SRC | Net

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;

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	Array Assembly
AAI	Authentication & Authorisation Interface
ADP	Advanced Data Product
API	Application Programming Interface
ART	Agile Release Train
FTE	Full-time Equivalent
HPC	High-performance computing
ODP	Observatory Data Product
PFLOPS	Peta Floating-Point Operations Per Second
SDP	SKA Data Processor
SKA	Square Kilometre Array
SKAO	SKA Observatory
SRC	SKA Regional Centre
SRCNet	SKA Regional Centre Network

1 Introduction

1.1 Purpose of the document

This document presents a top-level roadmap of the development phases of the SRCNet, in line with the requirements coming from external activities to the SRCNet context (e.g. SKA construction milestones or scientific events).

An estimation of the resources required at every stage is also presented, both at the level of personnel and hardware resources. The numbers provided are estimates of the required fully pledged FTEs and resources, not covering the possible decrease in performance of the development team due to fractional FTEs or shared hardware resources.

The current top-level roadmap only covers the development, operations and resources foreseen until the first SRCNet public version. Future versions of this document will cover the next phases.

2 General SKAO Roadmap

Construction of the SKA telescopes will be delivered in phases. The first major milestone, known as Array Assembly 0.5 (AA0.5), represents the completion of six SKA-Low stations and four SKA-Mid dishes. A further three Array Assembly milestones (AA1, AA2, AA*) will complete the construction.

The commissioning process, which tests that components work together as a system, will take place as we progress through each Array Assembly stage. Science verification will also begin while the telescopes are still under construction, carrying out end-to-end tests of the system based on proposals for astronomical observations from the SKAO user community. Commissioning and science verification ensures that the telescopes meet user needs and will be carried out as each Array Assembly is constructed.

Any required adjustments should be made as early as possible in the process, minimising the risks as much as possible. Following the completion, a final Operations Readiness Review will take place.

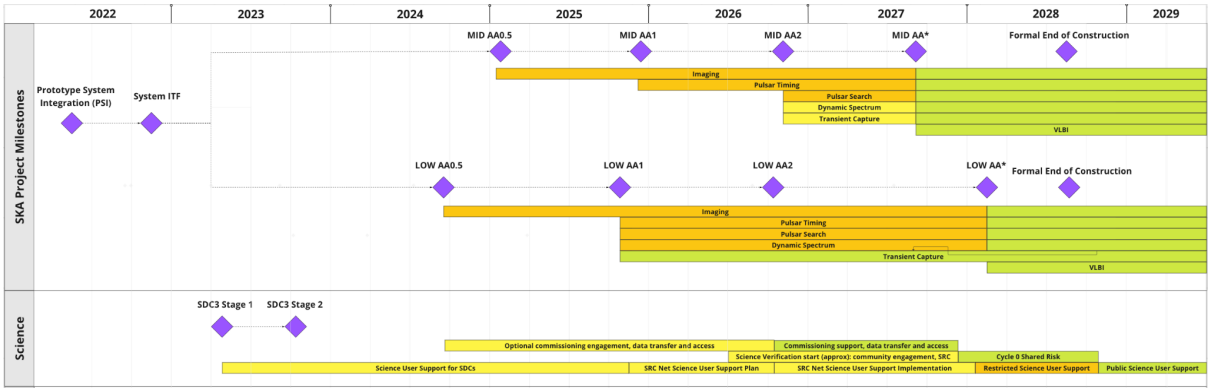

Figure 1: SKA project milestones and project dependencies for the SRCNet development

All these phases would require the deployment of software and hardware components of the SRCNet to provide support to the described activities. A parallel SRCNet development and deployment roadmap is defined below.

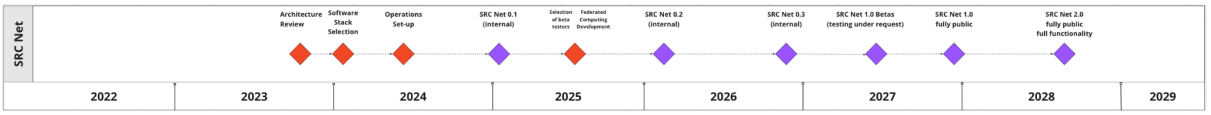

Figure 2: SRCNet development and deployment roadmap

Dates are relative to the general SKA construction milestones so any change to them will require synchronisation of the SRCNet roadmap.

2.1 Relevant Milestones Impact for the SRCNet

Milestone	Description	SRCNet Functionality	Scope (users)
SRCNet v0.1 First quarter of 2025	Opportunity to engage SRCNet with AA0.5 data transfer and access.	• Test data (and some precursor data) disseminated into a prototype SRCNet • Data can be discovered through queries to the SRCNet • Data dissemination to SRCNet 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
SRCNet v0.2 First quarter 2026	AA1 and Commissioning	• Test data (and some precursors data) disseminated into a prototype SRCNet • Data can be discovered through queries to the SRCNet • Data can be accessed through a prototype data lake • Data replication during processing • Unified Authentication and Authorisation system • 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. Storage • User storage areas • Visualisation of imaging and time series data through remote operations • Preparation of SRCNet User Support	Selected scientists from community Science Operations and Commissioning teams SRC ART members
SRCNet v0.3	Cycle 0 proposals,	• Data can be discovered through queries	Science

4th quarter 2026	AA2 and Science Verification	to the SRCNet • Data can be accessed through a prototype data lake • Data replication during processing • First Accounting model implementation. Storage • User storage areas • Unified Authentication and Authorisation system • Visualisation of imaging and time series data through remote operations • 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) • Spectral data visualisation and manipulation • Implementation of SRCNet User Support	verification community (public access) Science Operations and Commissioning teams SRC ART members
SRCNet v1.0beta 4th quarter of 2027	Science verification and Cycle 0	• SKA preliminary data (and some precursors data) disseminated into a prototype SRCNet • Data can be discovered through queries to the SRCNet • Data can be accessed through a prototype data lake • User storage areas • Unified Authentication and Authorisation system • Visualisation of imaging and time series data through remote operations • Data dissemination. Complete decision tree, including scientific program • Integrated portal with science analysis capabilities • Integrated federated computing. Workflows analysis	Increased Cycle 0 scientists Science verification scientists (public access) Science Operations and Commissioning teams SRC ART members

		• Complete SDP workflows runnable in the SRCs • Complete accounting model (storage and computational resources) • Monitoring system • Spectral data visualisation and manipulation • Data previews generation • Restricted SRCNet User Support	
SRCNet v1.0 First quarter 2028	Cycle 1	• SKA data disseminated into a prototype SRCNet • Data can be discovered through queries to the SRCNet • Data can be accessed through a prototype data lake • Data dissemination. Complete decision tree • Integrated federated computing • Complete subset SDP workflows runnable in the SRCs • Complete accounting model • User storage areas • Integrated federated execution • Unified Authentication and Authorisation system • Monitoring systems • Visualisation of imaging and time series data through remote operations • Spectral data visualisation and manipulation • Data previews generation • Full support to PI and program science tasks • Complete portal with science analysis capabilities • Public portal restricted to incoming public data • Not restricted SRCNet User Support	PIs and science program members Increased number of selected scientists from community Science Operations and Commissioning teams SRC ART members

3 SRCNet nodes simplified architecture

All the SRCNet nodes should share a common architecture view in order to present the SKA data and analysis capabilities in a consistent way. In a classical way, we make use of a three-tier architecture that comes from a presentation layer (including an API to bypass the user interface in case of scripting analysis), a server layer composed of management modules in charge of different aspects of the platform (data, metadata, resource allocation, user handling, etc), and a resources tier composed of databases, repositories and computing resources.

All the modules should have a certain level of abstraction using defined APIs so different SRCs could implement them with different software stacks (in particular, for the ones closer to the “resources tier”), although the SRCNet should produce a default implementation of all the modules so SRCs with less development time resources and appropriate implementation flexibility could just reuse and deploy them locally without significant adaptations.

The simplified architectural view shown below presents a traditional representation of the needed modules of typical science archives. The main difference with other known science platforms comes from the federated aspects applicable to the SRCNet:

- Data Lake: Data will be distributed in a data lake with repositories at different SRCNet nodes. Every SRCNet node will ensure the redundancies, backups and maintenance of their data storage.
- Authentication/Authorisation: Common authentication system is required linked to a shared authorisation management system that implements, among other things, a common security policy accessing data and computing resources
- Federated computing: Due to the SKA data size and due to the federated data lake, the SRCNet could require the execution of workflows and other execution entities into different SRCNet nodes, what requires a shared execution planner and abstraction to the pledged the computing resources. This will be implemented by the creation of an assembled interoperable computing infrastructure.
- Visualisation: Due to the size of some of the SKA data products, efficient visualisation will require a remote data parser to eliminate preliminary data movement

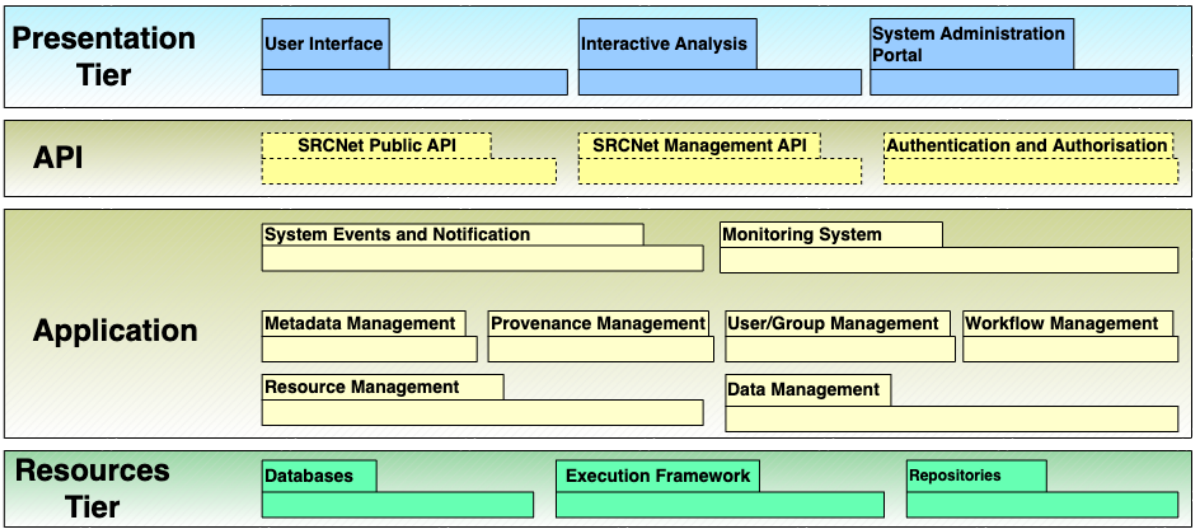

Figure 3: Condensed SRCNet node architecture footprint, including layers

SRCNet Architecture Document

3.1 Main Software Threads

This section contains a description of the main software threads that will compose the SRCNet, in order to have a top-level vision of the roadmap. The implementation of these software threads usually involves the implementation of different modules and integrations of APIs. In this document, we will not decompose the involved modules as they could vary due to the implementation approach. For a full description of all the modules that compose the SRCNet and SRCNet nodes, please refer to the SRCNet Architecture Document.

3.1.1 SRCNet Portal

The SRCNet Portal is composed of all the possible user entry points to the SRCNet platform. This SRC Portal will be deployed in different SRCNet nodes, allowing different extensions for different SRCNet nodes but providing common basic functionalities. The Portal will redirect users to the closest SRCNet node but users may also be allowed to select a specific node on which to work. The SRCNet Portal will be composed, among others, of a generic User Interface portal, an interactive data analysis portal, a system administrator portal, and a public and a management API that could be used by user layer applications or by clients that directly interact with it, bypassing the user interface (e.g. using a command-line client or a Virtual Observatory one).

An interactive analysis portal will be also part of the SRCNet Portal, containing a set of client libraries to interact with the data of the SRCNet and, also, a client application that allows the execution of this analysis on the SRCNet servers. A traditional implementation could be a notebooks portal where users can invoke different execution environments and analyse the data resources interactively using a limited set of programming languages.

Also, a set of specific libraries and methods to facilitate the exploration of the data should be provided and can be considered a sub-package of this one.

Also, the SRCNetPortal will allow the visualisation and analysis of SKA data. This is composed of server modules to allow client applications to connect to the system, parsing libraries at the server-side, visualisation components inside the browser, pre-computed data with preview images and maps, and, finally, visualisation applications prepared to interact with the SRCNet data lake.

3.1.2 Data Management Service

The Data Management system is composed of the different libraries in charge of the data logistics, including the distribution of data within the SRCNet, access to the data, replication of data during analysis, etc.

Composed of the Data Access Management module, responsible for access to the digital objects present in the SRCNet, and the Data Ingestion Management module responsible to ingest digital objects into the SRCNet and create requested copies across the network

3.1.3 Metadata Management Service

The metadata management system is composed of different libraries to access metadata from the SRCNet. The metadata management system uses database access methods and query languages to access them. Metadata that could be provided includes not only information related to SKA or closer missions like catalogues or observational metadata but also other metadata information relevant to the system like user and groups information, saved system events and notifications, etc.

This system also includes a metadata ingestion system that is able to parse different sources of input metadata (e.g. data files, catalogues, provenance, etc) and insert it into the metadata repository (usually a database) to be used by other systems.

The access to the metadata management system also includes a queue and connection pool system to control the access to it and it also should produce statistics metrics of the access to be consumed by the monitoring system.

Finally, a set of collaborative methods are also part of this system so the users could have their own table spaces to allow the implementation of science use cases that imply crossmatching or joining operations between tables with other SKA metadata while preserving good performance.

3.1.4 Authentication and Authorisation

All SRCNet nodes must have an authentication service entry, either a global service that connects to federated identity providers or local services implementation connecting to the global SRCNet authentication declared services.

In the case of the authorisation service, all SRCNet nodes must integrate consistent authorisation modules to ensure a secure access system following, e.g. the SKA data access policies. These policies would be a consistent combination of global policies provided by SKAO and the SRCNet with local policies at the SRCNet node (country/facility) level.

The SKA data access policies will be ensured by the use of a common module in all the SRCNet nodes. Changes in the data access rules (e.g. authorising a user or group to access a particular data set) could be done by the SRCNet Operations Team always following the agreed SKA data access policy. This team should be able to modify data access rules for data sets present in all the nodes of the SRCNet data lake.

Other kinds of authorisation policies, like access to computing resources, could have different levels for pledged resources and SRC local resources but the technologies used should be consistent or the interfaces properly abstracted to harmonise the SRCNet network as a whole. Modifications to the computing resources allocation rules could be also based on local SRCNet policies.

3.1.5 Federated Computing Service

The SRCNet Computing layer will be composed of an assembly of interoperable computing infrastructures. This layer will allow the federated execution of processes as close as possible to the data, using the best possible load-balancing of the resources available, including real-time information on the status of the resources, and directing the execution of workflows on the most appropriate computational resources (e.g. HPC systems).

This is why the need for a federated execution requires global/common services that make use of the description of the SRCNet topology (resources available per node, the status of the network, load and status of the resources, location of the input data, authorisation rules, evaluation of best possible execution plan, etc). These global/common services could be deployed in a certain number of SRCNet nodes.

All these metadata should be provided by the relevant SRCNet nodes by providing events to global services (e.g. a federated event bus) so the information can be compiled and used by a global Execution Scheduler. This submodule, part of the Workflow Management module, would decide on a good execution strategy on the SRCNet to prevent latency and good performance metrics.

Once the best possible strategy for execution is identified, the global/common workflow management system should request the execution of the process (either by a direct API invocation (active) or by using a federated event bus request to the required SRCNet node(s)).

In both cases, information on the federated execution will be provided by the execution SRCNet node(s) to the monitoring management system to update statistics, update user quotas and manifest status among others.

3.1.6 Operations Monitoring Portal

The Operations Monitor portal allows remote monitoring of the behaviour of the different SRCNet nodes and the SRCNet as a whole. That would include the invoking methods to query for load, resource allocation, network status, user notifications, etc. Also, it would include any needed server module to provide stats homogeneously and, finally, contain an interface to be used by the SRCNet Operations team to inspect the monitoring result, create reports and take actions on the system.

This is included in the System Administration Portal which will also compile other functionalities like the helpdesk, the user administration and the creation of reports.

Also, this system will take care of compiling stats on the usage of the network per user/group to handle the accounting of resources.

3.1.7 Security

The SRCNet should follow global and local security policies of different levels. For example:

- Data access rights
- Resources access rights and accounting
- GDPR and equivalent national and international rules
- National laws

To fulfil these rules, a high level of coordination between development, operations and local IT teams is essential. Identification and discussion of these policies will be done by the SRCNet Operations team, creating the relevant requirements on the system for development or implementation as procedures.

3.2 Main User Support Threads

3.2.1 Helpdesk

A Helpdesk portal is provided as an entry point for users to ask questions, report problems and ask for support. Also, the helpdesk system will contain training content like tutorials, videos and frequently asked questions (FAQ) on common problems or questions to support the community on the use of the SRCNet.

Questions from the community are distributed to a multi-profile team of radio astronomy scientists, data scientists and engineers to provide support on the different levels of the understanding of the processing of the SKA data and science workflows and in the use of the services of the SRCNet. Recurrent questions will be compiled in FAQ access and, in collaboration with the science workflow templates maintenance team, tutorials on access to the data will be created and published on this part of the portal.

3.2.2 Science Workflow Templates Maintenance

This team is in charge of creating, maintaining and compiling Science Workflows Templates and scientific use cases using the SRCNet. These templates should be properly documented and adapted to the SRCNet services for further adaptation and reuse by the scientific community. This team is composed of pure radio astronomers, experts on the SKA SDP pipelines, machine learning, HPC and SRCNet services.

4 SRCNet Roadmap

Miro board: https://miro.com/app/board/uXjVPyk3BzA=/

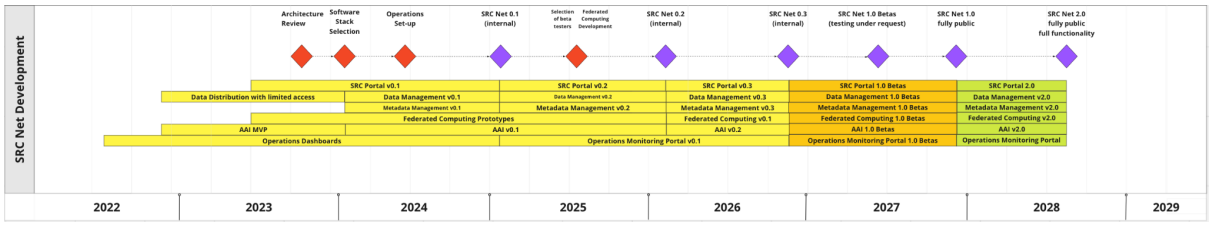

Figure 4: SRCNet development decomposition. Periods during which the principal SRCNet services are developed are indicated. Milestones imply the deployment of these services into SRCNet nodes. In yellow, are the development periods of intermediate versions. In orange, are the development periods of release candidates of fully implemented services. In green, development periods of fully functional operational services.

4.1 SRCNet Software Development View

4.1.1 SRCNet Versions Development

The development resources are expressed in FTEs during the period of the development of this specific version. These resources are corrected by several factors:

- Number of colocated efforts: Initial estimation is all the development of this particular development is in the same location. This initial estimation has been produced by comparing development efforts from other similar science platforms of other organisations with a dedicated local development team.
- Impact of geographically distributed teams: Distributed development resources are less efficient than colocated with the same hierarchical structure. **(Due to non-colocated effort) = 1.25**
- Adaptation to existing national platforms: Due to the nature of the SRCNet, we anticipate more than one implementation prototype per module. This would be higher than the expected duplication of efforts for a single development team where prototyping of new technologies only implies an overhead of 1.5 approx. This should be mitigated by a clear

division and acceptance of the distribution of work in areas by the national SRC teams. **(Expected different implementations) = 1.5**

- Organisational board effort of the development for this module, applicable for big modules, needed for the technical coordination and architectural decisions **(Organisational board) = 2**

The final calculation formula for every module would be:

Total = (Number of colocated effort)*(due to non-colocated)*(expected different implementations) + organisational board = (Number of colocated effort)*[1.25]*[1.5] + [2]

We will apply these corrections due to non-collocated teams for teams > 6 FTEs. For teams with size <= 3 FTEs, we will not add organisational board overheads.

The resources provided below are the ones needed to develop the SRCNet. Local SRCs could have bigger set-ups dedicated to national initiatives, special support to national communities, extensions of functionalities, etc. These resources would be described by the SRCs in the national proposals submitted to their respective national bodies.

4.1.2 SRCNet v0.1 Development

Version	SRCNet v0.1
Development Duration	March 2023-Nov 2024 = 20 months
Software Thread Versions	SRCNet Presentation Tier v0.1 Data Management v0.1 Metadata Management v0.1 AAI preliminary capabilities
Description	Preliminary version of the portal, enabling discovery queries on the data. That would also contain a basic local interactive analysis able to discover and access data from the SRCNet data lake, invoke local replicas and analyse data locally A preliminary version of the data management system, to allow the creation of digital object replicas on the SRCNet node selected by a user. A preliminary version of the metadata system to discover data in the data lake Minimum viable product of the authentication/authorisation system, in connection with the data management system to control access policies
Development Resources	Frontend Software Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs

	Server modules and data Management Development Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs Cloud Software Engineers - 2*1.25*1.5 = 3.75 FTEs
	Database Engineers - 3 FTEs distributed in different time zones (distributed database)
	Security Engineers - 2 FTEs
	Total 9.5 + 9.5 + 3.75 + 3 + 2 = 27.75 FTEs

4.1.3 SRCNet v0.2 Development

Version	SRCNet v0.2
Development Duration	Nov 2024 - Jan 2026 = 14 months
Software Thread Versions	SRCNet Portal v0.2 Data Management v0.2 Metadata Management v0.2 AAI v0.1 Federated Computing v0.1 Monitoring Module v0.1
Description	Enhanced version of the portal: - Data discovery portal - Embedded visualisation - Advanced interactive analysis including data processing threads Enhanced data management system: - Client library - Use of staging area for replicas - User personal storage areas and quotas Metadata management area: - Test catalogues - Table user spaces Federated Computing:

	- Simplified Federated Execution - Basic Execution Scheduler First accounting model
Development Resources	Frontend Software Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs
	Server modules and data Management Development Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs Cloud Software Engineers - 2*1.25*1.5 = 3.75 FTEs. No coordination overhead
	Database Engineers - 3 FTEs distributed in different time zones (distributed database). No corrections applicable
	Software and workflow repositories maintenance - 2 FTEs
	Computing and HPC Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs
	Total 9.5 + 9.5 + 3.75 + 3 + 2 + 9.5 = 37.25 FTEs

4.1.4 SRCNet v0.3 Development

Version	SRCNet v0.3
Development Duration	Jan 2026 - Sep 2026= 9 months
Software Thread Versions	SRCNet Portal v0.3 Data Management v0.3 Metadata Management v0.3 AAI v0.3 Federated Computing v0.1
Description	Advanced version of the portal: - Integration between discovery and analysis system - Computing resources interface - Helpdesk portal. More details at (Breen, 2021) - Embedded and spawned visualisation - Advanced interactive analysis including advanced data processing threads

	- VO services integration Enhanced data management system: - Client library - Staging Storage area - User personal storage areas and quotas Metadata management area: - Main astronomical catalogues ingestion - High-availability - Table user spaces - Provenance information Federated Computing: - Service Mesh and integration with monitoring system - Advanced Execution Scheduler - Advanced Federated Execution including HPC resources
Development Resources	Frontend Software Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs
	Server modules and data Management Development Engineers - 6*1.25*1.5 + 2 = 13.25 FTEs Cloud Software Engineers - 3*1.25*1.5 = 5.625 FTEs
	Database Engineers - 3 FTEs distributed in different time zones
	Software and workflow repositories maintenance - 2 FTEs distributed in different time zones
	Computing and HPC Engineers - 5*1.25*1.5 + 2 = 11.375 FTEs
	Total 9.5 + 13.25 + 5.625 + 3 + 2 + 11.375 = 44.75 FTEs

4.1.5 SRCNet v1.0 Betas Development

Version	SRCNet v1.0 Betas
Development Duration	Sep 2026 - Nov 2027 = 13 months

Software Thread Versions	SRCNet Portal v1.0 Beta Data Management v1.0 Beta Metadata Management v1.0 Beta AAI v1.0 Beta Federated Computing v1.0 Beta Operations Monitoring v1.0 Beta
Description	Complete version of the portal: - Integration between discovery and analysis system - Computing resources interface - Helpdesk portal - Embedded and spawned visualisation - Complete list of interactive analysis including advanced data processing threads - VO services integration Enhanced data management system: - Client library - Buffer Storage area - User personal storage areas and quotas Metadata management area: - Main astronomical catalogues ingestion - High-availability - Table user spaces - Provenance information Federated Computing: - Service Mesh and integration with monitoring system - Advanced Execution Scheduler - Advanced Federated Execution including HPC resources Administration System: - Events and Stats dashboards - Portal for operational procedures
Development Resources	Frontend Software Engineers - 4*1.25*1.5 + 2 = 9.5 FTEs
	Server modules and data Management Development Engineers - 6*1.25*1.5 + 2 = 13.25 FTEs

	Cloud Software Engineers - 3*1.25*1.5 = 5.625 FTEs
	Database Engineers - 3 FTEs distributed in different time zones (distributed database)
	Software and workflow repositories maintenance - 2 FTEs/month distributed in different time zones (distributed database). No corrections applicable
	Computing and HPC Engineers - 6*1.25*1.5 + 2 = 13.25 FTEs
	Total 9.5 + 13.25 + 5.625 + 3 + 2 + 13.25 = 46.625 FTEs

4.1.7 SRCNet Development Summary Table

March 2023 - November 2024	Number Of FTEs
Frontend Software Engineers	9.5
Server modules and data Management Development Engineers	9.5
Cloud Software Engineers	3.75
Database Engineers	3
Security Engineers	2
Total	27.75
November 2024 - January 2026	Number Of FTEs
Frontend Software Engineers	9.5
Server modules and data Management Development Engineers	9.5
Cloud Software Engineers	3.75
Database Engineers	3
Software and workflow repositories maintenance	2
Computing and HPC Engineers	9.5
Total	37.25
January 2026 - September 2026	Number Of FTEs
Frontend Software Engineers	9.5
Server modules and data Management Development Engineers	13.25

Cloud Software Engineers	5.625
Database Engineers	3
Software and workflow repositories maintenance	2
Computing and HPC Engineers	11.375
Total	44.75
September 2026 - November 2027	Number Of FTEs
Frontend Software Engineers	9.5
Server modules and data Management Development Engineers	13.25
Cloud Software Engineers	5.625
Database Engineers	3
Software and workflow repositories maintenance	2
Computing and HPC Engineers	13.25
Total	46.625

4.2 SRCNet Software Operations and Science Support

4.2.1 SRCNet Operations

Every SRCNet version implies a deployment for testing and use of the involved community (as described in 2.1 Relevant Milestones Impact for the SRCNet). The maintenance and support of these versions would include:

- SRCNet Operations team: Engineers in charge of the maintenance of the versions deployed. This team would incorporate a core team and personnel from different SRCs that could also have development roles. Remote operations on the SRCNet (e.g. procedures to recover systems or deployments of new versions) should be encouraged although some tasks would need to be executed locally. Example of services covered by the operations team:
    - Data Distribution: The operations team should be responsible for managing the distribution of the data produced by the SKA telescopes, including storing and, archiving into the SRCNet.
    - Data Access: The operations team should be responsible to maintain the access of the SKA data into the SRCNet following agreed policies.
    - Data Processing: The operations team should be responsible to maintain the data processing modules.
    - The operations team should be responsible for the monitoring of the SRCNet and the execution of health and safety procedures of SRCNet components.
    - The operations team should be responsible for the coordination of maintenance activities of the SRCNet nodes.
    - The operations team should be responsible for the handling of routine operations like the maintenance of users/groups quotas, creation of dashboards and stats,

coordination with IT teams on upgrades and security patches, maintenance of software repositories, etc (Quinn et al., 2020)

- SRCNet IT Team: Information technology engineers in charge of providing and maintaining hardware and network infrastructure for the SRCNet versions.
- Helpdesk Team: Scientists and engineers providing support to the relevant community of every version as explained in the scope column of 2.1 Relevant Milestones Impact for the SRCNet. Description of the function and concept of the helpdesk can be found at (Breen, 2021)
-

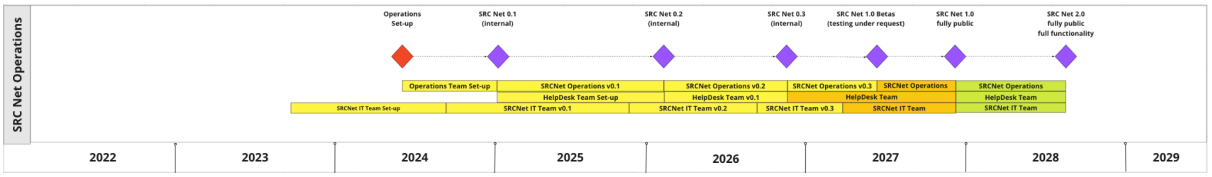

Figure 5: SRCNet operations decomposition. Every SRCNet deployment requires maintenance and support from the operations engineers, IT engineers and helpdesk members. During periods marked in yellow, a limited number of members with a limited allocation time form the teams. During periods marked in orange, the teams can be considered operative teams with capacity limitations. During periods marked in green, teams are operative and have enough members and allocation time to support the public community.

4.2.1 Local SRCNet Operations for running a node

From the previously described operations, two different types of resources could be identified:

- Participation in the global SRCNet Operations, as part of the Core Team: This team is in charge of the maintenance of the SRCNet global services and deployed components of the network. They are also in charge of module deployment, maintenance operations, definition of operational procedures, and technical support to the helpdesk team.
- Local Operations members: Teams in charge of the configuration of storage, local IT operations, network management, computing maintenance and, in general, all the maintenance activities that should be performed locally at the SRC Sites.

The nature of both sets of operations personnel is quite different due to the nature of the SRCNet. The first team are pure pledged resources provided by the different SRCs to the SRCNet to enable the SRCNet operations. The second team could be provided by the SRCs by delegation into national data centres, subcontracts to companies, sharing resources with other national projects, etc

In this section, the numbers in the different sections are only applicable to the first team (Participation in the global SRCNet Operations). However, we can enumerate some of the tasks that are applicable to the second team to help with the local resources required to run an SRCNet node. These tasks are

- Database Administration: The SRCNet nodes will have databases deployed to provide services to the SRCNet. These databases will be synchronised at several time zones and they need to be administered locally.
- AAI interfaces maintenance: Maintenance of the local communities registration and authentication and authorisation services maintenance and deployment.
- Transfer and network operations: Ensure connectivity and data transfers between nodes and network connections to other SRCNet nodes support.
- Data Management (storage): Configuration and management of the storage endpoints associated to the SRCNet data lake.
- Computing Resources and API: Deployment and maintenance of the computing resources available for the SRCNet node.
- Contribution to Monitoring: Implementation of monitoring services and provision of reports for internal services state.

As said, SRCs could provide these resources in many different ways depending on the national infrastructures so the estimation of FTEs needed to run an SRCNet node is difficult to estimate but, considering around 0.5 FTE for every skill/task, running a local SRCNet node could be estimated around 2.5-4 FTEs.

4.2.2 SRCNet v0.1 Operations

Operations Duration	Nov 2024 - Jan 2026 = 14 months
Description	- Set-up of the first SRCNet Operations Team composed of members of the ART development teams - Set-up of the first SRCNet IT Community of practices composed of IT members of the different SRCs - Services maintenance of version v0.1 - Data distribution and redistribution for v0.1 and preparation of environment for version v0.2 - Support on the deployment and maintenance of agreed computing resources (e.g. orchestrators, storage configuration, etc) - SRCNet testing, monitoring and accounting - Definition and execution of maintenance and recovery procedures
Operations Resources across the SRCNet	Operations Engineers Core Team - 4 FTEs
	Total = 4 FTEs [1]

[1] Estimations only for global operations. Running local nodes required local resources as per 4.2.1 Local SRCNet Operations for running a node. At this stage, 2 SRCNet nodes are expected to be deployed

4.2.3 SRCNet v0.1 Science Support

Operations Duration	Nov 2024 - Jan 2026 = 14 months
Description	- Set-up of the first SRCNet Helpdesk Team composed of scientists and engineers of the different SRCs - Templates generation for SRCNet compatible science workflows
Science Support Resources across the SRCNet	Helpdesk Engineers - 1 FTEs (fractions of scientists and engineers from different SRCNet nodes)
	Data Scientists - 5*1.25*1.5 + 2 = 11.375 FTEs
	Total = 1 + 11.375 = 12.375 FTEs

4.2.4 SRCNet v0.2 Operations

Operations Duration	Jan 2026 - Sep 2026 = 9 months
Description	- Services maintenance of version v0.2 - Data distribution and redistribution for v0.2 and preparation of environment for version v0.3 - Helpdesk support to Cycle 0 proposals, AA2 and Science Verification - Support on the deployment and maintenance of agreed computing resources (e.g. orchestrators, storage configuration, etc) - SRCNet testing, monitoring and accounting - Definition and execution of maintenance and recovery procedures
Operations Resources across the SRCNet	Operations Engineers Core Team - 6 FTEs

	Total = 6 FTEs [2]

4.2.5 SRCNet v0.2 Science Support

Operations Duration	Jan 2026 - Sep 2026 = 9 months
Description	- Helpdesk support to Cycle 0 proposals, AA2 and Science Verification - Templates generation for SRCNet compatible science workflows for science verification
Science Support Resources across the SRCNet	Helpdesk Engineers - 2 FTEs (fractions of scientists and engineers from different SRCNet nodes)
	Data Scientists - 5*1.25*1.5 + 2 = 11.375 FTEs
	Total = 2 + 11.375 = 13.375 FTEs

4.2.6 SRCNet v0.3 Operations

Operations Duration	Sep 2026 - Nov 2027= 13 months
Description	- Services maintenance of version v0.2 - Data distribution and redistribution for v0.2 and preparation of environment for version v0.3 - Support on the deployment and maintenance of agreed computing resources (e.g. orchestrators, storage configuration, etc) - SRCNet testing, monitoring and accounting - Definition and execution of maintenance and recovery procedures
Operations Resources across the SRCNet	Operations Engineers Core Team - 8 FTEs

[2] Estimations only for global operations. Running local nodes required local resources as per 4.2.1 Local SRCNet Operations for running a node. At this stage, 2 SRCNet nodes are expected to be deployed and maintained

	Total = 8 FTEs [3]

4.2.7 SRCNet v0.3 Science Support

Operations Duration	Sep 2026 - Nov 2027= 13 months
Description	- Helpdesk support to AA1 and Commissioning - Templates generation for SRCNet compatible science workflows for AA1 and commissioning
Science Support Resources across the SRCNet	- Helpdesk Engineers - 2.5 FTEs (fractions of scientists and engineers from different SRCNet nodes)
	- Data Scientists - 6*1.25*1.5 + 2 = 13.25 FTEs
	Total = 2.5 + 13.25 = 15.75 FTEs

4.2.8 SRCNet v1.0 betas Operations

Operations Duration	Nov 2027 - Jun 2028 = 7 months
Description	- Services maintenance of version v0.3 - Data distribution and redistribution for v0.3 and preparation of environment for version v1.0beta - Helpdesk support to pre-Cycle1 - Deployment and maintenance of agreed computing resources (e.g. orchestrators, storage configuration, etc) - SRCNet testing, monitoring and accounting - Definition and execution of maintenance and recovery procedures
Operations Resources across the SRCNet	Operations Engineers Core Team - 10 FTEs

[3] Estimations only for global operations. Running local nodes required local resources as per 4.2.1 Local SRCNet Operations for running a node. At this stage, 4 SRCNet nodes are expected to be deployed and maintained

	Total = 10 FTEs [4]

4.2.9 SRCNet v1.0 betas Science Support

Operations Duration	Nov 2027 - Jun 2028 = 7 months
Description	- Helpdesk support to Science verification and Cycle 0 - Templates generation for SRCNet compatible science workflows for a complete set of science use cases
Science Support Resources across the SRCNet	Helpdesk Engineers - 3 FTEs (fractions of scientists and engineers from different SRCNet nodes)
	Data Scientists - 8 * 1.25 * 1.5 + 2 = 13.25 FTEs
	Total = 3 + 13.25 = 16.25 FTEs

4.2.10 SRCNet Complete Operations

Operations Duration	June 2028 -
Description	- Services maintenance of version v1.0 public version - Data distribution and redistribution for v1.0 public version - Helpdesk support to Cycle 1 - Deployment and maintenance of agreed computing resources (e.g. orchestrators, storage configuration, etc) - SRCNet testing, monitoring and accounting - Definition and execution of maintenance and recovery procedures
Operations Resources across the SRCNet	Operations Engineers Core Team - 5 FTEs Operations Core Team
	Total = 10 FTEs [5]

[4] Estimations only for global operations. Running local nodes required local resources as per 4.2.1 Local SRCNet Operations for running a node. At this stage, up to 8 SRCNet nodes are expected to be deployed and maintained

[5] Estimations only for global operations. Running local nodes required local resources as per 4.2.1 Local SRCNet Operations for running a node. At this stage, up to 16 SRCNet nodes are expected to be deployed and maintained. This number of nodes is defined considering an initial approach of 1 node per country, although technical architecture enables alliances to create multinational nodes

4.2.11 SRCNet v1.0 Complete Science Support

Operations Duration	June 2028 -
Description	- Helpdesk support to Cycle 1 - Templates generation for SRCNet compatible science workflows for a complete set of science use cases
Science Support Resources across the SRCNet	Helpdesk Engineers - 5 FTEs (fractions of scientists and engineers from different SRCNet nodes)
	Data Scientists - 8 * 1.25 * 1.5 + 2 = 13.25 FTEs
	Total = 5 + 13.25 = 18.25 FTEs

4.2.12 SRCNet Operations Summary Table

November 2024 - January 2026	Number Of FTEs
Operations Engineers Core Team	4
Total	4
January 2026 - September 2026	Number Of FTEs
Operations Engineers Core Team	6
Total	6
September 2026 - November 2027	Number Of FTEs
Operations Engineers Core Team	8
Total	8
November 2027 - June 2028	Number Of FTEs
Operations Engineers Core Team	10
Total	10
June 2028 -	Number Of FTEs
Operations Engineers Core Team	10
Total	10

4.2.13 SRCNet Science Support Summary Table

November 2024 - January 2026	Number Of FTEs
Helpdesk Engineers	2
Data Scientists	11.375
Total	13.375
January 2026 - September 2026	Number Of FTEs
Helpdesk Engineers	2
Data Scientists	11.375
Total	13.375
September 2026 - November 2027	Number Of FTEs
Helpdesk Engineers	2.5
Data Scientists	13.25
Total	15.75
November 2027 - June 2028	Number Of FTEs
Helpdesk Engineers	3
Data Scientists	13.25
Total	16.25
June 2028 -	Number Of FTEs
Helpdesk Engineers	5
Data Scientists	13.25
Total	18.25

4.3 SRCNet Software Hardware View

4.3.1 SRCNet Versions Hardware

On the compute and storage resources, we will just indicate a certain percentage of implementation during different stages of development. The best global numbers indications for the SRCNet are (Bolton & Chrysostomou, 2018) (Bolton, 2021) (Hughes-Jones et al., 2019). As these figures will be updated, we will maintain the best-known calculations (Ratcliffe & Bolton, 2021) that, at the present time are:

- Sustained Compute (incl. efficiencies): 21.6 PFLOPs
    - 8.1 PF Re-processing + 13.5 PF Post-processing

    - In estimations, execution peaks of 35 PFLOPs were found (see (Bolton & Chrysostomou, 2018))
- Storage growth (steady state, usable capacity): 707 PBytes per year
    - 442 PB ODP + 265 PB ADP

It is expected that a gradual deployment of the SRCNet storage and computing resources will take place during the development phase in preparation for the public version. Availability requirements of the resources will be increased as per agreement between partners from the early phases to the public version, where an appropriate availability percentage will be used. The next tables show a proposed deployment plan using a percentage of the total needs for the public version, following a semi-exponential profile and percentual to the national share of the SKA project.

4.3.2 SRCNet Storage

The SRCs will store the SKA data products and make them available to users for analysis. High-performance storage is relatively expensive so typically sites adopt a hierarchical storage architecture with a limited amount of high-performance "buffer" storage (e.g. Solid State Drives (SSD)) (Tier 0), close to computing units to ensure high performance.

The "online" storage will be composed of most of the SKA data needed for analysis. The technology used for this storage will be a limited number of storage elements of Tier 0 and, mostly, cheaper, slower alternative technologies (Tier 1). The SDP cost model has assumed a costing based on SATA Hard Disk Drives, and we do the same here.

Due to budget limitations, a colder storage layer was also included (Tier 2) for old data products or data not usually needed for processing. A proposal of how to define these Tiers is presented in section 4.3.2.1 SRCNet Storage Tiers. The exact definition of what is present in the different layers could be modified in the future due to national budgets, storage prices and scientific analysis requirements. It has to be considered that data products in Tier 2 will degrade I/O access performance, having an impact on the science analysis so

We will cover here only a reservation of space of one year of ODPs (assuming two copies) at the moment of the first public version of the platform. As described in (Bolton & Chrysostomou, 2018) and (Hughes-Jones et al., 2019), some of the relevant assumptions applicable to storage extracted for this document (and applicable for the preparation of the first operations year) are:

1. We assume that image Observatory Data Products (from the SDP) are not deleted
2. We assume that image Observatory Data Products are used in the SRCs to generate Advanced Data Products. There are certain discrepancies in the percentage of the ADPs size in relation to the ODPs size but we maintain the modest calculation of 442 PB ODP + 265 PB ADP per year present on the confluence page (Ratcliffe & Bolton, 2021). As this document only covers the period until the first public version of the SRCNet, we are considering this limited percentage, creating a reservation of one year of storage for the first full operations

year. Corrections during the first year of operations could be needed if this assumption is exceeded

3. We assume that there are 2 copies of every data product within the global network of SRCs
4. We assume all data products are losslessly compressed with compression factors of 2. This would compensate point 3 of this table
5. In the following table, we could consider that one copy is “online” storage and the secondary copy could be in a cooler SATA storage
6. Finally, we apply a reservation margin[6] to cover possible significant deviations from this calculation of 50%. That implies total storage of 1060 PB for the first year.
7. These calculations are on usable storage, after possible reductions due to a selected RAID configuration.

[6] The initial reservation margin will cover early ADPs generation (final percentage of ADPs vs ODPs is still under debate), temporary data due to replicas invocations done by users to work locally (before a federated execution is implemented), extra replicas for popular datasets and testing environments. This could be refined in later phases when exact knowledge on the user needs and data creation is obtained.

		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

Figure 6: SRCNet storage allocations. Deployment of the SRCNet storage in line with the share per country and considering two copies of the SKA data. These figures are just indicative considering a possible share of the countries on the SKA construction (these figures could change with new partners or new sharing percentages) and a deployment plan until the first public version of the SRCNet. This table only covers a reservation of storage for the first year of operations (one copy in hot storage and a secondary copy in cooler storage). A possible increment of the percentage of ADPs vs ODPs could be corrected by a gradual increment of the storage during the operations phase. It is expected that the online storage (hot storage) will be maintained more or less stable during operations at 530 PB (depending on the ADP production rate), moving data to the near-line storage for data products older than one year, which implies a global increment of 1060 PB of the near-line storage (cold storage) per year.

4.3.2.1 SRCNet Storage Tiers

Whenever the data is older than one year, data would be moved to the hot storage to colder storage. In general, the SRCNet Storage will follow a paradigm similar to a Hierarchical Storage Management (HSM), making use of different storage tiers with different performances. The definition of rates between storage technologies per node will be done in a coordinated way depending on performance, budget and other boundary conditions (e.g. energy consumption). A proposed tier structure could be as follows:

- **Tier 0 - Cached data:** Technology SSD. In principle, due to its costs, this tier could be only used to temporarily cached data accessible for ongoing analysis. Also, this hot storage layer is the one used for buffering for certain data analyses. Although, ideally, the primary copy of some of the popular data would be in this kind of tier, it would be optional to be present in all the SRCNet nodes due to its costs. If during the SKA project evolution, the costs of SSD storage are more affordable, this layer could be extended to include also popular datasets.
- **Tier 1 - Online Storage:** Technology HDD. Two copies (to allow high availability) in two different SRCNet nodes of the last year of data. That includes ODPs and ADPs. The storage size needed for this tier is in line with the calculations done in the previous section (around 1060 PBs in the whole SRCNet data lake). As said for Tier0, the technology used for this Tier could be evolved in future extensions and upgrades whenever possible.
- **Tier 2 - Near-Line Storage:** Cold storage technology that could include tapes if it is not affordable a hotter technology. Data older than one year will be moved to this cold storage tier to leave space for newer data. As tapes are not optimal for science analysis, the type of technology will involve a balance between science analysis performance and the budget available at the different SRCNet nodes.

4.3.3 SRCNet Computing

As per (Bolton & Chrysostomou, 2018), we assume that data products are processed multiple times in the SRCs, initially up to 20 times (whilst the arrays are small the data rates are so low that this is not a driver) but reducing to 3 times only once the arrays are fully built. Using the SDP parametric model, this reference estimates the total computing of the SRCNet to be around 35 PFLOPs. For good performance, high-performance storage close to the computing units could be needed (“hot” buffer) as described in 4.3.2.1 SRCNet Storage Tiers.

		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

Figure 7: SRCNet computing resources. Deployment of the SRCNet computing resources in line with the share per country. These figures are just indicative considering a possible share of the countries on the SKA construction (these figures could change with new partners or new sharing percentages) and a deployment plan until the first public version of the SRCNet.

4.3 Expected resources profiles

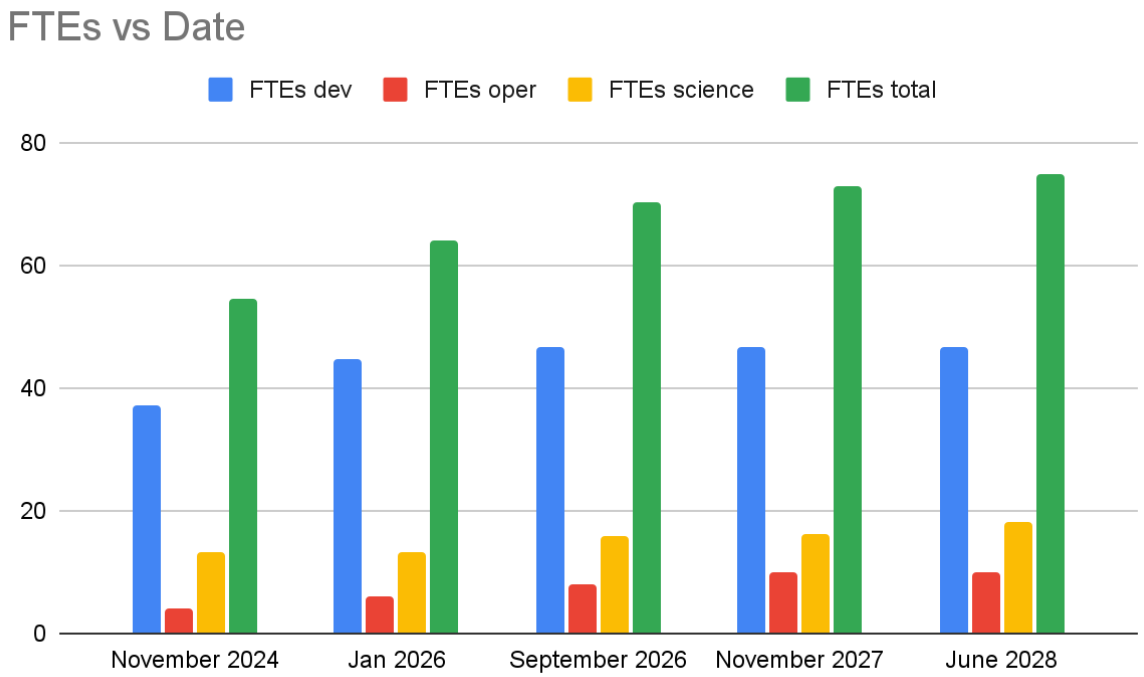

Figure 8: Profile of FTEs/months for the different implementation phases until SRCNet version 1.0. Numbers for the operation and upgrade of the SRCNet are maintained stable after version 1.0. Local Operations FTEs are not presented in this chart, what would increase the operations support

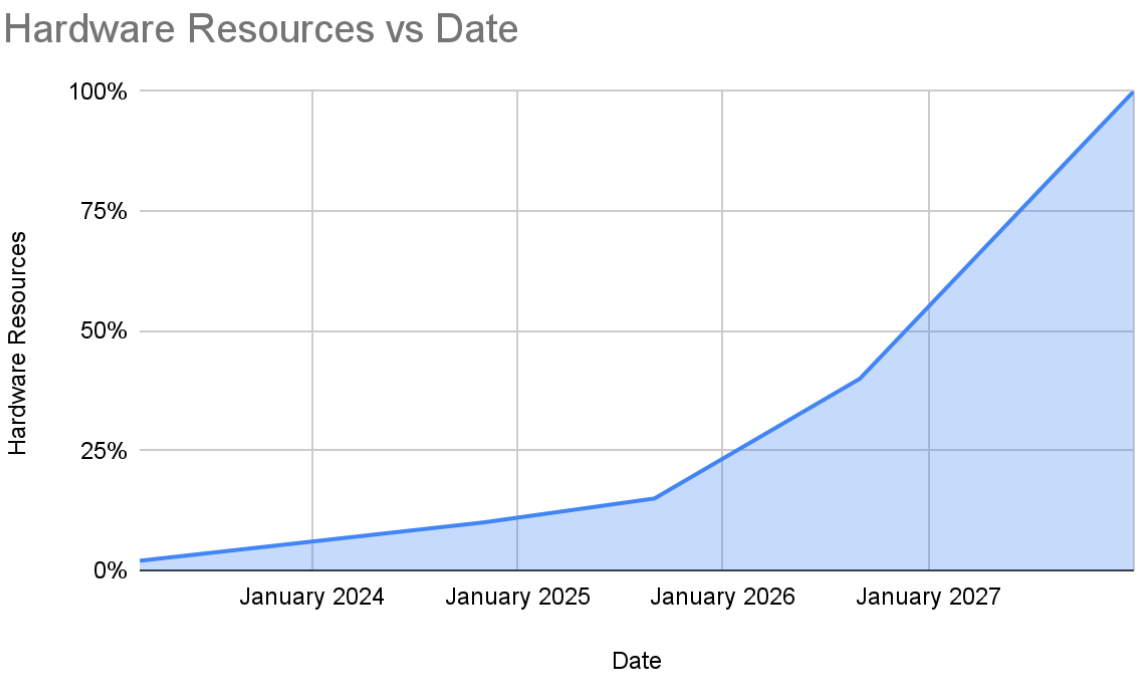

Figure 9: Profile of hardware resources percentage relative to the one required for the first completed version during the different implementation phases until SRCNet version 1.0

A References

Bolton, R., & Chrysostomou, A. (2018, 12 03). *A Model for SKA Regional Centre Global Size Estimation*. https://astronomers.skatelescope.org/wp-content/uploads/2019/05/SKA-TEL-SKO-0001000-01_AModelforSKARegionalCentreGlobalSizeEstimation-signed.pdf.

Breen, S. (2021, 03 16). *SKAO Helpdesk Concept*. https://ska-aw.bentley.com/SKAProd/Search/QuickLink.aspx?n=SKA-TEL-SKO-0001795&t=3&d=Main%5ceB_PROD&sc=Global&r=01&i=view

Hughes-Jones, R., Vicinanza, D., & Rayner, T. (2019). Architecture and cost model for a World-wide network for SKA. *AENEAS – H2020 – 731016*, *AENEAS GA DELIVERABLE*(D4.4). https://cordis.europa.eu/project/id/731016/results

Quinn, P., van Haarlem, M., An, T., Barbosa, D., Bolton, R., Chrysostomou, A., Conway, J., Gaudet, S., Klöckner, H.-R., Possenti, A., Ratcliffe, S., Scaife, A., Verdes-Montenegro, L., Vilotte, J.-P., & Wadadekar, Y. (2020, May). *SKA Regional Centres White Paper*. https://aussrc.org/wp-content/uploads/2021/05/SRC-White-Paper-v1.0-Final.pdf

Ratcliffe, S., & Bolton, R. (2021). *System Sizing, Costing and Scaling*. System Sizing, Costing and Scaling. Retrieved Mar 6, 2023, from https://confluence.skatelescope.org/display/SRCSC/System+Sizing%2C+Costing+and+Scaling