SRC | Net

SKAO Regional Centre Network

SRCNet Operational Concept

SRC-0000006	Revision 01B
Classification:	UNRESTRICTED
Document type:	PLN
Date:	2023-08-24
Status:	DRAFT

Role	*Name*	*Designation*	*Affiliation*	*Signature*	*Date*
Author	Rosie Bolton	Head of Data Operations	SKAO		
Author	Shari Breen	Head of Science Operations	SKAO		
Owner	Antonio Chrysostomou	Deputy Director of Operations	SKAO		
Approver	Michiel van Haarlem	On behalf of the SRCSC	ASTRON		
Released by	Lewis Ball	Director of Operations	SKAO		

TABLE OF CONTENTS

LIST OF FIGURES

No table of figures entries found.

LIST OF TABLES

No table of figures entries found.

1 Introduction

1.1 Purpose of the Document

This SRCNet Operational Concept document is intended to provide a first overview of the Operations of the SRC Network. This document has not been extensively reviewed by the SRCSC and should be considered as a mature draft.

1.2 Scope of the Document

The operational model for the global SRC Network ("SRCNet") has been discussed previously in the Observatory Establishment and Delivery Plan (OEDP) [AD1], with the endorsement of SKAO Council. In the time since that document was written the SRC model has matured - most significantly with the development of a set of SRCNet requirements, principles, a documented global architecture, a science use case analysis, and a full year of SRCNet software prototyping. In this document we revisit the SRC Operational model in light of these developments, noting where the SRCNet Principles [AD3] help inform the operational reasoning.

2 Operational concept

2.1 The Functions of SKA Regional Centres

The SRCs are logical and/or physical centres through which users will interact with the SKA Data Products (both ODPs and ADPs). The existence of SRCs is required both by SKA and by the users. They are operationally critical as part of the data flow from SKA to users and they are mission-critical for the delivery of science results from SKA data.

Here the major functions provided by the global network of SRCs are outlined, split into those SKA-facing functions towards which SKA will play an operational role, and user-facing functions that are more decoupled from SKA operations, but which are essential to the SRC ecosystem.

2.1.1 Direct SKA facing functions

Data Ingestion	SKA must be confident that science data products generated within each SDP can be interrogated by users without impacting the operation of the SDPs themselves – that is to say, users will not have access to the SDPs,

	since these are a fundamental and schedulable part of each SKA telescope. The Observatory must be confident when scheduling a particular Scheduling Block (SB) that the capability to ingest the data products is available somewhere across the SRCNet and that there is sufficient data network bandwidth available to actually perform the transfers. This is crucial to maintain the throughput of the Observatory's science programme. This is because the space in the data staging area of the SDPs is limited, and it would be very inefficient (and slow) to need to copy data out of the SDP's long term preservation systems rather than transferring data products whilst they still reside in the relatively fast data staging area of the SDP's data buffer. Ultimately, if an SRC cannot be identified as able to receive SKA data products as they are created in the SDP, there could be scheduling constraints imposed on the telescopes themselves.
Project-level Data Product Generation	Some projects will require only Observation-level data products (OLDPs), an ODP generated by SDP workflows based on data obtained from a single execution of a SB. Other projects may need, for example, deep images or stitched mosaics generated by combining several, related, OLDPs, delivering the requirements of the PI as outlined in their original proposal. Workflows required to generate these project-level data products (PLDPs) can be well-defined in advance, with estimates of the hardware resources needed to perform them. We anticipate that these workflows will be carried out within the SRCNet, by SKA Observatory staff. This will be done using software pipelines that the SKA Observatory develops and is responsible for. This is important to ensure data product integrity and traceability and for SKAO to be able to declare a project successfully complete. For example, if the Observatory gives time to a project requiring an image cube of sensitivity equivalent to 100hrs, then the SKAO has a responsibility to ensure that after committing telescope resources to collect the data, images of the appropriate quality are indeed created and made available to project teams in a timely manner.

There is also a suite of functions required in order to enable the SRCs to share the burden of storing data products and providing archive resources and supporting users for which collaboration between the SRCs can greatly increase overall efficiency and reduce cost. SKAO anticipates performing a coordinating, operational role in the day-to-day work needed to manage the network of SRCs, in collaboration with colleagues employed at SRCs (through the SRC operations group – see §4).

2.1.2 SRCNet operations functions

Data Indexing and Management	The SKAO and SRCs must have visibility of the location of each copy of SKA data products and allow a centralised data management service to manage these copies as required. All SKA science data products (observation-level, project-level and advanced (user-generated) data products) will be included in an index of data products that can be accessed by search tools.
SRC Service capability monitoring and accounting	Monitoring the availability and performance of services provided by each SRC, reporting service faults and tracking overall performance against pledged resources. Monitoring and maintaining central services.
Wide Area Network Health Monitoring and incident response	Checking the health and performance of the wide area network,fault reporting and responding to incidents.
SRC Common Software Management	To coordinate the development and deployment of shared software across the SRCs, and software to support SRC-collaboration activities. We would separate the software distribution (a service) from the software development work (an ongoing collaboration that has already started in the form of the ART and which would continue (albeit at a lower staffing level) once full SKA/SRC operations are established.
User support: provision and tracking	A full suite of user support will be provided across the SRC Network through the pledging process. Each SRC should be expected to dedicate effort to support

	users of their services, to allow SRC-specific enquiries to be dealt with. Tracking and following up on feedback from the community (gathered through e.g. the Helpdesk, community training days, and other meetings), allowing the implementation of improvements, development of new functionality, or reporting to the SKAO as appropriate
SKA proposal technical reviews	Provide input into the technical review process of SKA proposals to ensure that the request of the Network is feasible.

In addition to the SKA-facing functions, SRCs will provide users with capabilities and support to perform their science analyses which are essential to delivering scientific return. Although the immediate scientific return of SKA is in the provision of users with the data products they need to enable the science cases as described in their observing proposals, it must be recognised that in the long term it is expected that at least as much science impact will come from the re-use of SKA data products after they become public. So, the term "user" here includes not only those PI/KSP teams with active SKA projects but also scientists making use of public SKA data.

2.1.3 User-facing functions

Users will be provided with a suite of tools, resources and support by the SRCNet to enable the efficient delivery of science. Accessibility and usability will be at the forefront of the design considerations (see P13[1]). The provided tools, resources and support include:

Provision of a Science Analysis Platform	SRCs will serve as analysis facilities for users, enabling them to develop and run science analysis workflows, and generate advanced data products, on computing

[1] The SRC Network will be committed to providing, and abiding to, accessible and equitable tools, practices and processes. *Accessibility and use of tools, and SRCNet practices and processes will be based on fairness and be unbiased towards any particular group. This is especially important for user-facing tools provided or hosted by the SRC Network. This principle should be embodied within an Accessibility Policy*.

	platform(s) provided in the SRC ecosystem. An SRCNet Science Analysis Platform Vision [RD1] has already been elaborated, and includes provision for: • Access to SKA data products (both OLDPs and PLDPs) • collaboration on shared projects, to save their workflows and preserve data and software provenance information, to have access both to interactive compute sessions on appropriate (possibly very large) scale resources, and to submit offline jobs. • Accessing appropriate visualisation packages to interrogate their large data products • SKA archive searching, multiwavelength searches • Project, resource and team management functionality • Downloading data products if small or planned • Upload ADPs for inclusion in the archive if created outside the network • Workflow template library • Project, resource and team management functionality This Science Analysis Platform will be designed with accessibility and usability in mind and will be uniformly deployed across the Network enabling a persistent user profile used regardless of which SRC is accessed
Data Search (Science Archive) Functionality	Provision of science archive functionality: Allow users to search for and extract public SKA data products (and protected data products, where they have access rights), including supporting external query requests (e.g., via IVOA protocols).

Compute allocation proposal for archival data	A process to submit a compute proposal to conduct projects on science archive data.
SRCNet compute requirement estimator	Tools to estimate required SRCNet compute to deliver science (could be based on AI taking into account similar past projects longer term). This will need to be estimated at the point of proposal and will therefore need to be integrated into the Proposal Preparation and Scheduling Tool (PPT) which forms part of the SKAO's Observatory Science Operations tool suite (i.e. it will be a shared responsibility). Needs to account for: • Creation of project level data products • Estimates of compute required to arrive at the desired, publishable ADPs which satisfy science requirements (including provision of potential "feedback loop").
Provision of workflow templates (see P14[2])	Some types of ADPs will be frequently created within the SRCNet. For these ADP workflow templates will be provided, building to a more extensive library over time. This work will be supported by workflow contributions from users which can be adapted.
General User Support	Provision of a centrally managed repository of extensive and up to date documentation, user guides, FAQs, videos and tutorials to guide users through their SKA data experience.

[2] P14: The SRC Network will provide workflow templates to carry out basic and standard processing tasks. *This will lower the bar for non-expert users to come to the SRC Network to access and use SKA data for their scientific research*.

	Current and prospective user training (e.g. SKA and SRCNet capability, proposal submission, radio astronomy, HPCs, advanced data reduction, data archive mining, astronomy tools, the SRCNet Science Analysis Platform) through hybrid meetings, workshops, town hall meetings, SKA community days, combined training events with SKAO. Facilitating user interactions through community events, meetings, and conferences. The totality of these needs will be provided jointly by SKAO and SRCNet.
Helpdesk Provision (see P5[3])	Problem-solving, proposal development and project guidance/assistance. Users should access a single Helpdesk already described in [AD2], including an up-to-date knowledgebase, staffed by both SRCNet and SKAO employees (i.e. there should be only one Helpdesk to serve the entirety of user needs).
Support workflow development	Provide assistance developing, adapting and executing workflows and pipelines (work should be managed through the Helpdesk but since this is expected to be a major effort it is noted separately).
Authentication (see P3[4])	SRCs must enable registered SKA users to use their SKA credentials for SRC

[3] P5: There will be one Helpdesk system for the SRC Network and the SKAO. *The Helpdesk will be the cornerstone of the SKAO and SRC user support model, providing end-to-end support to users on topics from proposal preparation through to the access and processing of SKA data in the SRC Network.*

[4] P3: There will be a common SKAO/SRC Network user account that allows users access to SRC Network resources. *The same account that users of the SKA will use to access SKAO resources (e.g. proposal preparation tools, Helpdesk) will provide authorised access to SRCNet resources.*

	activities, and support data access according to SKAO policies.
SDP feedback loop	Engage with SKAO staff to enable project feedback loops allowing users to refine selected SDP workflows and parameters. Some large projects may be able to justify the delivery of a small fraction of the calibrated visibility data alongside their requested ODP to test and fine tune the selected SDP pipelines (which will be available within the SRCNet) and ultimately amend the requested SDP workflow for the full project.
User notifications	Notify the users of any activity on their projects including, for example, the availability of ODPs for access, and processing job completion.

The implementation of the user-facing functions should be guided by an overall desire to streamline and simplify the user experience, to appeal to and be accessible to the broadest possible SKA user community. This includes the need to hide (or provide appropriate guidance on) complexities of the overall system as much as possible, lowering barriers to the delivery of transformational science. The SRCNet user experience should be of a consistently high standard no matter which SRC is serving their needs. It should not be obvious to the user which SRC resources they are using.

3. Roles and Responsibilities

In Figure 1, a 'swimlane' diagram is used to show where the responsibilities of the Observatory, the SRCNet, and the SKA user (whether as a PI/CoI, or an archive user) lie with respect to delivering the SKA science programme of successful projects. It shows two phases: the project execution phase and the science extraction phase. The Observatory is responsible for the project execution phase, which includes the generation, calibration, and delivery of OLDPs and PLDPs into the SRCNet. The SRCNet is then responsible for supporting the SKA community of users in extracting the science from the ODPs delivered to them for publication and dissemination.

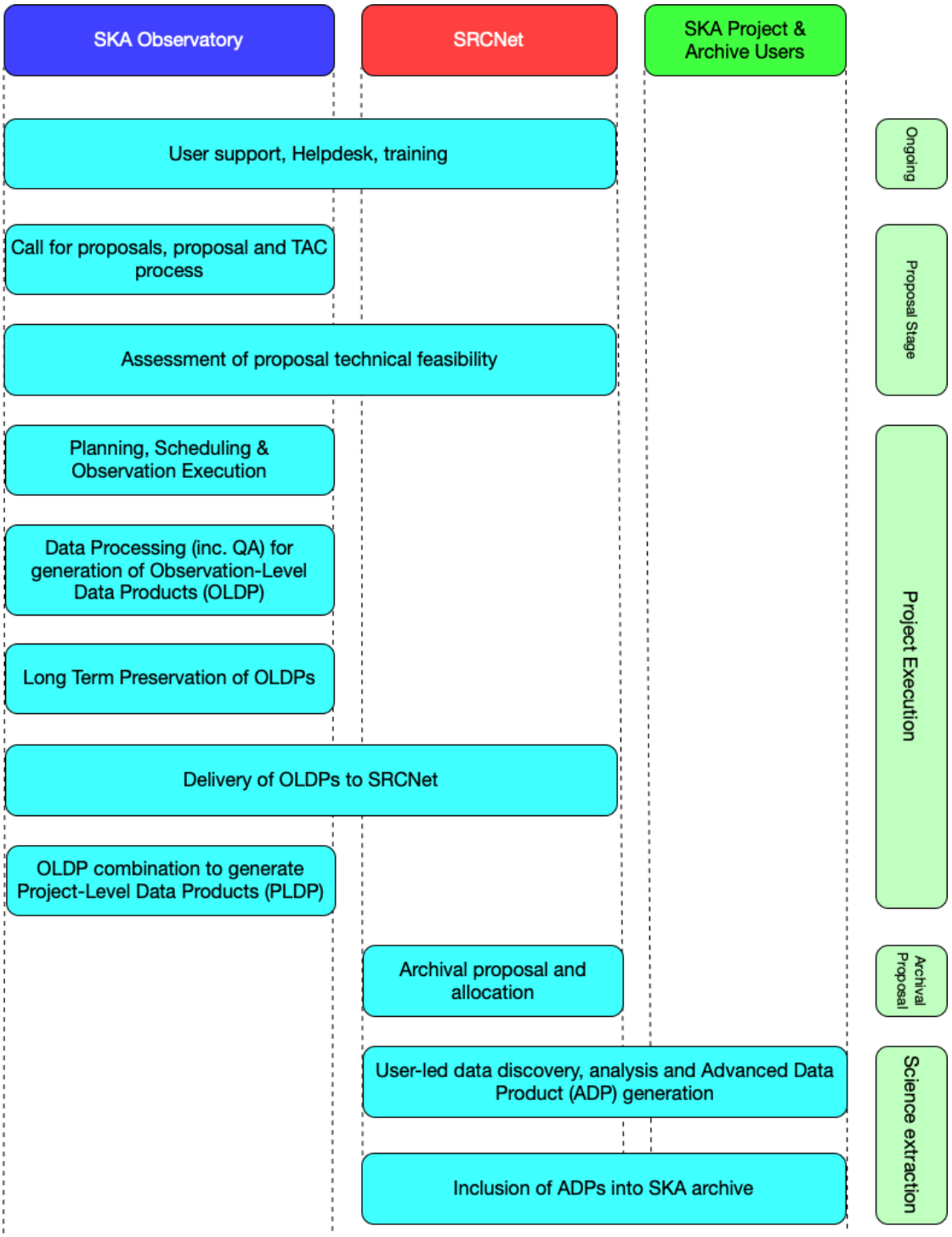

Figure 1: Swimlane diagram showing the responsibilities for the generation of science data products during the project execution and science extraction phases of a science project. The left lane depicts the Observatory's responsibilities, the middle lane those of the SRCNet, and the right lane for the PIs and Co-Is of SKA projects and general archive users. Observation-Level and project-level data products are generated by the Observatory, while advanced data products are produced by users within the SRCNet.

The science extraction phase will generally result in additional, advanced data products (ADPs, see AD1 §6.4.4.6.1 for definition) to be generated as a consequence of the advanced analysis and modelling techniques that will be employed by the science community.

The boundary between SKAO's responsibility for data products and the users' responsibility requires striking a balance between centralisation and the desire for SW quality and data traceability, and the ability to declare that the Observatory's responsibilities to a specific PI or KSP team have been met, against the somewhat competing need to encourage scientific freedom and innovation.

As Figure 1 shows, the SKAO has responsibility for data product generation up to (and including) the project level. However, though maintaining and running the software pipelines to generate the PLDPs from OLDPs will be the responsibility of the Observatory, the actual processing to generate these (for the projects that need it) will be undertaken within the SRCNet. This is a core SKAO-facing function. This extension of PLDP creation on top of SDP's OLDP generation is a relatively small change – it allows the Observatory to perform QA metrics on more complete data products, and of course, feedback difficulties and failures immediately to affect pipelines not just for combining OLDPs but also for their creation in the SDP. Since all OLDPs are in any case already delivered to the SRCs, users will also have access to those data products if they wish to combine OLDPs in other ways. This has the effect of reducing risk by bringing responsibility for PLDP generation within the SKAO's remit, without reducing opportunities for scientific innovation and user input and feedback.

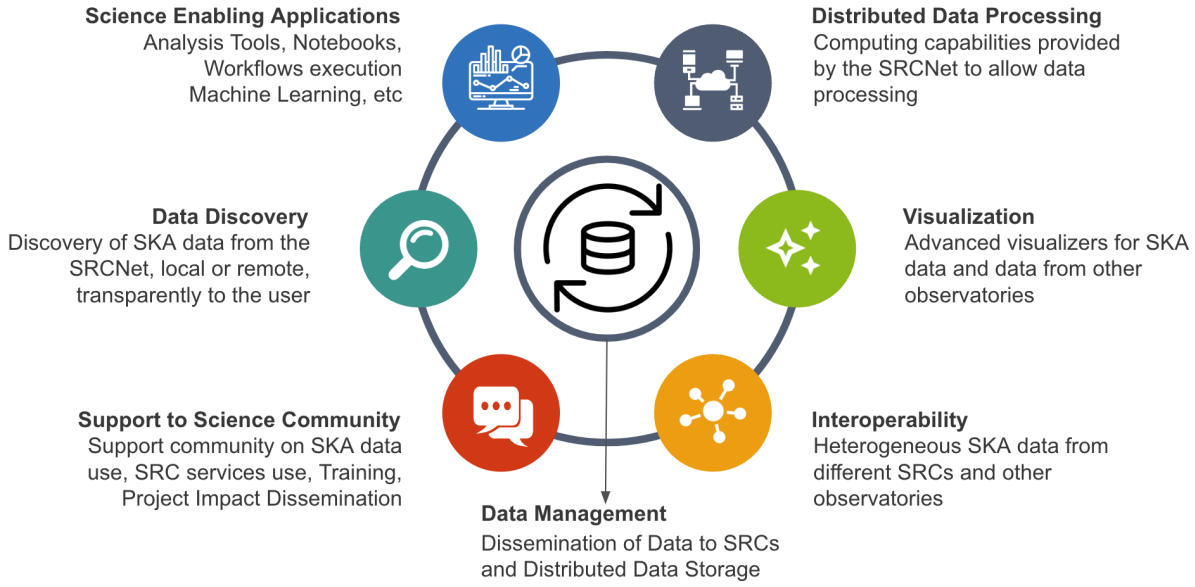

Figure 2: The SRC blueprint, reproduced from the Architecture document [AD4]

Data management, discovery and access is at the heart of the SRC concept, as shown in Figure 2 and as reflected in the architecture and in SRC principles P15[5], P18[6], P19[7], P20[8], P21[9].

As shown in the swimlane diagram (Figure 1), once the OLDPs are delivered to the SRCNet the SDP sites will store copies of them in the long term preservation (LTP) system - this is a high-latency (i.e. slow) data storage system existing only as a backup of the data products so that they can be re-delivered to the SRCNet if all copies in the SRCNet are lost. In other words, the LTP system is a back-up of last resort and not an actively managed storage element in the network of SRCs. It will not be possible to delete items from the LTP on the grounds that there are copies in place in one or more SRCs.

Within the SRCs, the management of data products can be more flexible. By agreeing to share the burden of data storage and access to users, SRCs can provide coverage of the whole SKA archive of Observatory and advanced data products without requiring that each SRC must keep a full (and fully backed-up) copy of each data product. Instead, there can be a global data management service that applies rules to data products or collections of data products and manages data transfer between SRCs, or to different storage classes in order to maintain adherence to these rules. The service can tag

---

[5] P15 The SRC Network will embrace FAIR (Findable, Accessible, Interoperable, Reusable) and Open Science principles whenever possible and appropriate. *Users will be encouraged, and supported, to include basic provenance and metadata information to abide by this principle. This includes not just the Advanced Data Products (ADPs), but also the workflows and data collections that generated those ADPs.*

[6]P18 The physical location of SKA data products will be determined to optimise access and minimise data redistribution within the Network. *The efficiency of the SRC Network will be improved if data distribution is minimal (or zero) in order to execute processing jobs for science projects. Individual SRC nodes will not be able to choose which data products they host in their pledged resources.*

[7] P19 SKA Observatory Data Products (ODPs) will be pushed from the SKA telescopes into the SRC Network. *SKA ODP ingest into the SRC Network will be managed by the SOG. Data will be delivered to individual nodes within the Network and optimised for cost and efficiency. SRC nodes will receive data at the maximum rate that they can be transferred into that node from the SKA telescopes.*

[8]P20 The lifecycle of science data products will be managed by the SRC Operations Group. *There will be a policy to determine when science data products should be moved to high-latency data storage devices rather than remain active in the SRC Network. For instance, this could be informed by the "popularity" of those data products, ie. how often they have been accessed over a period of time.*

[9]P21 Data processed within the SRC Network will automatically propagate all metadata and provenance information in support of FAIR principles. *There will be a data policy to specify required metadata and provenance information, and the auto-population of metadata within workflow processing.*

"spare" copies of data products for deletion (without necessarily performing the deletion) so that individual SRCs can clear space when resources become limited.

Using a global data management service to perform this function is essential to avoid confusion - SKAO's role in this will be to provide coordination (through operations group personnel) and the necessary software and hardware (e.g. servers) to run the data management service and maintain the catalogue of data product locations.

4. Coordination and Operations

A model for how the SRCNet will be operated and governed is under active development. There is good agreement that pledges of resources, and not financial contributions, should be put forward by representative agencies from Member countries to support SRC efforts.

In the current model, there will be an SRC coordination committee (SCC), comprising representatives from the SKAO and each Member state. The SCC will be responsible both for tracking the anticipated required SRC resources needed to support core, communal, SRC activities, and for tracking the pledging of SRC resources into this common pool.

The national representatives of each Member state will be responsible for pledges of resources and capabilities to the SKAO as part of the global SRCNet. Pledges will be made publicly, as will information on their delivery and the performance of the network as a whole. This model has worked very well for the CERN's WLCG (The Worldwide LHC Compute Grid), which operates a global collective of processing and data storage over 150 computing centres across more than 40 countries. The WLCG delivers over one million computer cores to physicists and makes use of over an exabyte of storage, (e.g., see http://wlcg-cric.cern.ch/core/pledge/list/) so this model is already functioning well at scales similar to those required by the SKA.

Each SRC pledging cycle would coincide (TBD) with the call for proposals for each SKAO observing cycle.

Anticipated SRC resource requirements will need to draw from results of previous years and look forward to plans regarding Observatory upgrades, newly available modes and long-term projects such as KSPs. Once the target level for the total pooled SRC resource is agreed upon, individual members would, through appropriate funding agencies, pledge resources through individual SRCs to be made available as part of the SRCNet. This

collective pool of resources and capabilities will then be used to support the SKA science programme of KSPs and PI projects (see P16[10], P17[11]).

The reason the definition of a shared pool is so important is that it gives the SKA proposal review process freedom to recommend a science program (allocating telescope and SRCNet resources) without the need to attempt to manage national SRC resource allocations.

Broad agreement between SRC resource needs and SRC resource pledging per nation can (to first order) be ensured if each member's SRC resource pledge, as a fraction of the total SRCNet, aligns with that Member's share of the SKA Project (over time, the allocation of telescope resources to projects will align with Member's share of the SKA Project).

Any SKA member country could choose to develop SRC resources locally which go beyond the agreed communal pledge amount if they desire. For example, more compute or storage capacity than the amount pledged may be provided only to local/regional users within that SRC's community.

Several areas of management are needed on various timescales to support the functioning of the SRCNet and its components, these are listed below, ordered according to the timescales and group responsible.

Annually / Ongoing SCC	• Oversee and develop the pledging of SRC resources and personnel into the SKA-facing roles and for user-facing purposes. Interaction between the SCC and the SKA's time allocation process will be required to ensure sufficient capacity to support the next cycle's science programme. • Manage a roadmap for the development of SRC-related technologies and pledging of (human) resources to enable this – for example, SRC collaboration tools, data management and transfer tools, improving best practice. • Develop and maintain a long-term roadmap for managing data transfer, data storage, and data processing capabilities of the SRCNet in line with the prospective long-term science programmes of the SKA Observatory.

[10] P16 Resources pledged into the SRC Network will enter, and be allocated from, a global federated pool.
There will be a policy to describe how resources are pledged and enter into the SRC Network. The physical allocation of those resources to projects that request them will be centrally managed by the SRC Operations Group (following other principles outlined in this document).

[11] P17 The allocation of resources will be per project. *Storage and processing resources will be allocated to projects and not to individual users. The use of those resources by the project team will be the responsibility of the project PI(s). This applies to all projects, whether they are SKA observing projects, or archival (public data) projects.*

Quarterly SCC	• Agree on priorities in development areas for subcommittees to work on. • Track performance of SRCs against pledges and KPIs, both for SKA-facing and user-facing activities and of the data transfer network links. • Review data management policies to ensure appropriate quality of service (trading off performance and cost/capacity needs).
Continuously (day to day operations) SOG	The SCC will be supported by an operations team of SKA and SRC staff, the SRC Operations Group (SOG see P4[12]), to implement the policies and priorities that have been agreed and provide the required monitoring and reporting capabilities to alert stakeholders to any difficulties. The SOG will ensure the continuous availability of the SRCNet by: • reporting to SCC the performance of SRCs against pledges, both for SKA-facing and user-facing activities and on the performance of the data transfer network links; • updating data management policies to ensure appropriate quality of service (trading off performance and cost/capacity needs); • maintaining a global data management service implementing the data management policies; • monitoring the ability of each site storage element to accept data products from SKA sites; • monitoring the ability of appropriate sites to accept batch processing jobs and to provide interactive sessions for users • monitoring the network link availability and performance (e.g., through continuous monitoring of links in use and liaison with network providers). • monitoring user support, tracking Helpdesk metrics and other feedback The SOG will report any problems to the SRCs concerned through a formal ticketing system. It will be the responsibility of each SRC to perform its own internal operational procedures to ensure its availability to the SRCNet is in agreement with the terms in its MoU with the SKAO and the terms and conditions of its accreditation as an SRC. The SRC operations group will manage the delivery of ODPs into the managed data storage in the SRCNet to specific SRC sites.

The SRCNet architecture is built on a set of components that offer functionality to the global SRCNet.

[12] P4: SRC Network Operations will be supported in a distributed fashion and coordinated by a single body, the SRC Operations Group.
The SRC Operations Group (SOG) will coordinate the operational support for the SRC Network. The SOG will be distributed across the SRC Network but coordinated from the SKAO. The SOG will not have a role to play in local operations that pertain to the maintenance and operation of local data centres.

The high-level services needed to support the SRCNet architecture are [reference]:

1. Data Management Service
2. Metadata management service
3. Workflow management service
4. Authentication and Authorisation Service
5. Services Discovery Service

At the site level, physical resources need to be purchased or set aside for use within an SRC to meet pledges. On top of this hardware the relevant service software will need to be deployed by the team at each SRC in order to enable those resources to contribute towards the multiple federated services needed for SRCNet - basically allowing those local services to be discovered by the "Services Discovery Service" and used by the federation-level services (points 1-4) above.

The SRC Operations group will need to coordinate monitoring of the health of the services running at each site and of the centralised services. Responsibilities for maintaining the centralised services are yet to be fully determined - but these federating entities are absolutely critical to the SRC mission - they turn a heterogeneous set of nodes into a coherent science analysis ecosystem.

Individual sites and the local services they offer can become unavailable but the SRCNet system will remain functional - albeit at reduced capacity, and with some inconvenience to users using local services. The responsibility for keeping individual sites, and their associated services, up and running will rest with the local site teams, but the responsibility for maintaining the critical federating services will be a collective effort managed by the SOG.

A References

A.1 Applicable Documents

The following documents are applicable to the extent stated herein. In the event of conflict between the contents of the applicable documents and this document, **the applicable documents** shall take precedence.

[AD1] SKA-TEL-SKO-0001722, SKAO Establishment & Delivery Plan, Rev01

[AD2] SKA-TEL-SKO-0001795, SKAO Helpdesk Concept, Rev 01

[AD3] SRC-0000005, SRCNet Vision and Principles, Rev 01

[AD4] SRC-0000001, SRCNet Software Architecture Document, Rev 01

A.2 Reference Documents

The following documents are referenced in this document. In the event of conflict between the contents of the referenced documents and this document, **this document** shall take precedence.

[RD1] SRC-0000003, SRCNet Science Analysis Platform Vision, Rev01

[RD2] Reference Document 2

LIST OF ABBREVIATIONS

AD	Applicable Document
ADP	Advanced Data Product
LTP	Long Term Preservation
ODP	Observatory Data Product
OEDP	Observatory Establishment and Delivery Plan
OLDP	Observation Level Data Product
PLDP	Project Level Data Product
RD	Reference Document
SB	Scheduling Block
SCC	SRC Coordination Committee
SDP	Science Data Processor
SKA	Square Kilometre Array
SKAO	SKA Observatory
SOG	SRC Operations Group
SRC	SKA Regional Centre
SRCNet	SKA Regional Centre Network
SRCSC	SRC Steering Committee

DOCUMENT HISTORY

Revision	Date Of Issue	Engineering Change Number	Comments
01	2021-05-14	N/A	Initital Release
01B	2023-08-24		For SRCNet Review

DOCUMENT SOFTWARE

	Package	Version	Filename
Word processor	MS Word	Office 365	SKAO-TEL-0000000-01B_GenDocTemp_Unclassified_EmptyTemplate.docx
Block diagrams			
Other			

ORGANISATION DETAILS

Name	SKA Observatory
Registered Address	Jodrell Bank Lower Withington Macclesfield Cheshire, SK11 9FT, UK
Fax	+44 (0)161 306 9600
Website	www.skao.int