SRC Network: Executive Summary

SRC-00000XX	DRAFT
Classification:	UNRESTRICTED
Document type:	PLN
Date:	2024-01-19
Status:	DRAFT
Authors:	

Role	*Name*	*Designation*	*Affiliation*	*Signature*	*Date*
Author	SRC Steering Committee		Various		
Owner & Approver	M. van Haarlem	SRC Steering Committee Chair	ASTRON		
Released by	L. Ball	SKAO Director of Operations	SKAO		

TABLE OF CONTENTS

LIST OF ACRONYMS

AA	Array Assembly
AAI	Authentication & Authorisation Interface
ADP	Advanced Data Product
API	Application Programming Interface
ART	Agile Release Train
Co-I	Co-Investigators
FTE	Full-time Equivalent
HPC	High-performance computing
KSP	Key Science Project
ODP	Observatory Data Product
OLDP	Observation-Level Data Product
PFLOPS	Peta Floating-Point Operations Per Second
PI	Principal Investigator
PLDP	Project-Level Data Product
SDP	Science Data Processor
SKA	Square Kilometre Array
SKAO	SKA Observatory
SOG	SRC Operations Group
SRC	SKA Regional Centre
SRCNet	SKA Regional Centre Network
WLCG	World-wide LHC Computing Grid

1 Introduction

1.1 Purpose and context of the document

This document presents an overview and executive summary of the SKA Regional Centre (SRC) Network. It is based on previous documents prepared and published by the SRC Steering Committee at either past Council meetings, or for the SRC Network review by the SEAC in September 2023. This Executive Summary provides a high level view of the need for an SRC Network, a description of the operational concept and the roadmap for delivering that Network. The list of documents (and the links to them) that this draws from are:

- SRC Network Concepts and Planning (2022)
- SRCNet Operational Concept
- SRCNet Top Level Roadmap
- SRCNet Software Architecture
- SRC Network Vision and Principles
- SRCNet Science Analysis Platform Vision
- SKA Observatory Establishment and Delivery Plan

The organisation and management of the operational SRC Network will be the subject of a separate document to be considered by the SKAO Council, and will not be addressed in this Executive Summary.

For further context, it is noted that the software development work that follows the roadmap described in §6 has already begun and is ongoing, with specific management processes in place.

2 What is the SRC Network and why it is needed

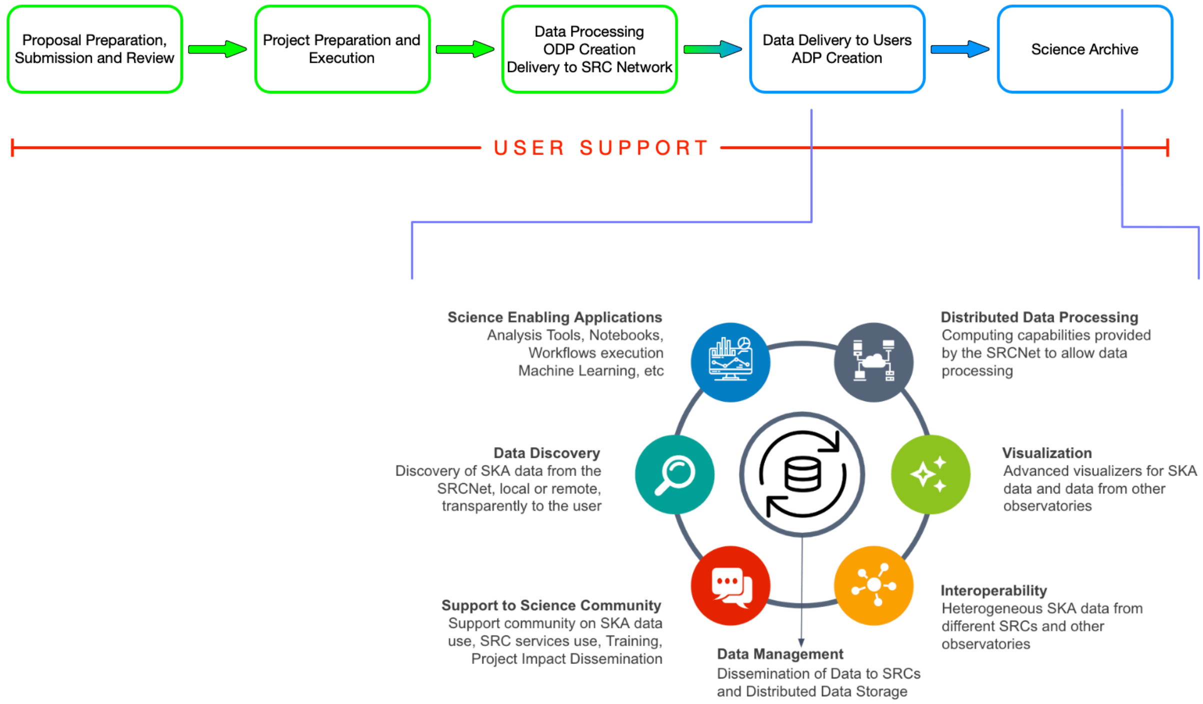
In 2016, the SKA Board of the SKA Organisation requested the SKA and its partner states to form a collaborative network of SKA Regional Centres to provide those essential functions and capabilities that are not funded within the scope of the SKA Project, thus completing the end-to-end system from telescope proposal to science delivery.

The aim is to establish and deliver an operational, science-driven SRC Network as a collaborative e-infrastructure with a common interoperable software platform of services and tools that federate the distributed computational and data resources provided by locally funded and operated SKA Regional Centres. The SRC Network will deliver to the international SKA science community global capacity and capability to transfer, archive, manage, process and analyse tera- to petabyte-scale SKA data within and across the SRCs, together with user support, enabling the community to extract impactful science from SKA data.

Work began to describe the functions, capabilities and architecture of the SRC Network and the service layer that will need to be locally deployed, supported and funded to interoperate across a global collaborative network of SRCs. This network will provide the SKA user community with access to SKA data via a common platform of tools and services that provide the compute and user support to enable SKA science. The collection of tools, services and infrastructure that constitute a global SRC capability is referred to as the SRC Network (SRCNet).

The SRC Network is driven by the high degree of internationalisation of the SKA science programme – membership of large Key Science Projects (KSPs) will, by design, be drawn from across the international SKA community (and it is anticipated that many PI projects will also be international collaborations). This, together with the scale of SKA data to be transported, archived, managed and processed needs a persistent, scalable and operationally resilient solution that is compatible with the SKA data flow for the lifetime of the SKA Observatory. The total storage and compute requirement for SKA science drives the need for an SRC Network that provides the resources to globally distribute observatory data products (OPDs) to SKA scientists. In turn, SKA scientists will require an SRC Network in order to process these data into Advanced Data Products (ADPs), enable the visualisation and analysis of tera- to petabyte scale data products, as well as operate a science archive facility.

The scale of both storage and computational requirements, more than any one country can do alone, and the international nature of SKA science, requires a global and collaborative solution. Rather than a set of loosely connected national SRCs that each support their national users, the solution needs to be delivered through predictable, reliable and stable federated resourcing, together with a collaborative SRC Network governance and business model between the member states and SKAO. This latter aspect will be addressed in a separate document.



The figure above illustrates the end-to-end science operations to be delivered by the SKA Observatory and the SRC Network (green is in the SKAO, blue is in the SRC Network). The cycle of tools and services to be provided by the SRC Network are also shown.

2.1 Incentive for participation in the SRC Network

Building and operating the SKA represents a significant collaborative international investment. The scientific return from that investment is maximised by providing researchers with a platform from which to access and process scientific data, with the ability to store, explore and reuse those data products. Without an end-to-end infrastructure, the SKA will not achieve the maximum potential that it can on those national investments.

Incentives for participation by SKA partners in the SRC Network are detailed in the document *SRC Network Concepts and Participation*. These are listed as:

- Reduced cost and duplication of effort
- Broader availability of resources
- Expanded SKA capability with an end-to-end e-infrastructure
- Expanded Global collaborative opportunities
- New research areas
- Technological and Industrial opportunities
- Development of skills

In the main, the benefits come from operating and having access to a global infrastructure at scale, as opposed to each country needing to provide resources to their communities. This has advantages to both countries with relatively small and large investments in the SKA. Not all SKAO partners will have ready national access to the scale and range of capabilities needed to achieve the best return from the SKA. Through participation in the SRC Network effort, individual nations will all have access to SKA-scale capabilities, contributing their unique national skills and capabilities in an optimal manner through partnerships within the global SRC effort.

From the perspective of this global cooperation, it is beneficial that engagement in SRC initiatives lies not only with current SKAO members, but also with those countries who are considering joining the SKAO in the future.

More discussion on incentives for participation in an SRC Network is in the accompanying Organisation and Management document.

3 Vision and Principles

The Vision for the SRC Network is presented in the *SRCNet Vision and Principles* document. Its purpose is to steer an international strategy that informs and guides the development work of the SRC Network, as well as the operations of that Network.

3.1 Vision statement

The Vision statement for the SRC Network is:

> We will develop and deploy a collaborative and federated network (SRC Net) of SKA Regional Centres, globally distributed across SKA partner countries to host the SKA Science Archive and to provide the resources to scientifically process and analyse these data . 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.
>
> 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 federation and operations;
> - growing the prototype SRC Network, as new SRCs become available, leading towards a fully operational and global SRC Network.

This vision is presented at Program Increment planning meetings for SRCNet development. While presenting a long term vision for the SRC Network, there is an initial focus on the immediate work to deliver the first prototype SRC Network together with the minimal set of core services.

3.2 Principles of the SRC Network

The *SRCNet Vision and Principles* document presents a list of 21 principles for the SRC Network. These principles are not reproduced here, but it is noted that they fall under the following categories:

- Operational
- Sustainability and Environmental
- Security and Federation Identity system
- Fairness, Equity and Inclusion
- User-facing
- Compute and visualisation
- Data Storage and Archive
- Data Processing and analysis

4 Core capabilities and functions

The table below presents a list of the core capabilities and functions that the SRC Network needs to deliver in order for the SKA Observatory to present a complete end-to-end capability to its users. This list (presented in no assumed order of priority) informs and drives the delivery roadmap of the SRC Network.

SRC Network Capability/Function	Description
Access to SKAO data for project users	Users with access rights to SKAO data must be able to access that data.
Public Science Archive	Users will have the ability to browse the archive, searching and filtering on publicly available data.
Science Analysis Platform	A basic science analysis platform will provide (authorised) access to project data, processing and visualisation tools and services. Users, with the appropriate access rights to SKA data, must be able to process, visualise and analyse that distributed data.
Receive data from SKAO	The SRC Network must be able to receive data from the SKAO and place them within the SRC Network when the Observatory is ready to transfer that data.
Replica management	Ensures that there are sufficient copies of ODPs accessible from the SRC Network.
Monitoring of performance against pledges	To monitor performance/usage/capacity, at each SRC node, against the resources pledged to the SRC Network.

SRC Network Capability/Function	Description
Monitor SRC availability	At any point in time, ascertain whether a particular SRC node in the SRC Network is available and capable of receiving data, and whether that availability will remain for the duration of any data transfer that is planned or scheduled to that node. Monitor, and report, SRC availability for users to access data and services.
Single Sign-on and Secure, Federated AAI	One login for all users with an SKAO account. There is a need to abide by SKAO data access rules and also global/national/local cybersecurity rules.
User support and community training	Helpdesk support, training and events.
Creation of Project-Level Data Products	Resources within the SRC Network so that PLDPs can be generated by SKAO staff.
Execute SDP pipelines/workflows on the SRC Network	To facilitate the user community to define/refine their SDP workflow parameters. This also creates a pathway for user engagement with the SKAO towards the improvement of SDP workflows and pipelines.
SRCNet Software development coordination, product management and architecture development	Essential tasks for setting up and managing the software development SAFe train. Ensure that the development work is aligned with the needs of the global SRCNet and on the key areas that are essential functions of the SRC Network.

5 Operational Concept

The existence of a collaborative network of SRCs is required both by SKA and its science users. The SRC Network is operationally critical for the delivery and access of data from SKA to users and they are mission critical for the delivery of science results from the SKA. The SRC Network will serve as a federated analysis facility for its user base, allowing for collaboration on shared projects, whether smaller PI projects, substantially larger Key Science Projects (KSPs), or archival projects. Federating entities are absolutely critical to the SRC mission - they turn a heterogenous set of HPC facilities into a coherent science analysis ecosystem.

Through a shared and federated pool of resources, the SKA proposal review process will have the freedom and flexibility to recommend a science programme without the need to attempt to manage individual national SRC resources. Compute resources on the SRC Network are allocated to projects from that pool, rather than resources from a country, to projects from a country. Given that the majority of projects will be collaborative across national boundaries, such a model is very appealing.

Such a federation also introduces robustness and high availability of the SRC Network to its users. Individual nodes and the local services they offer can become unavailable but the SRC Network as a whole will remain functional and available. As such, one of the principal concepts of the SRC Network is that it should not be necessary for a user to know which SRC node they are using.

The responsibility for maintaining the critical federating services will be a collective effort (under a governance structure that is to be defined). Nevertheless, it is expected that the support and management of these SRC services across the SRC Network will be by an SRC Operations Group (SOG), who will perform a coordinating, operational role in the day-to-day work needed to manage the SRC Network. The effort of these staff is provided through the pledging process.

5.1 Data management

Data will flow from the Science Data Processors at the two SKA telescopes and be delivered and placed into the SRC Network. As such, and in order to maintain the throughput of the Observatory's science programme, the capability to ingest data products must be available somewhere across the SRC Network. To enable this, and the data processing and analysis, there needs to be sufficient data network bandwidth for those transfers.

The SRC Network will operate a global data management system to provide coverage and reliability of the whole SKA archive, to enable implementation of data integrity, data location and data life cycle policies and to give visibility into the location of all SKA data products (ODPs and ADPs) in the Network.

5.2 Data processing

One of the key aspects of the SRC Network is to bring high-performance distributed data processing of large data sets (terabyte to petabyte scales) to a broad and globally distributed community.

SKA must be confident that science data products generated by each SDP can be post-processed, analysed and interpreted by users without impacting the operation of the SDPs themselves. In fact, there will be no access to the SDPs by the user community. The tools, workflows and pipelines must therefore be made available to the community by the SRC Network.

The SKA Observatory Establishment and Delivery Plan defines three general types of data products in two categories:

- Observatory Data Products (ODPs): These data products come in two forms.
    - Observation-Level Data Products (OLDPs): calibrated data products generated by SDP workflows and based on data obtained from a single execution of a scheduling block.

    - Project-Level Data Products (PLDPs): calibrated data products generated by combining several, related, observation-level data products, delivering the requirements of the PI as outlined in their original proposal.
- Advanced Data Products: user-generated products, produced through the detailed and rigorous analysis and modelling of Observatory data products (either at the observation or project level). The generation of ADPs will usually require some level of interactive visualisation and examination of data, as well as comparison to data from other SKA observations and/or other facilities.

OLDPs will be delivered into the SRC Network for archiving and community access (abiding to the SKAO data access policy). PLDPs will be generated using workflows and pipelines that will be executed within the SRC Network by SKA Observatory staff, using software that will be developed and maintained by SKAO. PLDPs will also enter the science archive.

Users will still have access to their OLDPs, and so retain the freedom to combine data in different ways. In fact, it is anticipated that this will be an important feedback loop from the community to the Observatory for the continued improvement of workflows and algorithms.

5.3 Development and architecture

The development and deployment of shared software across the SRC Network will be coordinated and organised within an Agile Release Train (ART), part of the Scaled Agile Framework for large-scale software development.

All SRC nodes should share a common architecture view in order to present SKA data and analysis capabilities in a consistent manner to its global user base. The community should not refamiliarise themselves with the SRCNet platform as they move around the globe, or as they work with collaborators around the world. In addition, to implement 24/7 support (taking advantage of the global nature of the project), common systems across the Network are necessary.

As such the development effort of the SRC Network should produce a default implementation of all core modules so SRC nodes with fewer development resources, and implementation flexibility, can just reuse and deploy them locally without significant adaptation. Moreover, this makes it easier to onboard new SRC nodes into the SRCNet as the Network grows.

The architectural view of the SRC Network is typical of modern science archives. The significant difference with other platforms comes from the federated aspects of the SRCNet. These include (but are not limited to):

- data lake - distributing data from repositories at different SRC nodes;
- authentication/authorisation - a common, network wide, federated authentication system linked to a shared authorisation management system providing common security for access to data and computing resources. The SRCNet will follow global and local security policies;

- federated computing - enabling deployment and execution of workflows across different SRCNet nodes;
- visualisation - efficient visualisation to accommodate the size of the distributed SKA data products while minimising data movement.

For a full description of the modules that compose the SRC Network and individual nodes, refer to the SRCNet Software Architecture document.

5.4 User support

User support will be distributed across the SRC Network and staffed through the pledging process. The term "user" includes not just PI and KSP teams, but also scientists making use of public SKA data through the science archive.

The user support function will be provided out of the SRC Network by local staff primarily serving their regions. A degree of user support will also be provided by SKAO. There will be a centralised Helpdesk (most likely based on Jira) to be staffed by both SKAO and SRCNet personnel that will cater for the entirety of user needs.

As well as responding to user queries and bug reports, the user support function will manage a repository of extensive and up-to-date documentation on instrumentation, how-to guides, procedures, rules and regulations, important dates, schedules and timelines, and will also create and deliver user training events to local communities (either face-to-face or online).

An SRCNet portal will be deployed across the SRC Network, providing a common user-interface (but allowing for regional language settings), to access the services and functions as described in Section 4. There will also be a system administration version that will allow for common user admin functions and the generation of reports.

There is an implicit assumption that there will be a Users' Forum where representatives of the SKA community will be able to offer advice and feedback to the SKAO and SRC Network. This is not elaborated on here but is discussed in the *SKA Observatory Establishment and Delivery Plan*, as well as the accompanying Organisation and Management document.

5.5 Resourcing and pledging

The exact mechanism and process that will be followed for pledges to be made to resource the SRC Network development and its operations is still to be determined. However, it is strongly expected that the SRC Network will receive pledges of resources and not financial contributions to spend. National representatives of each participating country will be responsible for making available those resources to the SRC Network.

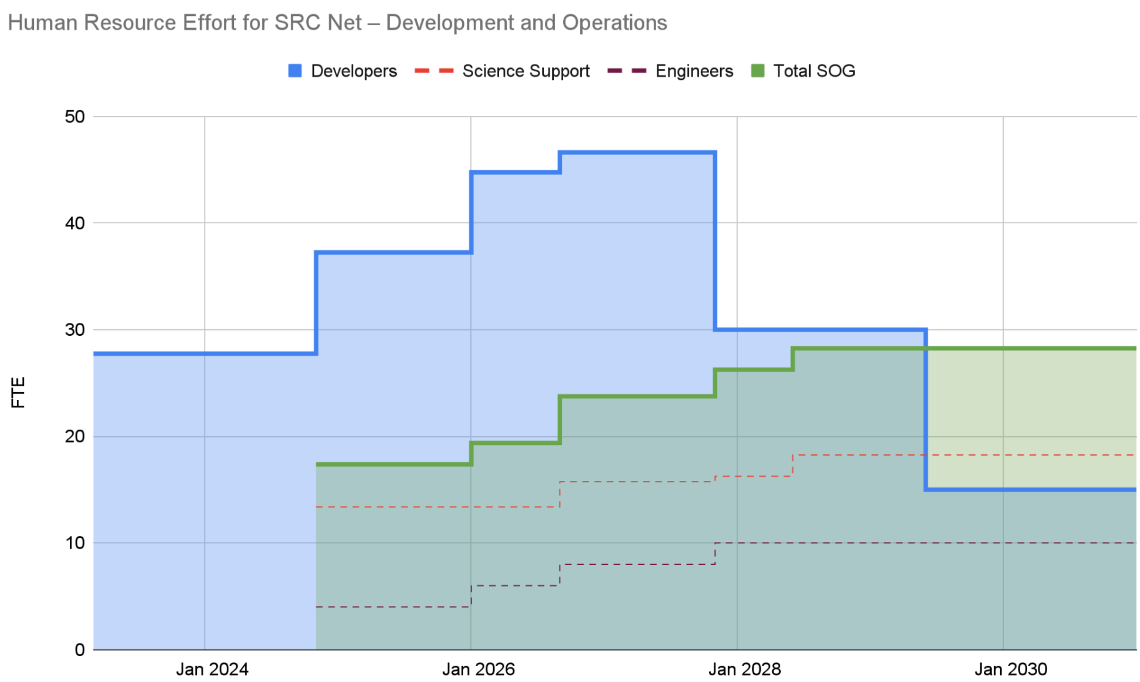
We note that the WLCG operates a similar model as is planned for the SRC Network and functions at a similar scale to the SKA, although of course, the SKA is an Observatory with many active projects at a time that change over time.

The mechanism and model(s) for pledging resources into the SRC Network is still to be determined and forms a part of the discussions regarding the governance of the Network.

5.6 Size of the SRC Network

5.6.1 Development needs

The figure below shows the estimated level of Developers (blue) and Operations staff (green) through the development phase of the SRCNet, and beyond. The *SRCNet Top Level Roadmap* document explains in more detail how these numbers are estimated.



The Developers consist of roles across:

- software and data management engineers;
- cloud software engineers;
- database and security engineers;
- software and workflow repository maintainers; and,
- computing and HPC engineers.

These roles are aggregated into the Developers (blue) line in the graph above. The drop in numbers from 2028 onwards is an estimate of the change in development needed as the focus of the SRC Network changes from construction and establishment to operations.

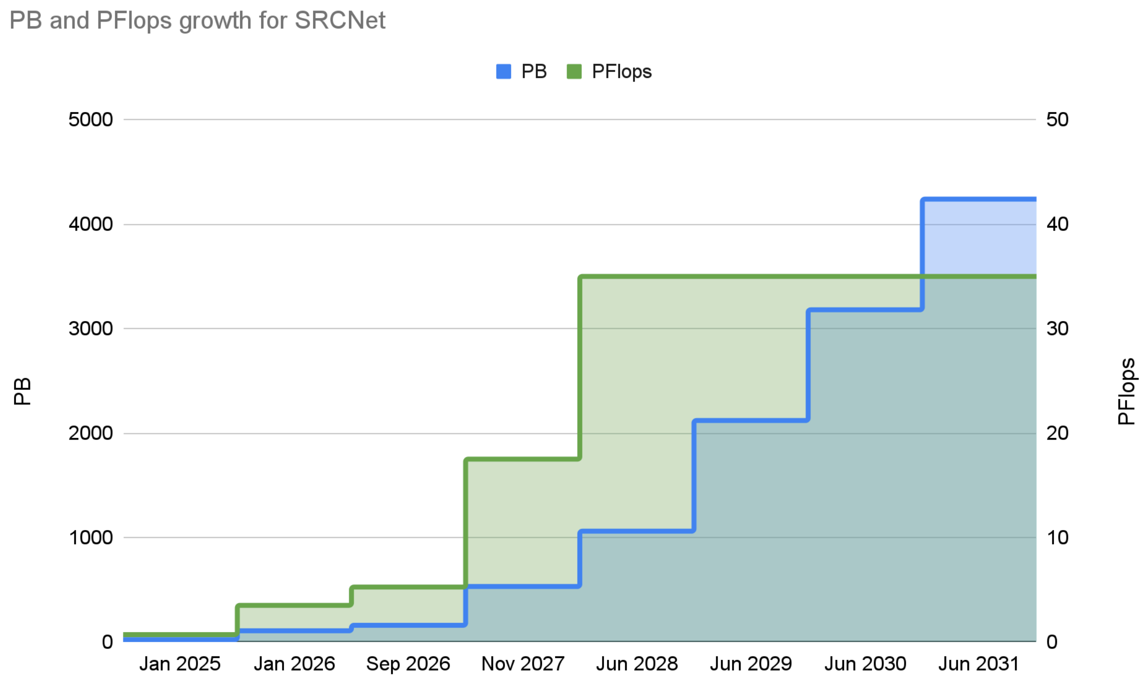
The SRC Operations Group consists of roles across:

- helpdesk engineers and science staff;
- science workflow support;
- data scientists; and
- operations engineers.

Note that user support functions for the community will come from some of these technical and scientific functions. At the time of writing the full user support requirement has not been investigated.

5.6.2 Hardware resources

The method used for estimating the hardware resources (storage and compute) is explained in more detail in the *SRCNet Top Level Roadmap* document. A brief description is given here.



The diagram above shows the growth of the storage (blue, left axis) and the processing (green, right axis) requirement for the SRC Network through construction and into the first couple of years of SKA operations. The storage and processing needs are based on work carried out by Bolton and Ratcliffe (2021) for estimating the system size of the SRCNet. The following are assumed:

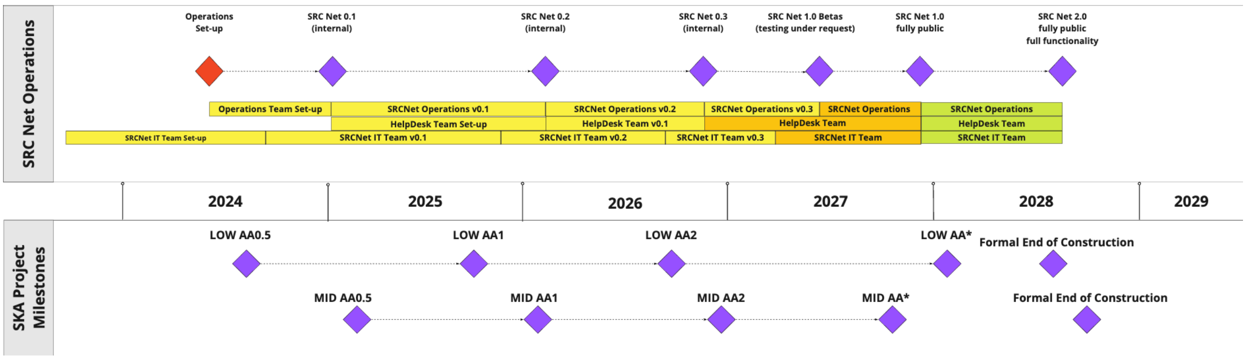
- no ODPs are deleted;
- there are two copies of SKA data products across the SRC Network;
- on-line (hot) storage will be maintained at more or less the same level (approx 530PB) once in full Operations;

- the annual increment of the storage is approx. 1060PB (mixture of near- and off-line);
- data older than a year will be moved to near-line (warm) storage;
- unpopular[1] data products will be moved to off-line (cold) storage with a relatively high latency to recover any of those data products if they are requested;
- processing peaks of 35 PFlops with sustained compute of 22 PFlops;
- data products are processed multiple times (especially in the early years), which with time may fall to an average of 3 times (note that this is just a crude estimate at this time and is an average across all active projects).

These numbers are based on the Operational phase of the SKA Observatory (and also assuming the SKA design baseline) and SRC Network. During the growth phase, there will be a gradual deployment of resources leading to the first public version with the start of SKA Operations. Increases leading up to that point will happen in agreement with SRCNet partners.

6 Roadmap for delivery

The roadmap for delivering the major, high-level, milestones (blue diamonds) of the SRC Network, and how they align with the SKA Project (through construction), is illustrated in the figure below:



Highlights of what is delivered with each SRCNet milestone, and its connection to SKA project milestones, is given in the table below. Further details are available in the SRC Top Level Roadmap document.

SRCNet Operations Milestone	Connected SKA milestone	Added SRCNet Functionality
SRCNet v0.1 First quarter 2025	Opportunity to engage SRCNet with AA0.5 data transfer and access	Data discovery, access and replication services Interactive analysis portal (notebooks) Test data (including precursors)
SRCNet v0.2 First quarter 2026	AA1 and Commissioning	Data dissemination using interfaces at telescope sites SDP workflows runnable in the SRCs Preparation of SRCNet User Support

[1] a policy to determine what parameters and metrics make a data product unpopular is to be determined.

SRCNet v0.3 4th quarter 2026	Cycle 0 proposals, AA2 and Science Verification	First SRCNet version required by SKAO with SKA preliminary data Science verification subset of SDP workflows Start-up of SRCNet User Support
SRCNet v1.0beta 4th quarter of 2027	Science verification and Cycle 0	Integrated portal with science analysis capabilities Complete SDP workflows runnable in the SRCs Restricted SRCNet User Support
SRCNet v1.0 First quarter 2028	Cycle 1	Operational version and complete portal with science analysis capabilities First non-proprietary data available publicly Operational SRCNet User Support