



REGIONAL
CENTRE
NETWORK

SRCNet project update

Rosie Bolton
rosie.bolton@skao.int

Head of Data Operations

09/04/2024



Summary - the SRCNet Project

- **Vision**
- Roadmap
- Architecture
- The team of teams
- Implementation of SRCNet0.1
 - Locations
 - Potential implications for network



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.**



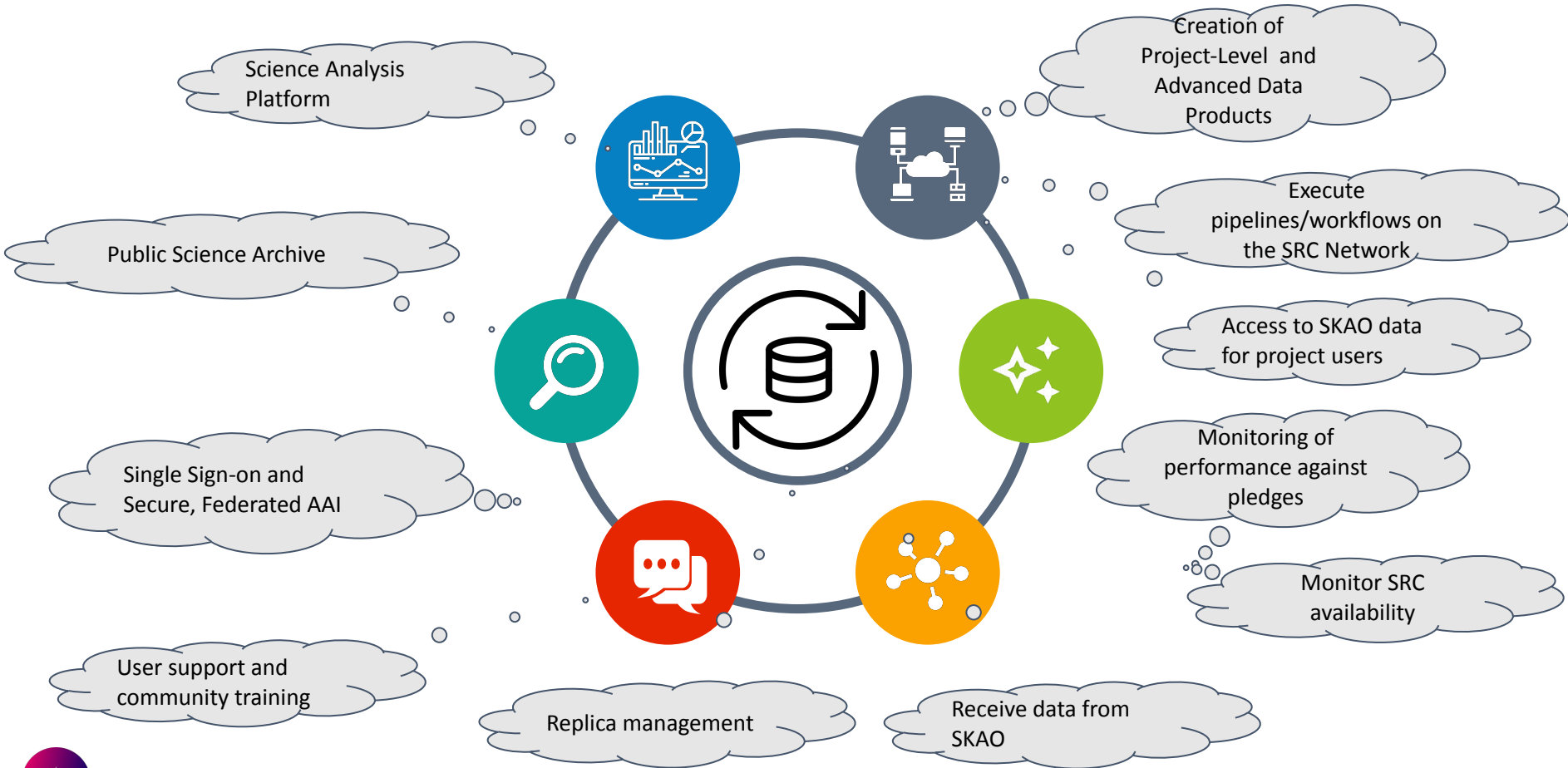
SRC Network Vision

Initially, we will do this by:

- developing a scalable, prototype SRC Network that allows authorised users and teams to access and analyse SKA data;
- developing the software, architecture, policies and processes necessary for SRC Network operations;
- growing the prototype SRC Network, as new SRCs become available and expanded or new functionalities are developed, leading towards a fully operational and global Network.



SRC Network is critical to SKA Science



SRC Network Principles (some of them!) - written and agreed by SRCSC

- 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.**



SRC Network Principles (highlighted for SRCNet0.1, From Dec 2024)

- There will be a **common SKAO/SRC Network user account** that allows users access to SRC Network resources
- Security of the SRC Network is the responsibility of the SRC Network.
- The SRC Network will be committed to providing, and abiding to, **accessible and equitable tools, practices and processes.**
- 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 **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.



SRC Network Principles (highlighted for SRCNet0.1, From Dec 2024)

- 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.

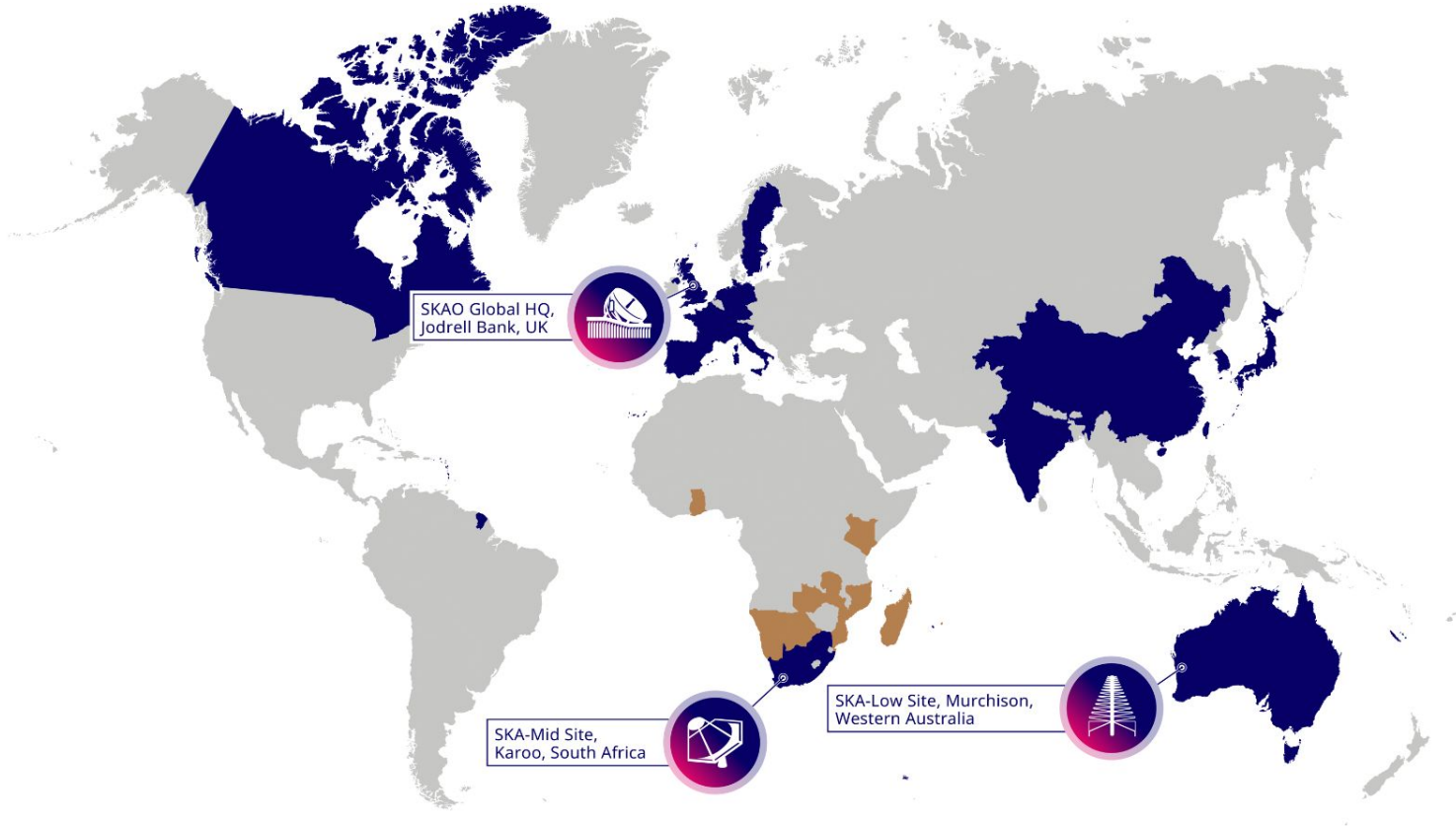
Single AAI System used by all SRCNet0.1 sites and services

- 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.

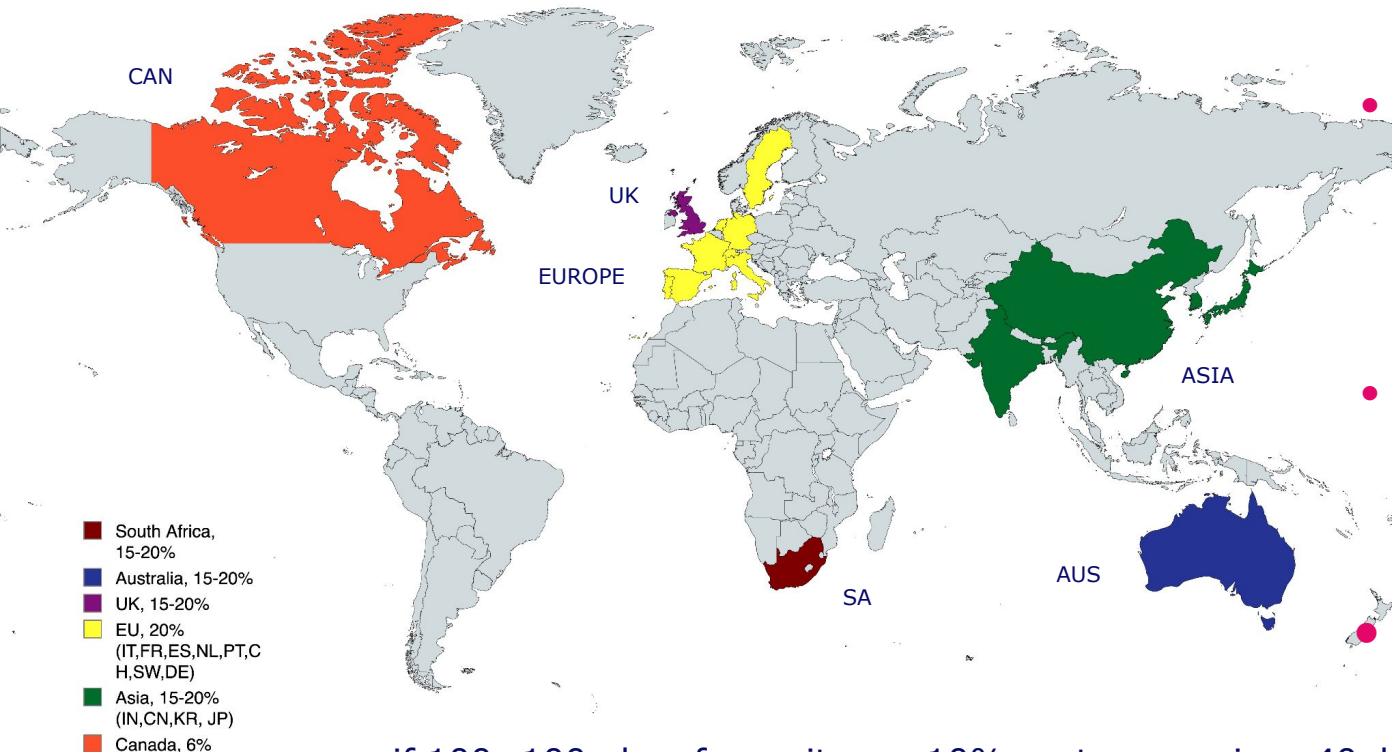




- SKAO Partnership - includes SKAO Member States* and SKAO Observers (as of April 2023)
-  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *
 -  *

- African Partner Countries
- 
 - 
 - 
 - 
 - 
 - 
 - 
 - 
 - 
 - 

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

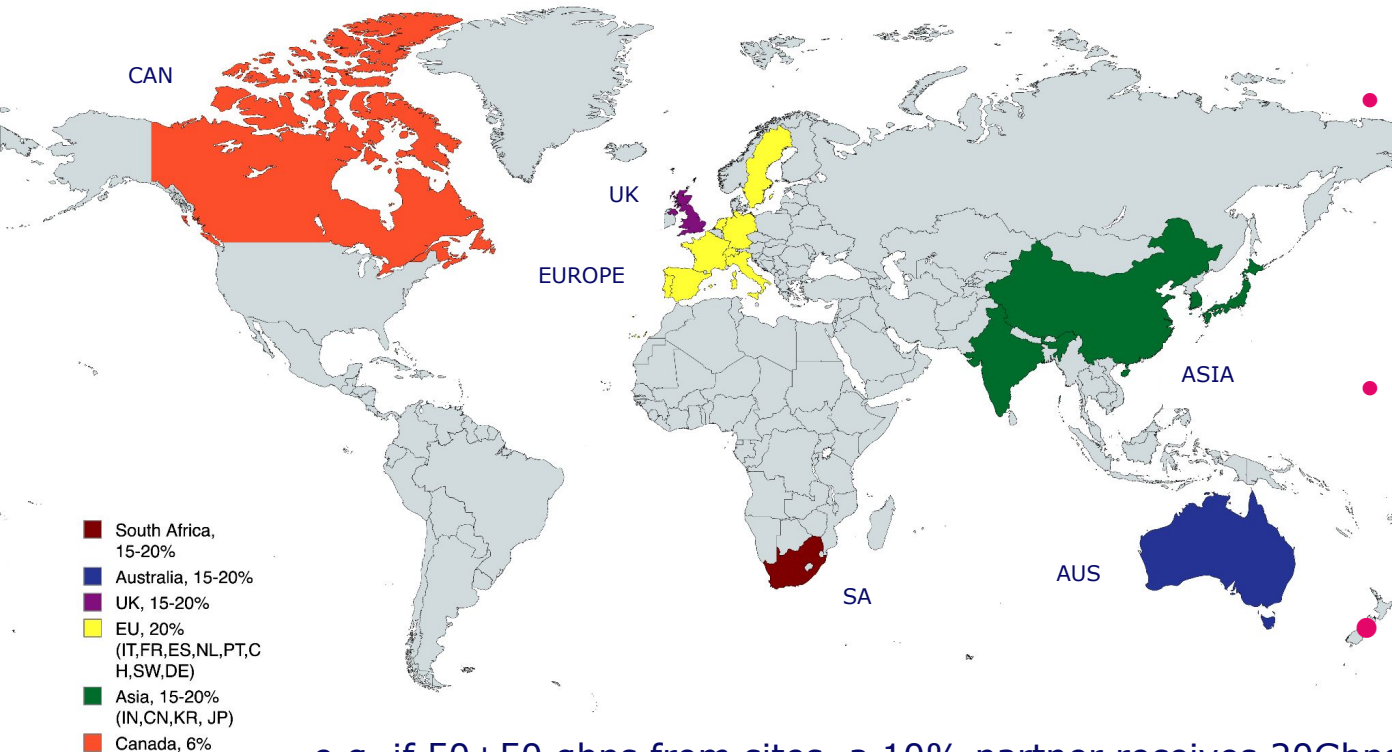


- Roughly, 6 global zones of equivalent size (Canada smaller)
- **Distribute two base copies** of each data product to different countries, and perhaps insist to different regions
- Average incoming rate per (20%) region not more than 2x40 Gbit/s = 80Gbit/s (~2x12 Gbit/s for Canada)
- **Max 100 Gbit/s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

e.g. if 100+100 gbps from sites, a 10% partner receives 40gbps data (400 TBytes per day, 140 PBytes per year)



SKA Regional Centre Broad Distribution: Fair Share (if 50 Gbps per SKAO site)



- Roughly, 6 global zones of equivalent size (Canada smaller)
- **Distribute two base copies** of each data product to different countries, and perhaps insist to different regions
- Average incoming rate per (20%) region not more than 2x20 Gbit/s = 40Gbit/s (~2x6 Gbit/s for Canada)
Max 50 Gbit/s out of SA and AUS

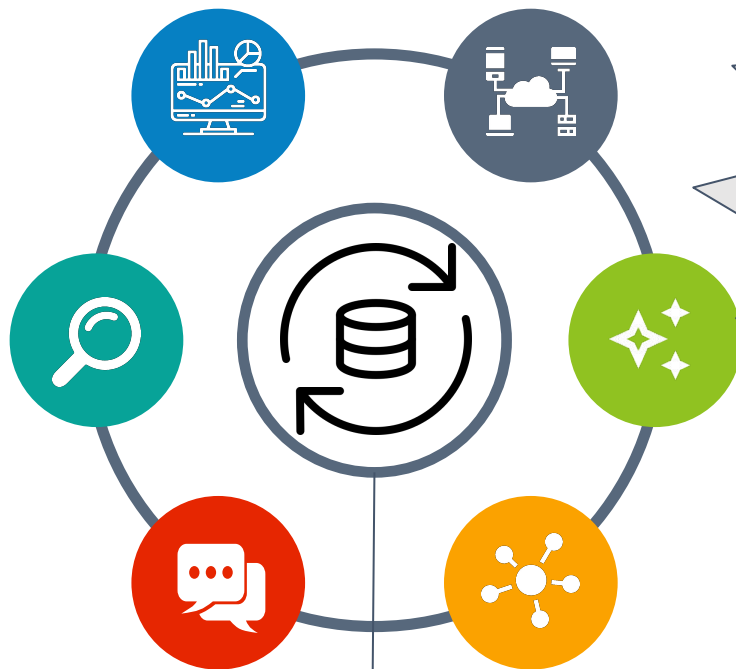
e.g. if 50+50 gbps from sites, a 10% partner receives 20Gbps data (200 TBytes per day, 70 PBytes per year)



SKA Regional Centre Capabilities

Archive of data products continues to grow over the 50yr+ lifetime of the Observatory.

Eventually similar in scale to LHC / HL-LHC



Data Management

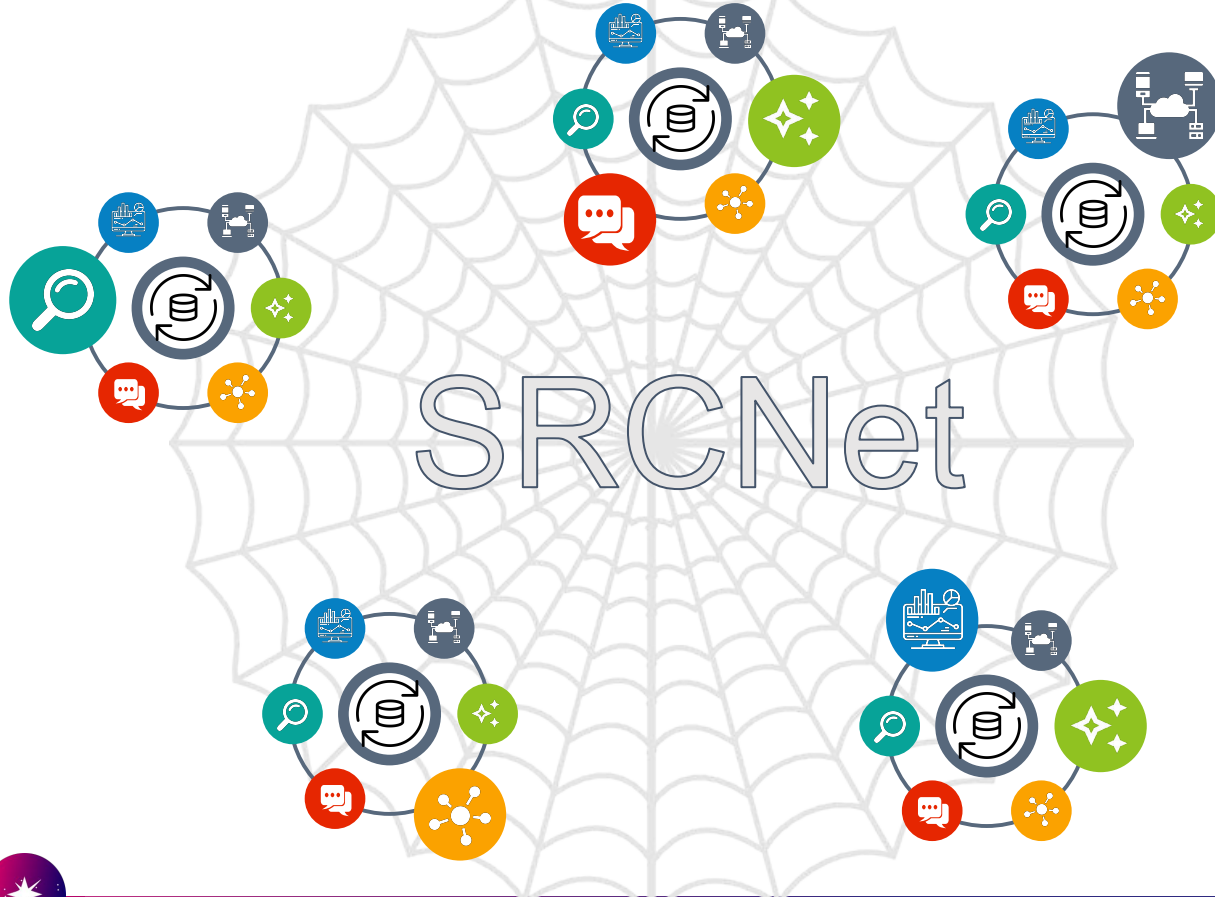
Dissemination of Data to SRCs and Distributed Data Storage

Distributed data management is at the heart of SRCNet

appropriately managing replicas to avoid duplication is the best way to achieve long term cost savings



SRC Network global capabilities



Collectively meet the needs of the global community of SKA users

Anticipate heterogeneous SRCs, with different strengths



Summary - the SRCNet Project

- Vision
- **Roadmap**
- Architecture
- The team of teams
- Implementation of SRCNet0.1
 - Locations
 - Potential implications for network



Top Level Roadmap



SKAO Regional Centre Network

SRCNet Top-Level Roadmap

SRC-0000002	Revision 01
Classification:	UNRESTRICTED
Document type:	PLN
Date:	2023-08-21
Status:	RELEASED
Authors:	Salgado, Jesús; Bolton, Rosie; Swinbank, John; Joshi, Rohini; Sánchez, Susana; Villote, Jean-Pierre; Gaudet, Séverin; Yates, Jeremy; Barbosa, Domingos; Taffoni, Giuliano; Frank, Bradley; van Haarlem, Michiel; Breen Shari; Conway, John; Akahori, Takuya; Yates, Jeremy; Tolley, Emma Elizabeth; Wadadekar, Yogesh; Lee-Waddell, Karen; de Boer, Janneke;

Signed document (August 2023) written by SRCSC members and co-opted helpers from the SRCNet ART

This sits with Architecture Document (SRC-0000001) and Science Platform Vision (SRC-0000003) to say **Why?** (Vision, User driven), **When?** (Roadmap) and **What?** (Architecture).

(We are here to talk about some of the **How?** and **Where?**)

These are unrestricted documents, please ask for a copy



What is covered?

- The Top-Level roadmap declares:
 - Required SRCNet version in-line with other SKA milestones
 - It tries to identify the needs of the scientific community on the use of the SRCNet until first public version (SRCNet1.0)
 - It tries to identify the required resources to implement this intended version on terms of:
 - FTEs and skills
 - Storage
 - Computing



What is not covered?

- The Top-Level roadmap **does not declare**:
 - The ***real*** resources that the SRCNet will have (these would depend on contributions)
 - Governance aspects
 - Operation plan
 - Platform and Software stack solutions
 - It could be considered the seed of an implementation plan but more information on budget, governance, operations set-up, etc is needed
 - It only covers until first operational version (although some information has been added to extrapolate figures)



Roadmap Timeline

First quarter of 2025



SRCNet v0.1

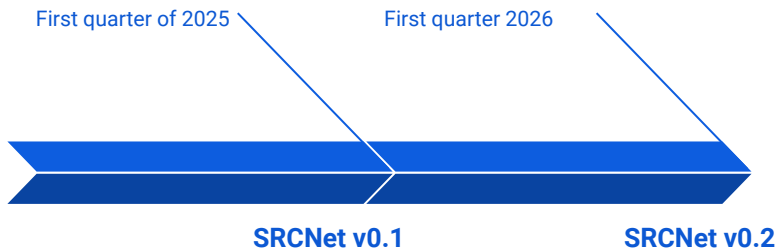
SRCNet0.1 is an internal release Not intended for external users Motivation is to enable testing

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] (Potentially Opportunity to engage SRCNet with AA0.5 data transfer and access.)	<ul style="list-style-type: none">• 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

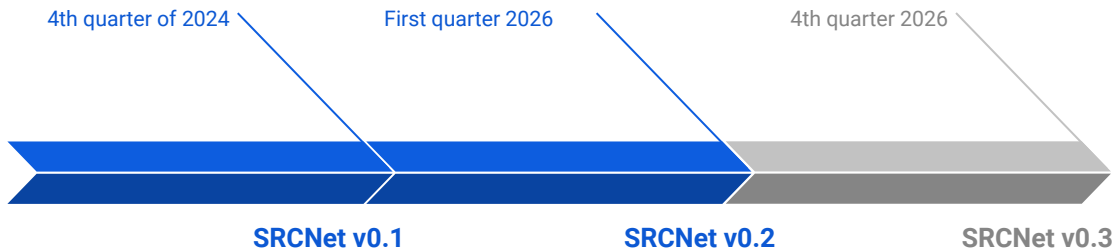


Not generally public
Small amount of science
commissioning interaction
Most SRCNet users are within
the project or SKAO

Milestone	Description	SRC Net Functionality	Scope (users)
SRCNet v0.2 First quarter 2026	AA1 and Commissioning	<ul style="list-style-type: none">• 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



Roadmap Timeline

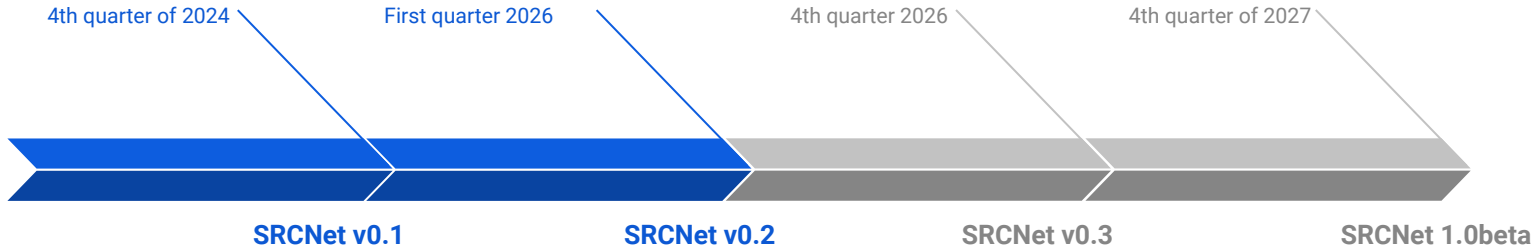


First public access intended for SRCNet0.3 community scientists undertaking Science Verification (AA2)

Milestone	Description	SRC Net Functionality	Scope (users)
4th quarter 2026	Cycle 0 proposals, AA2 and Science Verification	<ul style="list-style-type: none">● Improved data dissemination. Use of available storage● SKA preliminary data (and some precursors data) disseminated into a prototype SRCNet● Upgraded federated computing. Basic execution planner implementation and move execution to a selected SRC● Upgrade of subset SDP workflows runnable in the SRCs● Provide access to the first set of workflow templates for science analysis (light ADPs)● ADPs ingestion system● Spectral data visualisation and manipulation● Implementation of SRCNet User Support	Science verification community (public access) Members of Science Operations SRC ART members



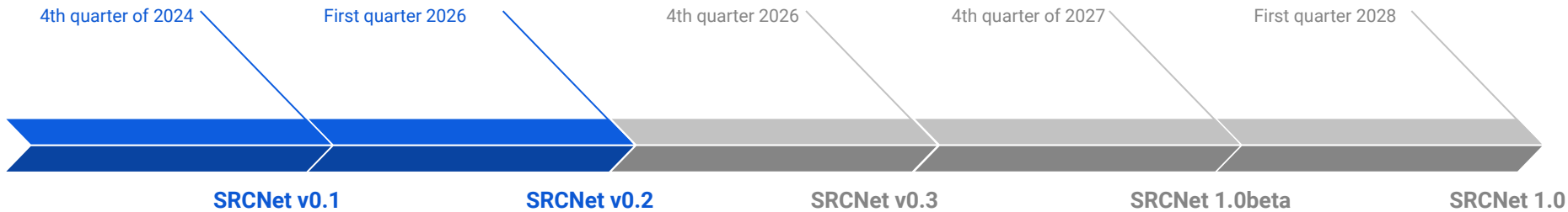
Roadmap Timeline



Milestone	Description	SRC Net Functionality	Scope (users)
SRCNet v1.0beta 4th quarter of 2027	Science verification and Cycle 0	<ul style="list-style-type: none"> • Data dissemination. Complete decision tree, including scientific program • Integrated portal with science analysis capabilities • Integrated federated computing. Workflows analysis • Complete subset SDP workflows runnable in the SRCs • Complete accounting model (storage and computational resources) • Monitoring system • Spectral data visualisation and manipulation • Data previews generation • Restricted SRC Net User Support 	<p>Increased Cycle 0 scientists</p> <p>Science verification scientists (public access)</p> <p>Members of Science Operations</p> <p>SRC ART members</p>



Roadmap Timeline

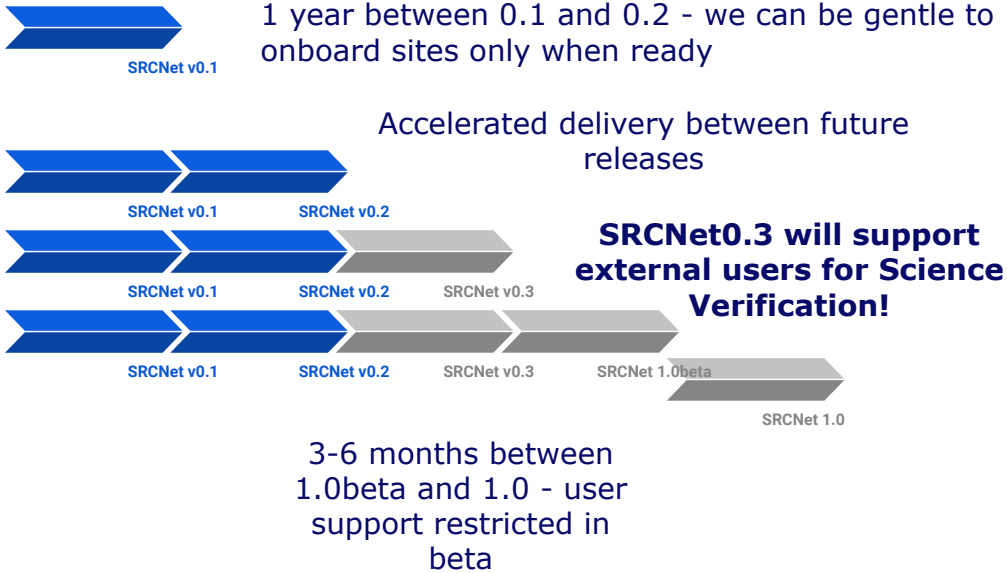


Milestone	Description	SRC Net Functionality	Scope (users)
First quarter 2028	Cycle 1	<ul style="list-style-type: none"> • Full support to PI and program science tasks • Complete portal with science analysis capabilities • Public portal restricted to incoming public data • Not restricted SRC Net User Support 	<p>PIs and science program members</p> <p>Increased number of selected scientists from community</p> <p>Members of Science Operations</p> <p>SRC ART members</p>



Staged Delivery and SRCNet releases side by side

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 May	2025 Nov
AA2	64 dishes 64 stations	2027 Apr	2026 Dec
AA*	144 dishes 307 stations	2028 Jan	2028 Mar
Operations Readiness Review		2028 Apr	2028 Aug
AA4	197 dishes 512 stations	TBD	TBD

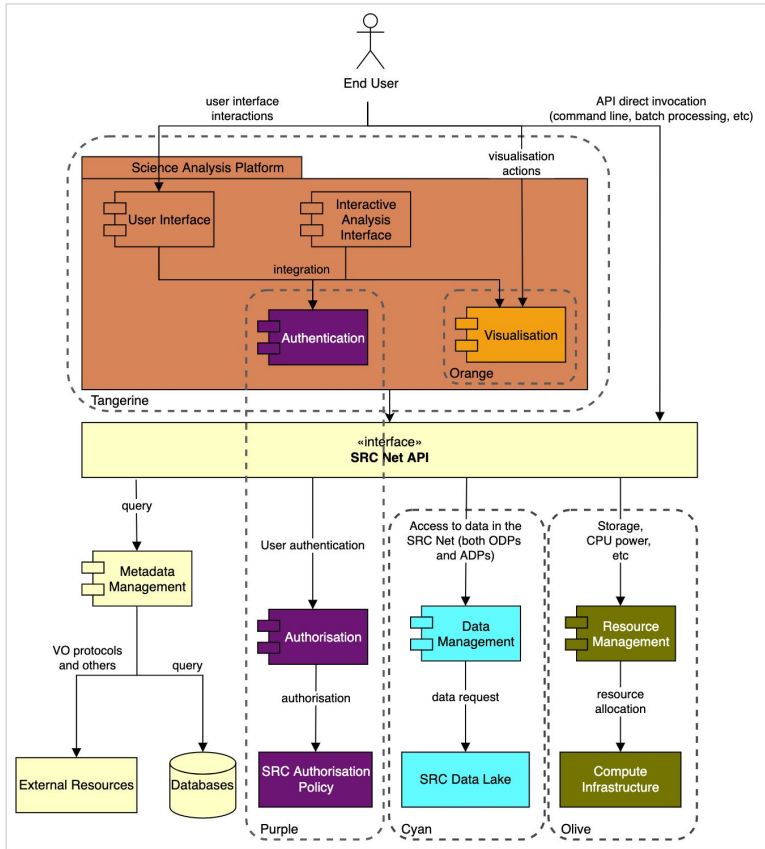


Summary - the SRCNet Project

- Vision
- Roadmap
- **Architecture**
- The team of teams
- Implementation of SRCNet0.1
 - Locations
 - Potential implications for network



Defining an architecture



Users



Science Platform



Interface (API layer)



Metadata query - Science Data Discovery



Authentication Who? Permissions?



Data Logistics Globally distributed storage sites



Compute Resource Management Work sharing



Summary - the SRCNet Project

- Vision
- Roadmap
- Architecture
- **The team of teams**
- Implementation of SRCNet0.1
 - Locations
 - Potential implications for network



The *current* SRCNet teams

Since June 2022 we have been working as a team-of-teams
Engagement from across most SKA countries
25 persons-worth of effort
About 70 contributors



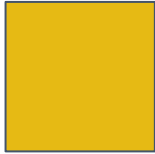
Coral



Purple



Lavender



Gold



Orange



Red



Magenta

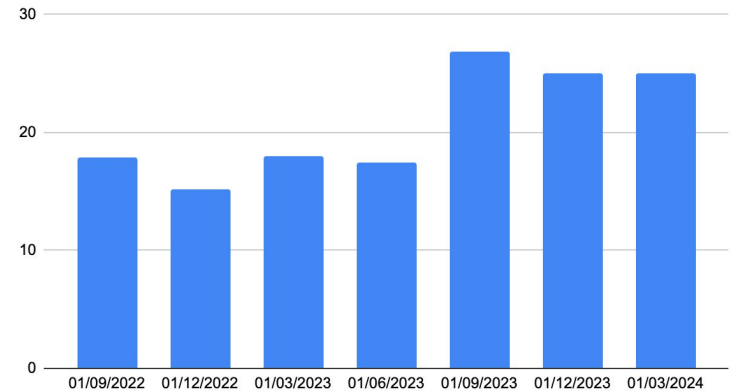


Teal(s)

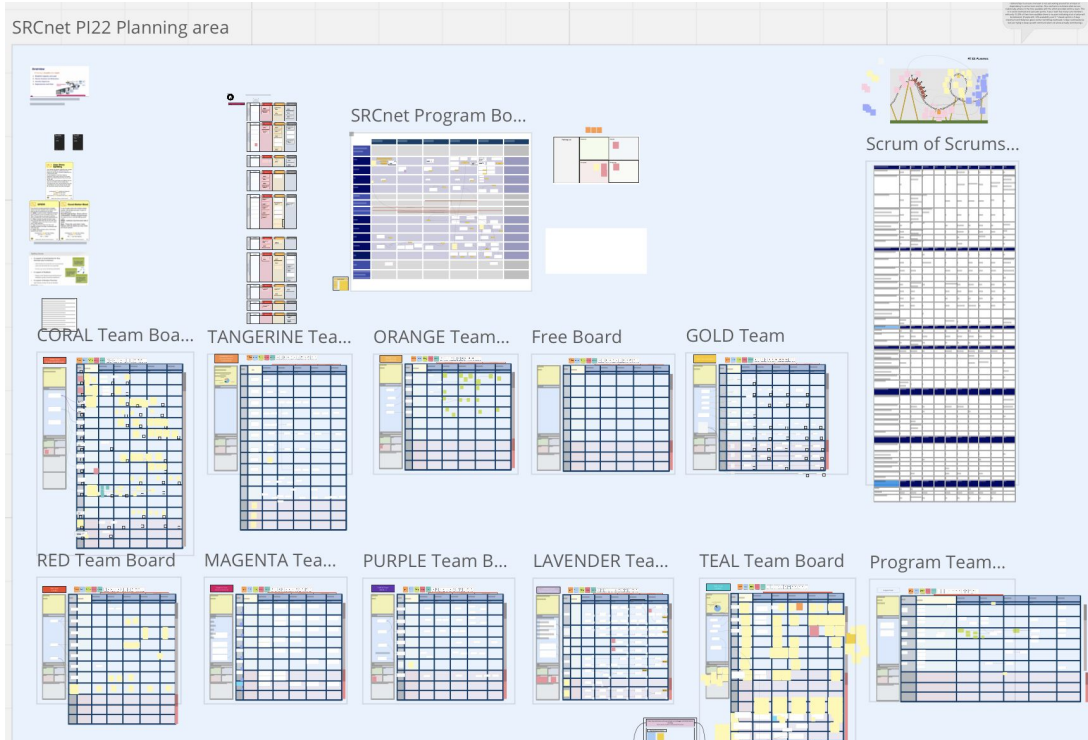


Tangerine

FTE in SRCNet SW development effort by date



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.



Summary - the SRCNet Project

- Vision
- Roadmap
- Architecture
- The team of teams
- **Implementation of SRCNet0.1**
 - Locations
 - Potential implications for network



Big picture progress to 0.1

Science gateway under development well:

Login with SRC IAM

Integrate with APIs

Distributed data management - Rucio and CADC SI prototypes, downselect for SRCNet0.1 by end April



Welcome to **SKA IAM Prototype**

Sign in with your SKA IAM Prototype credentials

Username

Password

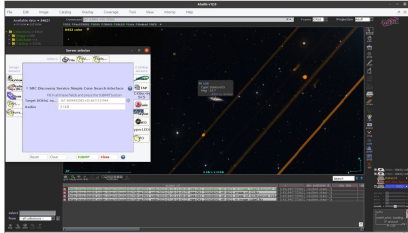
Sign in

[Forgot your password?](#)



Lots of technical progress across many areas

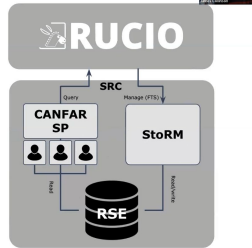
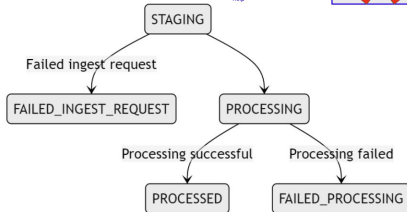
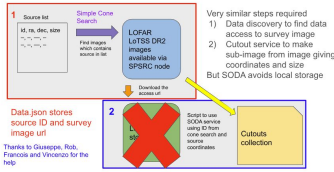
M108 dataset discovery in Rucio data discovery service – Apertif data



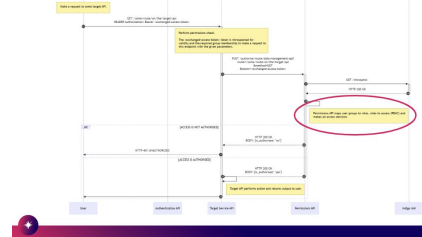
The Demonstrator Collections

LECTION	EXECUTION LOCATION	PRODUCT TYPES	DIRECTION
'S	LEGACY	ODP	LEGACY
'S	LEGACY	ODP	LEGACY
SS	CADC Operations	ODP	Pull from https
'S	Science Platform	ODP, ADP	Push from Science Platform User Storage
LABY	Science Platform	ADP, ODP (future)	Push from VOspace (vault)
POSSUM (future)	AusSRC, Science Platform	ODP, ADP	Push from AusSRC cache, Science Platform User Storage

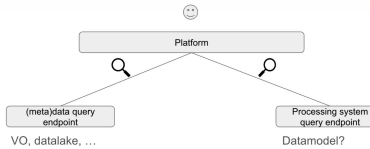
The Modified Workflow for SODA



Authentication and authorisation (3/3)



Feature and context



Deployments of CANFAR

3 (Coral Team) + 1 (Magenta Team) independent deployments of CANFAR

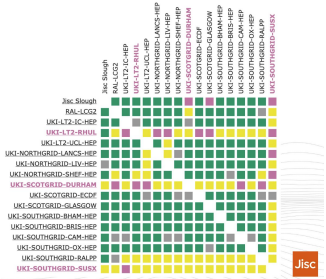


davix-put (http): 5 Flows DTNlon → CamDTN1 & 5 Flows DTNpar → CamDTN1 10 Disk-Disk Concurrent Flows

- DTNlon → CamDTN1 5 flows
- DTNpar → CamDTN1 5 flows
- 100 GByte files 1 TB total data transferred
- On average a file took 3 min
- 1 T Byte in 8 min 20 sec

Time_startup	Time_sending	Total_Time	Data_rate	Data_rate	total_TCP
_s	_s	_s	GBytes/s	Gbit/s	retrans
0.03	141.38	141.40	0.71	5.66	0
0.03	138.28	138.30	0.72	5.79	0
0.02	134.62	134.65	0.74	5.94	6
0.02	130.21	130.24	0.77	6.14	0
0.02	139.98	140.01	0.71	5.72	0
0.05	134.81	134.86	0.74	5.93	0
0.04	130.54	130.58	0.77	6.13	0
0.03	139.64	139.67	0.72	5.73	0
0.04	140.19	140.22	0.70	5.60	0
0.03	139.05	139.08	0.72	5.75	0

- One flow took a long time
- Scaling good 54 Gbit/s with 10 flows

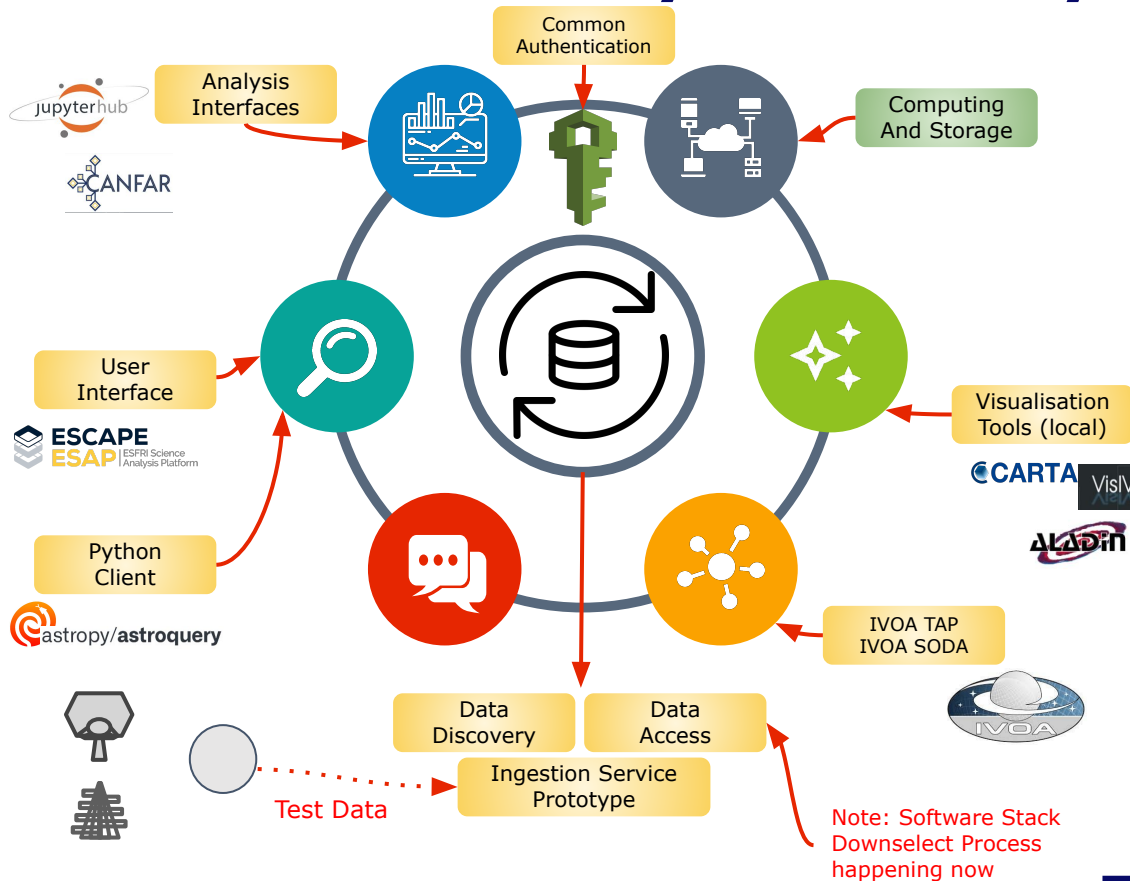


Conclusions and next steps

- Dockerfiles are quite simple to create
 - Need to set up DL, VNC, ...
 - Minimal Dockerfiles to meet skata requirements
- Debugging: rebuild and try again
- VAI/VO and Adata will be accessible in the mini SRCNet Demonstrator mode with the next release of the science platform coming before the end of the current PI.



Basic Functionality Covered by v0.1



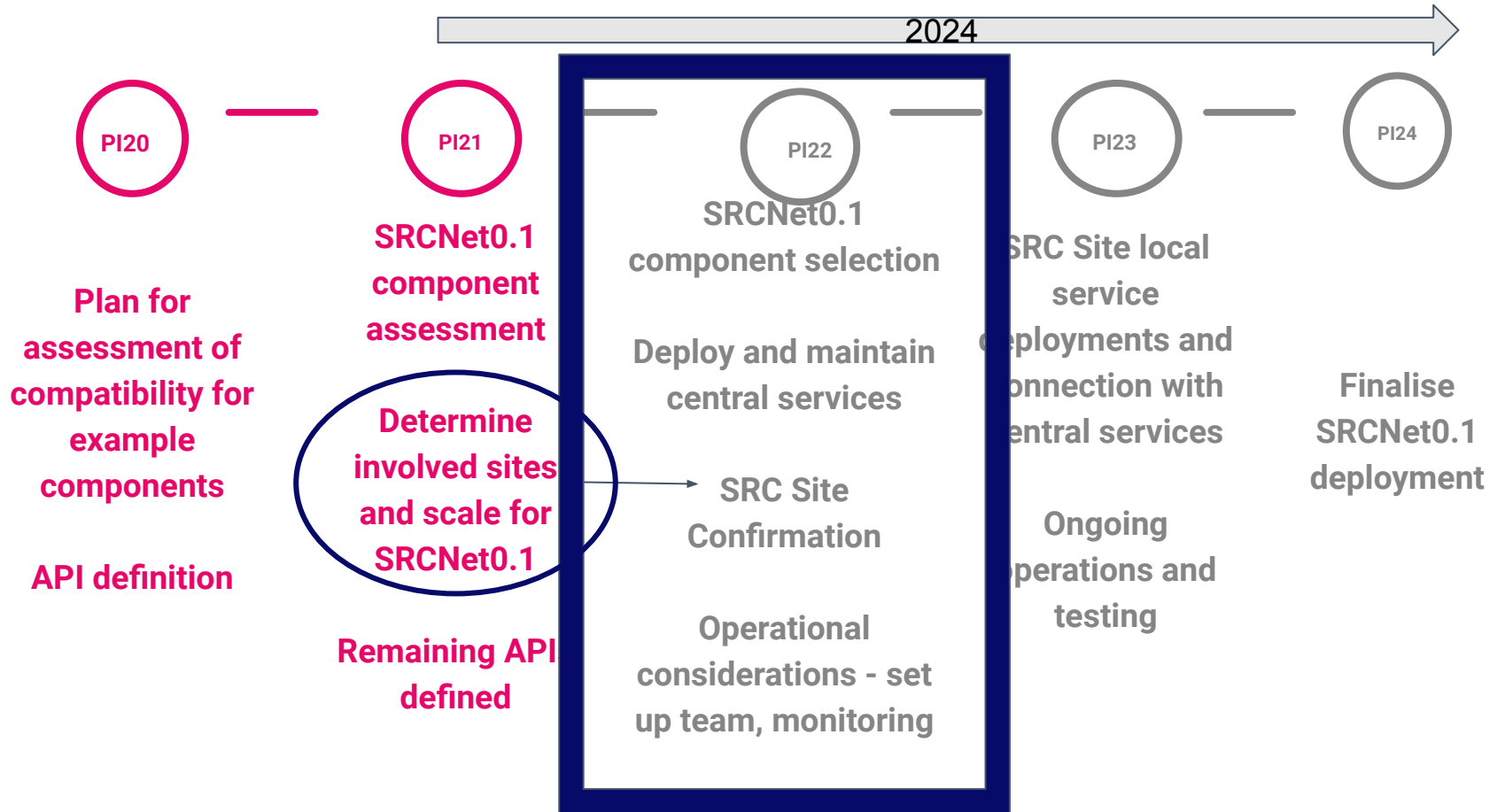
- Common Authentication
 - IAM
- Visualisation Tools (local)
- IVOA Protocols
 - TAP, SODA
- Data Discovery and Access from Data Lake
- Ingestion Service Prototype
- Python Client
 - Astroquery Module
- User Interface
 - ESAP
 - <https://esap.srcdev.skao.int/>
- Analysis Interfaces
 - JupyterHub
 - CANFAR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

Note: Software Stack Downselect Process happening now

Target ~20PB storage

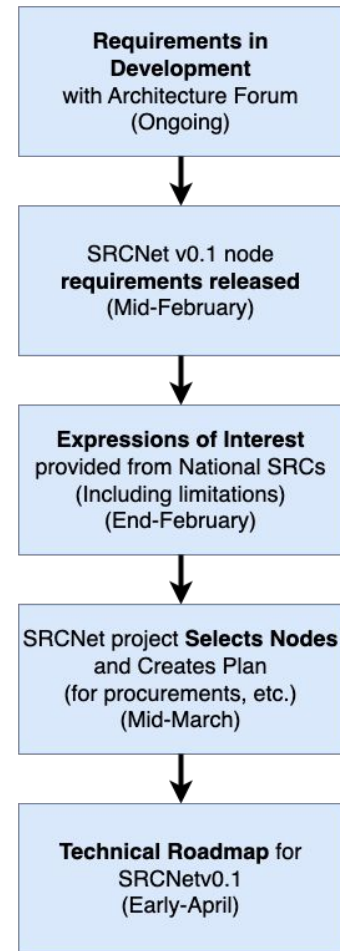


SRCNet0.1 deployment vision



SRCNet v0.1 deployment

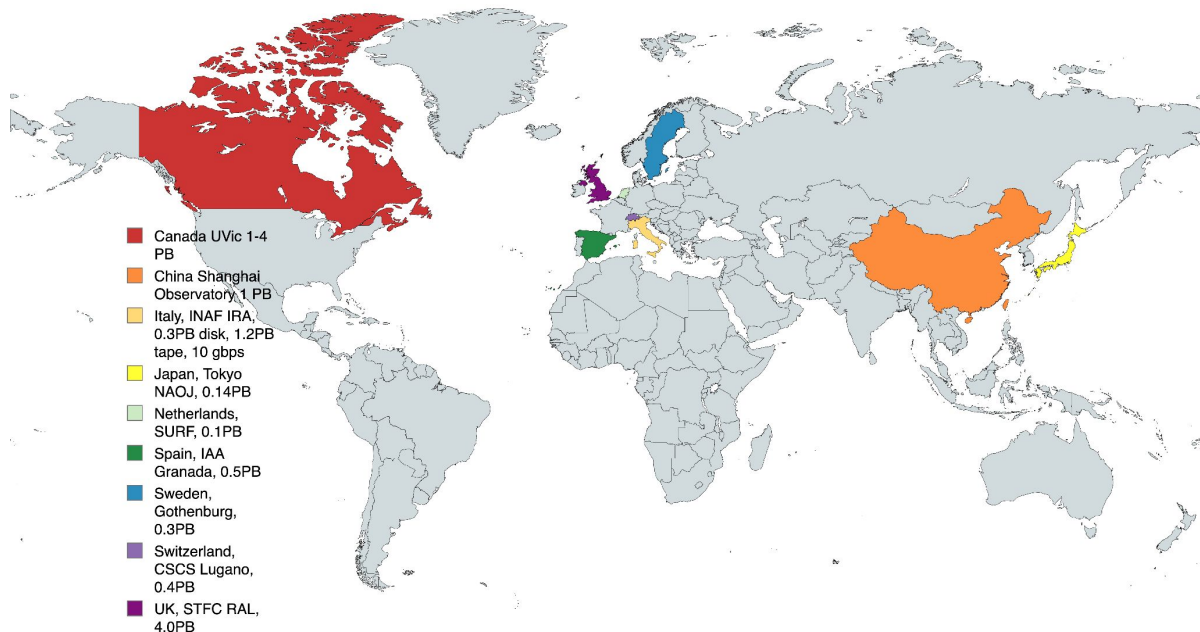
- Deployment of first operational version of v0.1 foreseen for Nov/Dec-2024
- Stable and maintained nodes (target 4) with enough human and hardware resources
- Document describing requirements to have a node has been developed
 - 9 SRCs have expressed interest to be a node, including available resources
 - Workshop held mid March (Shanghai) to discuss path forward



SRCNet0.1 intended sites

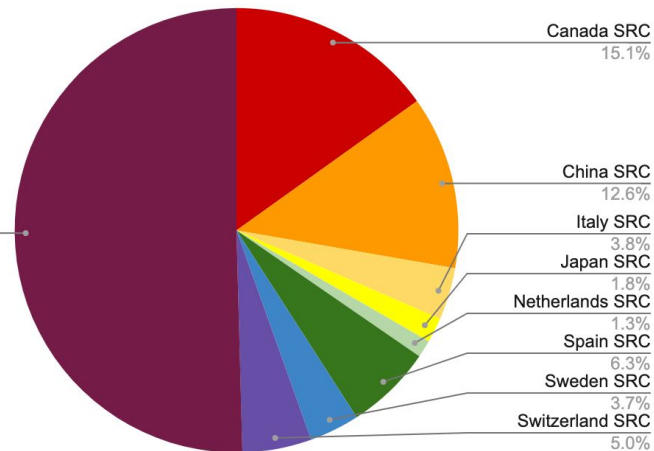
8 PBytes total storage offered for SRCNet0.1 (c.f stated target of 20 PB)

Potentially more subject to usage



Storage (PB)

UK SRC
50.4%



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

Some sites quite new and teams will learn by being involved

SRCNet0.1 testing

Implementation and testing plans are being developed now

Data moving challenges will be a key part of testing the network.

What might be reasonable?



Considering potential use of global data network

Thought experiment (**to trigger conversation only!**)

Imagine SRCNet runs a series of testing challenges, e.g. for a week every 3 months attempt to fill / transfer 20% of offered SRCNet0.1 storage

	Canada SRC	China SRC	Italy SRC	Japan SRC	Netherlands SRC	Spain SRC	Sweden SRC	Switzerland SRC	UK SRC	Total
Data rate if fill 20% storage in 7 days, Gbps	3.2	2.6	0.8	0.4	0.3	1.3	0.8	1.1	10.6	21.0
Current connectivity or allowed share	? (but UVic has high connectivity globally, not sure of SRCNet allowance)	1	10	10	100	10	100	10	100 as part of coordinated plan, need to discuss	

Based on information from SRC technical leads, all sites could accommodate these tests currently apart from China. Significant share to RAL / UK connections



Automated testing

Screen grab here from the dashboard for the current Rucio prototype

Some science data in Rucio but most of dashboard events are from automated test suite (**Rucio task manager**), just running as a heartbeat currently with all-to-all transfers every few minutes across 12 different locations, but also used for data throughput tests. Very powerful to complement lower-level test work



<https://gitlab.com/ska-telescope/src/ska-rucio-task-manager>



Automated testing

Rucio task manager is extensible and has been used to test specific ideas

Builds on the experience gained from within the ESCAPE project

Well suited to running planned data challenges and a likely component for SRCNet testing



<https://gitlab.com/ska-telescope/src/ska-rucio-task-manager>



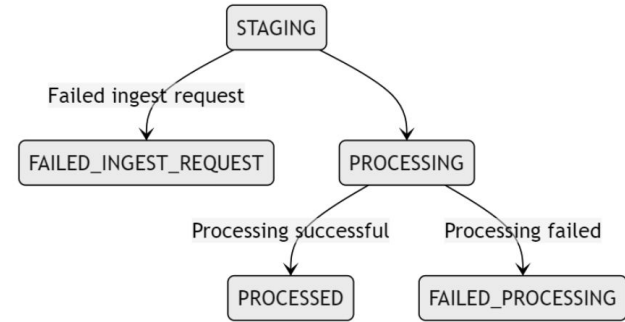
Data Ingestion

We have a prototype data ingestion service, detects new files and their metadata, adds file to Rucio (triggers replication) and adds metadata to the auxiliary metadata database

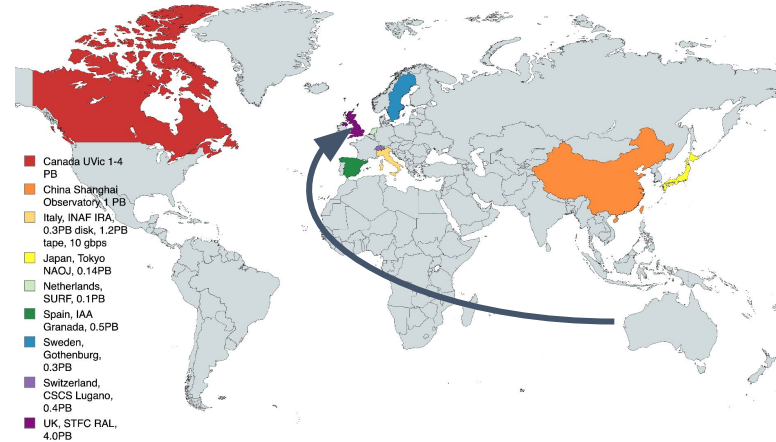
In AUS we have a small Rucio storage element and a small team in the SRCNet project. It would be great to test operational long-distance data management.

Question: If we were to run intensive data ingestion tests for (e.g.) a week from Australia (Perth) to Europe (UK) what bandwidth could we hope to use?

For continuous tests what bandwidth would be acceptable / below impact?



<https://gitlab.com/ska-telescope/src/ska-src-ingestion>



Created with mapchart.net



SRCNet0.1 testing

Implementation and testing plans are being developed now

Data moving challenges will be a key part of testing the network.

What might be reasonable?

How can we connect the right people to find necessary capacity? Are you all in the NREN forum? Are other formats useful?



Questions / discussion

