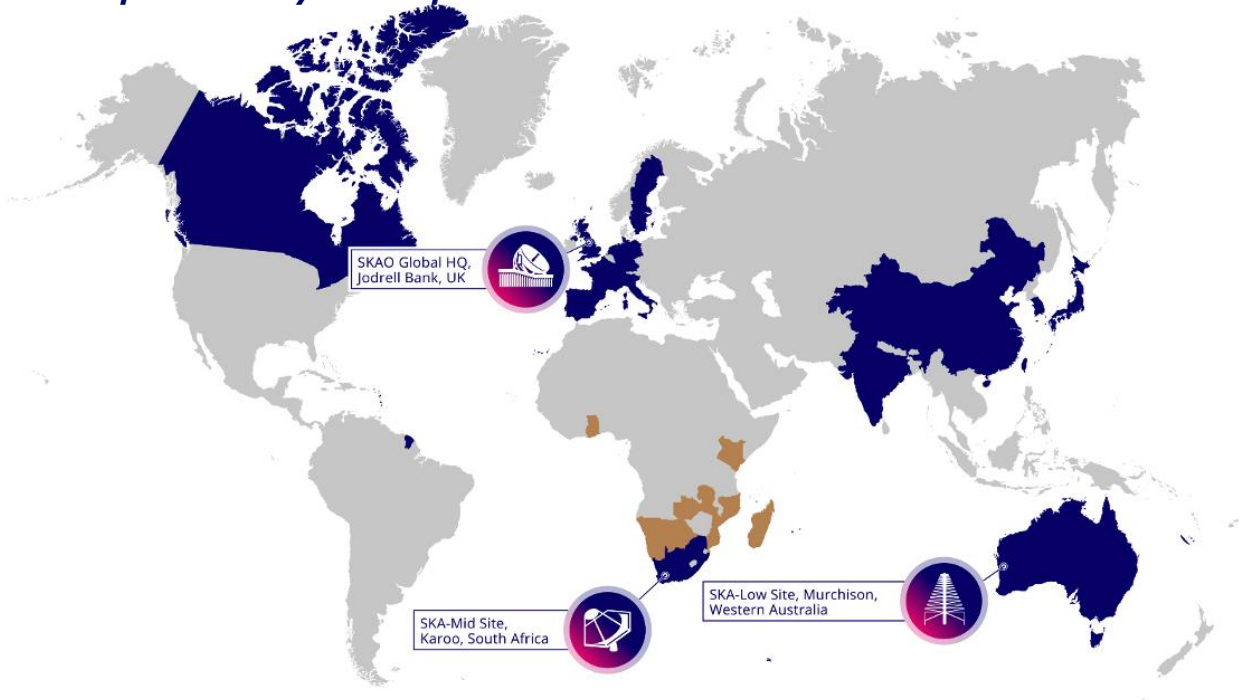
SRC Science Analysis Platform Vision Document de Boer, et al (2023)

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

The SRC Network

- The need for a network of SKA Regional Centres formed around ~ 2016:
  - Distributed compute, storage and expertise to store, process and disseminate data to the communities

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.***



A Naive mapping between LHC and SKA

Don't take too literally

LHC/WLCG	SRCNet
Cern + Experiements	~ SKAO + SKA-MID + SKA-LOW
WLCG	~SRCNet
GridPP	~UKSRC

SRC Network global capabilities

Collectively meet the needs of the global community of SKA users

Anticipate heterogeneous SRCs, with different strengths

SRC Network is critical to SKA Science

Science Analysis Platform

Creation of Project-Level and Advanced Data Products

Public Science Archive

Execute pipelines/workflows on the SRC Network

Access to SKAO data for project users

Single Sign-on and Secure, Federated AAI

Monitoring of performance against pledges

Monitor SRC availability

User support and community training

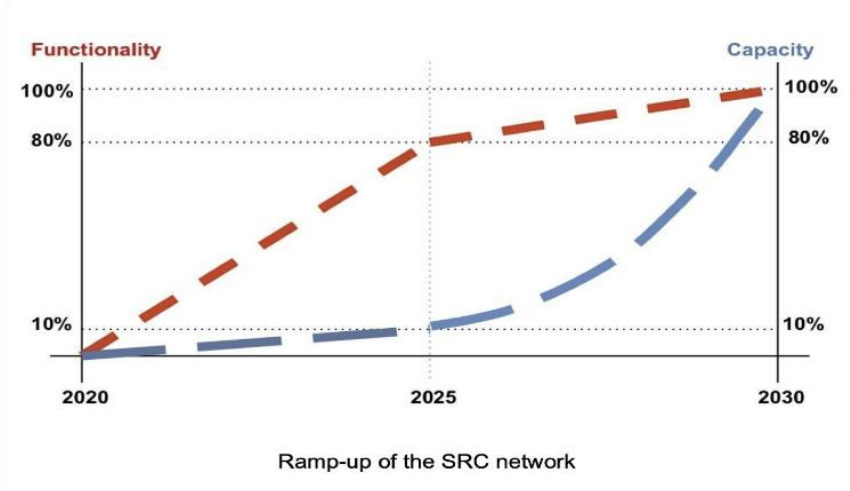
Replica management

Receive data from SKAO

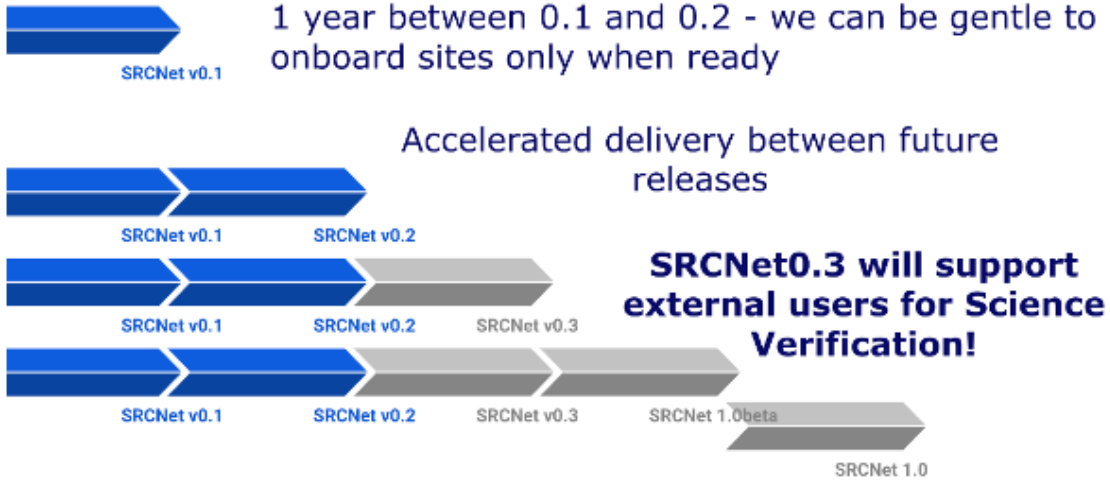
SRCNet Timeline

- SRCNet timeline as mapped to the construction roadmap.
  - Increased capabilities; then scale out

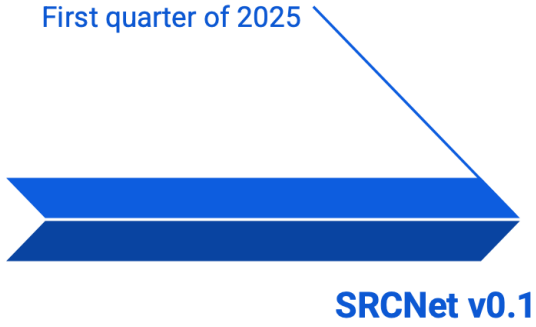
Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 Apr	2025 Nov
AA2	64 dishes 64 stations	2027 Mar	2026 Oct
AA*	144 dishes 307 stations	2027 Dec	2028 Jan
Operations Readiness Review		2028 Apr	2028 Apr
AA4	197 dishes 512 stations	TBD	TBD



Ramp-up of the SRC network



Roadmap Timeline

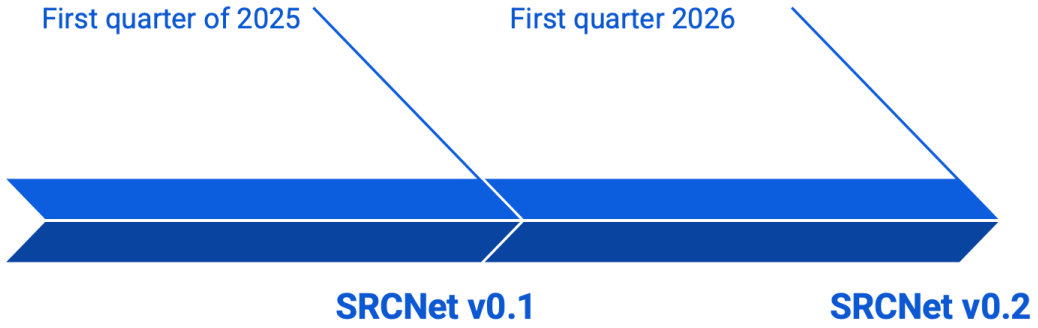


SRCNet0.1 is an internal release
Not intended for external users
Motivation is to enable testing

SRCNet0.1 is an agreed milestone (first of five) on our top level roadmap

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] (Potentialy Opportunity to engage SRCNet with AA0.5 data transfer and access.)	• 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)**

Roadmap Timeline



Not generally public
Small amount of science commissioning interaction
Most SRCNet users are within the project or SKAO

Milestone	Description	SRC Net Functionality	Scope (users)
SRCNet v0.2 First quarter 2026	AA1 and Commissioning	• **Data dissemination using telescopes sites interface** • First version of federated execution. Access to remote operations on data using services and the possibility to invoke execution into a relevant SRC • Subset of SDP workflows runnable in the SRCs • First Accounting model implementation. • User storage areas • Visualisation of imaging and time series data through remote operations • Preparation of SRCNet User Support	Selected scientists from community Members of Science Operations SRC ART members

SKA expected data rates*

***these numbers should be used as a guide only - email Shari.Breen@skao.int for further information about ongoing work**

- Numbers refer to data to be delivered to the science community via the SRCNet (i.e. not data used internally for commissioning etc.)

Milestone	Year	Primary activity	Estimated data rate	
			Low	Mid
AA2 • 64 Mid dishes • 64 Low stations	2026 - 2027	**Science Verification** - observed in dedicated ~week long blocks + single observations interspersed throughout. A higher rate of raw data products will be included at this stage.	1.5 PB/week^ 20 Gbps	2 PB/week^ 27 Gbps
AA* • 144 Mid dishes • 307 Low stations	2027 - 2029	**Science Verification** - observed in dedicated ~week long blocks + single observations interspersed throughout. A higher rate of raw data products will be included at this stage.	5 PB/week^ 66 Gbps	9 PB/week^ 119 Gbps
AA* • 144 Mid dishes • 307 Low stations	2029 +	**Operations** - Observation cycles, starting with shared risk observing, building to successful science observations ~90% of the time	173 PB/year 44 Gbps	280 PB/year 72 Gbps
Target is to deliver the SKA Baseline Design but the details of this transition between AA* and AA4 are TBD				
AA4 • 197 Mid dishes • 512 Low stations	2030 +	**Operations** - full SKA baseline design	216 PB/year 55 Gbps	400 PB/year 100 Gbps

^Data rates refer to dedicated Science Verification observing weeks, not an average over a year

SRCNet Top-Level Roadmap Resource Requirements (2023)

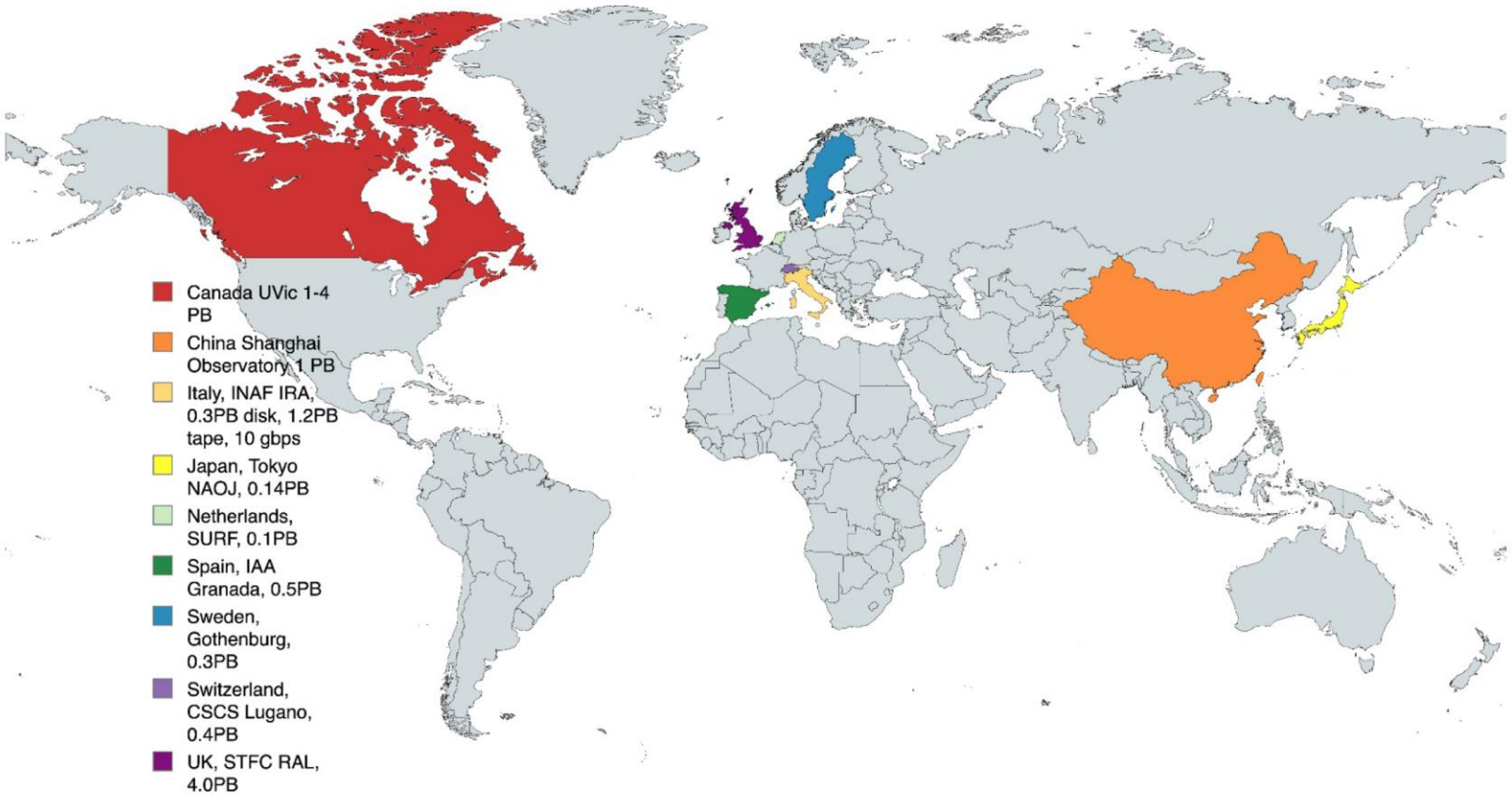
		SRCNet v0.1	SRCNet v0.2	SRCNet v0.3	SRCNet v1.0b	SRCNet v1.0
		Jan 2025	January 2026	Sep 2026	Nov 2027	Jun 2028
Deployment (%)		2.00	10.00	15.00	50.00	100.00
Country	Share (%)	Storage (PB)	Storage (PB)	Storage (PB)	Storage (PB)	Storage (PB)
UK	19	4.03	20.14	30.21	100.70	201.40
South Africa	18	3.82	19.08	28.62	95.40	190.80
Australia	18	3.82	19.08	28.62	95.40	190.80
China	10	2.12	10.60	15.90	53.00	106.00
Canada	7	1.48	7.42	11.13	37.10	74.20
Italy	6	1.27	6.36	9.54	31.80	63.60
India	5	1.06	5.30	7.95	26.50	53.00
France	3	0.64	3.18	4.77	15.90	31.80
Netherlands	2	0.42	2.12	3.18	10.60	21.20
Japan	2	0.42	2.12	3.18	10.60	21.20
Spain	2	0.42	2.12	3.18	10.60	21.20
Portugal	2	0.42	2.12	3.18	10.60	21.20
Switzerland	2	0.42	2.12	3.18	10.60	21.20
Sweden	2	0.42	2.12	3.18	10.60	21.20
South Korea	1	0.21	1.06	1.59	5.30	10.60
Germany	1	0.21	1.06	1.59	5.30	10.60
Total	100	21.20	106.00	159.00	530.00	1060.00

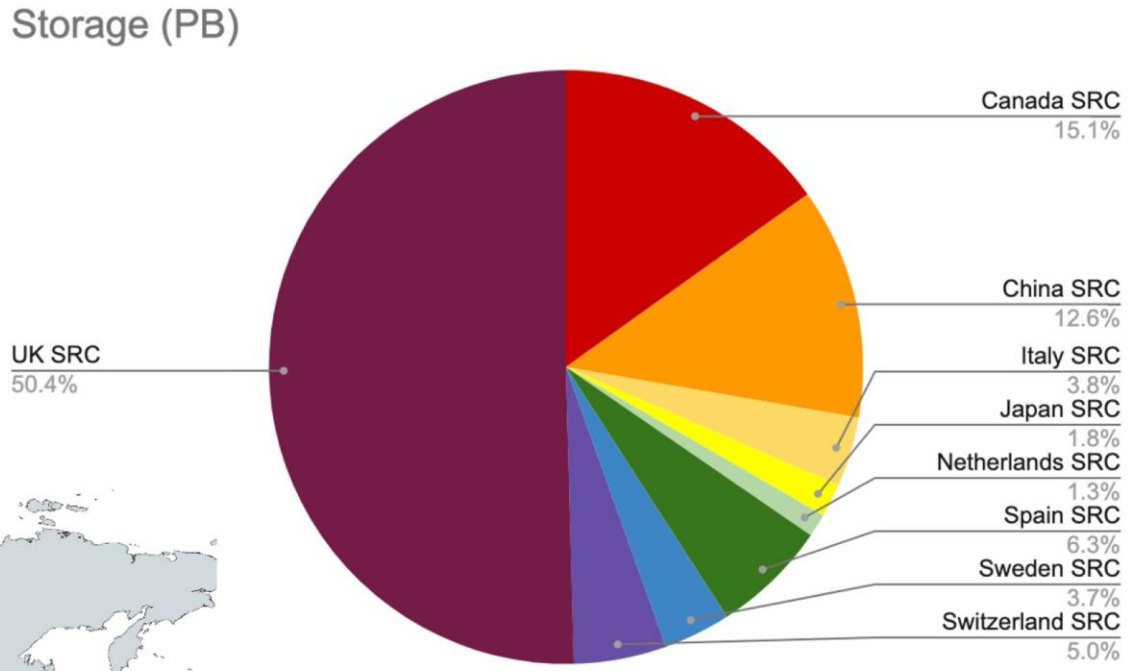
		SRCNet v0.1	SRCNet v0.2	SRCNet v0.3	SRCNet v1.0b	SRCNet v1.0
		Jan 2025	January 2026	Sep 2026	Nov 2027	Jun 2028
Deployment (%)		2.00	10.00	15.00	50.00	100.00
Country	Share (%)	Computing (PFLOPS)	Computing (PFLOPS)	Computing (PFLOPS)	Computing (PFLOPS)	Computing (PFLOPS)
UK	19	0.13	0.67	1.00	3.33	6.65
South Africa	18	0.13	0.63	0.95	3.15	6.30
Australia	18	0.13	0.63	0.95	3.15	6.30
China	10	0.07	0.35	0.53	1.75	3.50
Canada	7	0.05	0.25	0.37	1.23	2.45
Italy	6	0.04	0.21	0.32	1.05	2.10
India	5	0.04	0.18	0.26	0.88	1.75
France	3	0.02	0.11	0.16	0.53	1.05
Netherlands	2	0.01	0.07	0.11	0.35	0.70
Japan	2	0.01	0.07	0.11	0.35	0.70
Spain	2	0.01	0.07	0.11	0.35	0.70
Portugal	2	0.01	0.07	0.11	0.35	0.70
Switzerland	2	0.01	0.07	0.11	0.35	0.70
Sweden	2	0.01	0.07	0.11	0.35	0.70
South Korea	1	0.01	0.04	0.05	0.18	0.35
Germany	1	0.01	0.04	0.05	0.18	0.35
Total	100	0.70	3.50	5.25	17.50	35.00

** Specific numbers out of date – but the shares and the broad ramp up remain instructive*

SRCNet0.1 included sites

8 PBytes total storage offered for SRCNet0.1 (c.f original target of 20 PB)

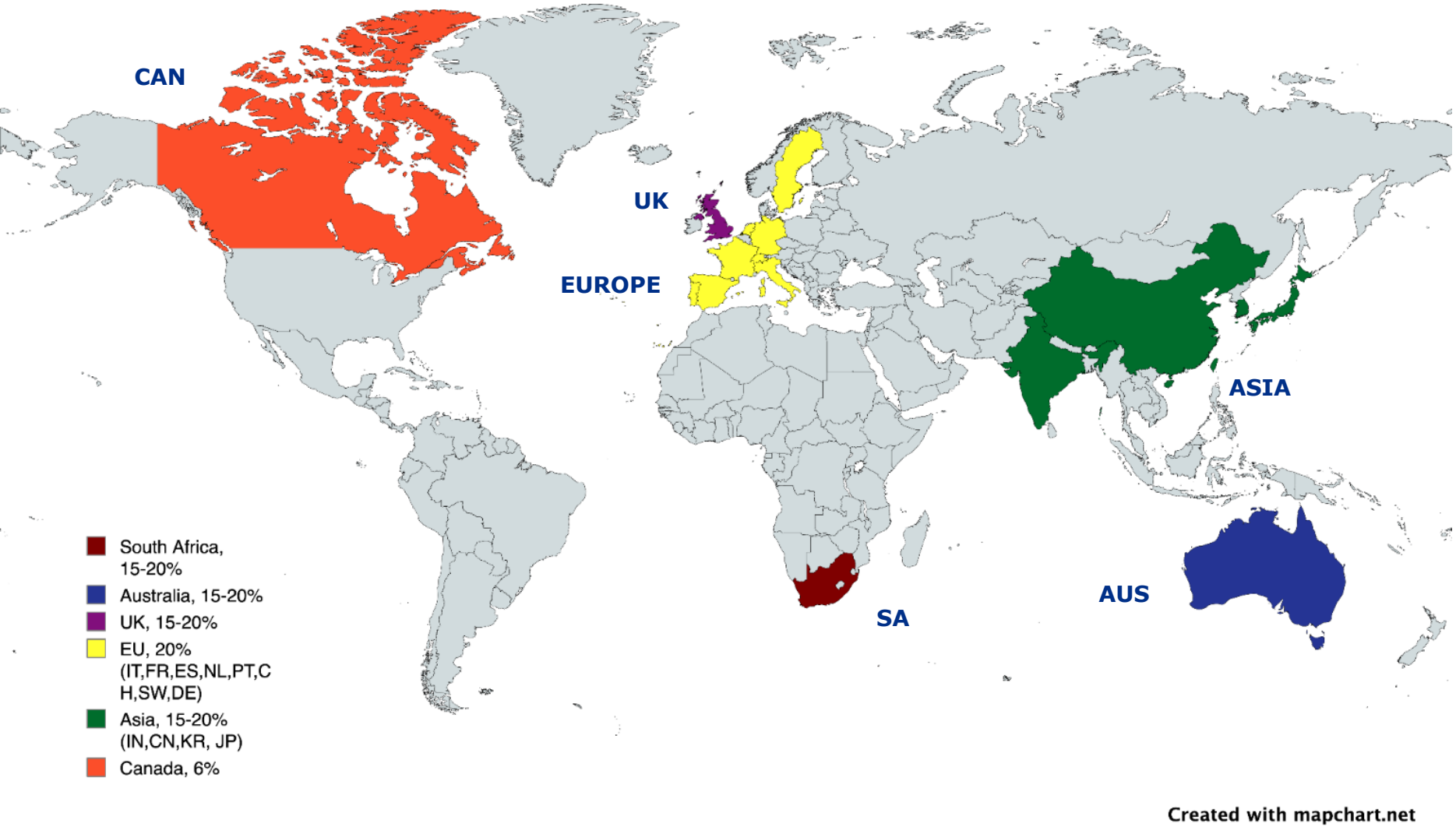




WLCG experience at some sites (Canada, Netherlands, Sweden, Switzerland, UK)

Several new sites and teams will learn by being involved

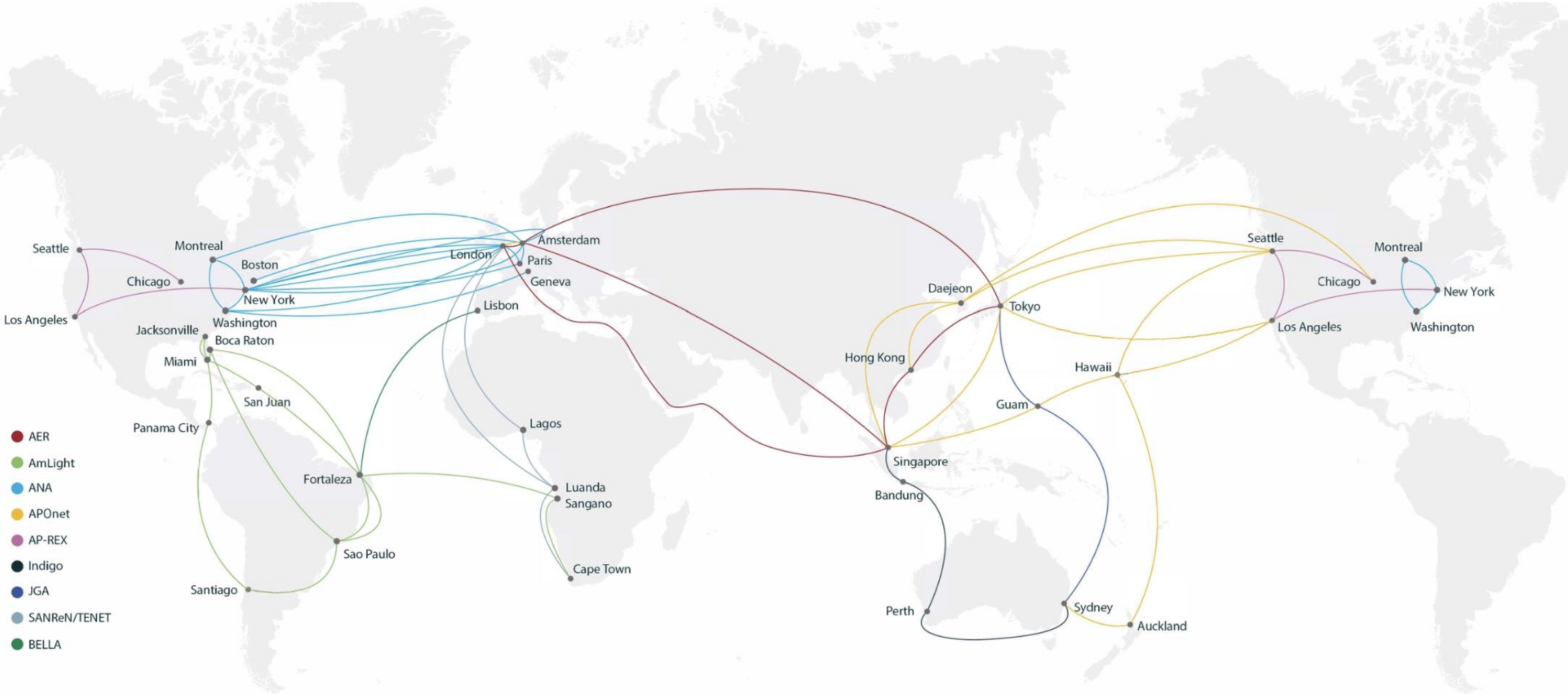
SKA Regional Centre Broad Distribution: Fair Share, AA4 data rates



- Roughly, 6 global zones of equivalent size (Canada smaller)
- **Distribute two base copies** of each data product to different regional centres
- Average incoming rate per (20%) region not more than 2x40 Gbit/s = 80Gbit/s (~2x12 Gbit/s for Canada)
- **Average 100 Gbit/s out of SA and AUS**
- **Actual bandwidth available much greater**

e.g. if average 100+100 gbps from each of mid and low, a 10% partner receives 40gbps data (400 TBytes per day, 140 PBytes per year)

The Global Research & Education Network



Science and Technology Facilities Council

With thanks to Lars Fischer from Nordunet

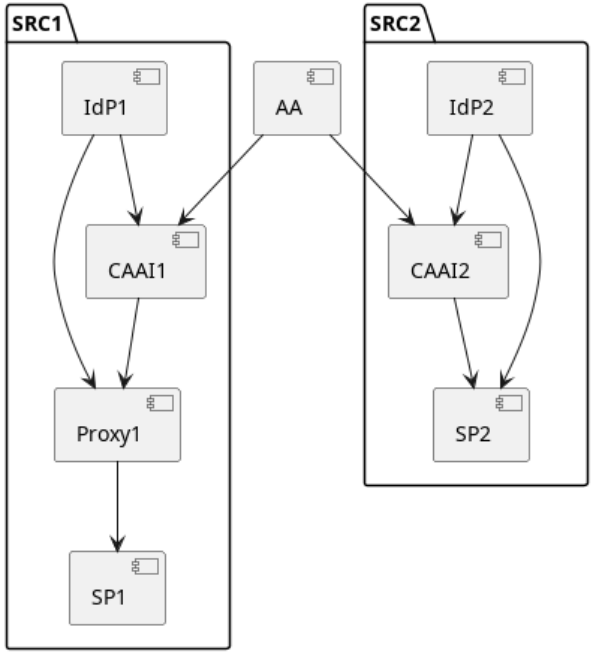
More on the network

- Network expected to ingest ~250PB/year from each of mid and low - ~500PB/year
  - Average ~100GB/s from each
  - Then gets copied to second regional centre
  - Comparable to one HL-LHC experiment
- SA (to London) and Australia (to Singapore) have theoretical capacity up to many TB/s
  - Utilization must be paid for, but capacity is in place
- Overall the Global R&E Networks can cope
- Specific places may need further investment
- Some SRCs are in countries with good connections but may be at campuses that require upgraded connections

- Tony Cass mentioned that SRCNet has an NREN forum
  - This is focused on technical work
- There is a new NREN Strategy group being convened to provide assurance on the global level capacity questions.
- Also talking to WLCG and other communities
  - SKAO will host next LHCOPN-LHCONE meeting in Manchester early next year
- SKA benefits hugely from the collaboration between WLCG and NRENs over the years

AAI Design & Data Management and Movement

Two particular areas where SRCNet is benefiting from WLCG's experience



- **AAI Design & prototyping**
  - UK has been running a dedicated Indigo IAM instance for SRCNet prototyping work for two years
  - Also working on token based federated AAI design to meet final requirements SRCNet
    - **Do not** want single central token issuer (resilience and latency concerns)
    - Central attribute authority (manage group memberships etc) run by SKAO
      - Also needed for eg telescope time etc.
    - Network of proxies (one in most regional centres) to provide resilience and performance
  - Identifying development requirements

- **Distributed Data Management**
  - See Rose Cooper's talk in Track 1
  - "FTS as a part of the SKA data movement pipeline"
    - **https://indico.cern.ch/event/1338689/contributions/6010770/**

Science and Technology Facilities Council

Summary

Moving to first operational phase

- Staged deployment of both instrument and global compute infrastructure
- Establishing infrastructure management and operations teams across partner sites
- Procuring and deploying compute and storage resources
- Establishing security policies etc.
- Development work moving on to next phase

Looking forward to

- Establishment of technical capabilities at SRCNet nodes and across network
- New formal technical roadmap flowing from SKAO's revised deployment roadmap
  - We know 2023 technical roadmap is out of date
- Coordination with other communities
  - For example, during upcoming WLCG data challenges
  - Much to learn and contribute on network management/monitoring

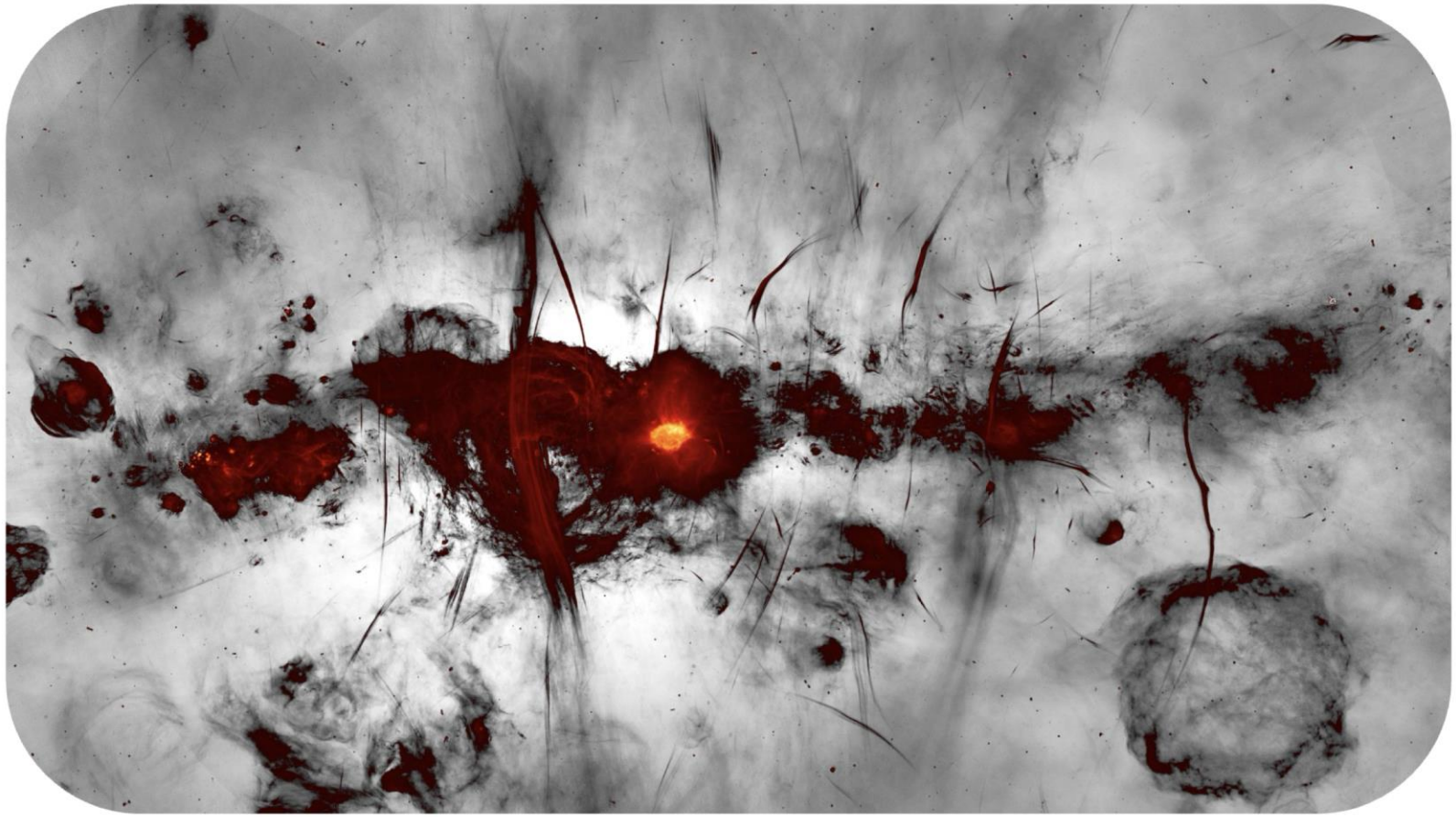
Science and Technology Facilities Council

Science and Technology Facilities Council

Questions

See also James Walder's talk at 16:15 today in Track 7 on the UK role in the deployment of SRCNet
https://indico.cern.ch/event/1338689/contributions/6011575/

We will deliver data products!



Credit: I. Heywood, SARAO

- Our data are BIG, expecting to deliver ~700 PB/year of *data products*
- Don’t need to be a radio expert to access the SKA!
- Transformational science increasingly relies on multiwavelength data, everyone with great science is welcome :)