SKA Regional Centres Network (SRCNet) Use Cases

SRC-0000004	Revision 01
Classification:	UNRESTRICTED
Document type:	NOT
Date:	2023-06-20
Status:	DRAFT
Authors:	Franzen, Thomas; Clarke, Alex; Breen, Shari; Bolton, Rosie

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

TABLE OF CONTENTS

1 Introduction 4

1.1 Purpose of the document 4

1.2 Scope of the document 4

1.3 Data type definitions 5

2 SRCNet Operational Use Cases 5

2.1 Interfacing with the data archive 6

2.2 Data inspection 6

2.3 Management of user account 6

2.4 Management of compute resources 7

2.5 Data visualisation 8

2.6 Software access and data processing 8

2.7 User support 9

2.8 Data management across the SRCNet 10

3 SRCNet Scientific Use Cases 11

4 References 31

4.1 Applicable documents 31

4.2 Reference documents 31

LIST OF ABBREVIATIONS

AD	Applicable Document
ADP	Advanced Data Product
ADQL	Astronomical Data Query Language
ART	Agile Release Train
DM	Dispersion Measure
DOI	Digital Object Identifier
EoR	Epoch of Reionisation
FAIR	Findable. Accessible. Interoperable. Reusable.
GPU	Graphics Processing Unit
GSM	Global Sky Model
GUI	Graphical User Interface
HECP	High Energy Cosmic Particles
HPC	High Performance Computing
ICRS	International Celestial Reference System
LOFAR	LOw Frequency ARray
LSM	Local Sky Model
MJD	Modified Julian Date
ODP	Observatory Data Product
OLDP	Observation-level Data Product
PI	Principal Investigator
PLDP	Project-level Data Product
PSF	Point Spread Function
PTA	Pulsar Timing Array
QA	Quality Assessment
QoS	Quality of Service
RA	Right Ascension
RD	Reference Document
RFI	Radio Frequency Interference
RM	Rotation Measure
RMS	Root Mean Square
SB	Scheduling Block

SDP	Science Data Processor
SED	Spectral Energy Distribution
SKA	Square Kilometre Array
SKAO	SKA Observatory
SNR	Signal-to-Noise Ratio
SQL	Structured Query Language
SRC	SKA Regional Centre
SRCNet	SRC Network
SRCSC	SRC Steering Committee
SWG	Science Working Group
TBD	To Be Determined
ToA	Time of Arrival
TPA	Thousand Pulsar Array
UMAP	Uniform Manifold Approximation and Projection
VLBI	Very Long Baseline Interferometry
VO	Virtual Observatory
WCS	World Coordinate System
WG	Working Group

1 Introduction

1.1 Purpose of the document

The purpose of this document is to provide a high-level overview of tasks an astronomer as well as SKAO/SRC Staff would be likely to perform within the SRCNet. This includes interfacing with the data archive, exploring the data using visualisation tools, running scripts for further data processing and managing compute resources. The SRCNet Operational Use Cases are presented in Section 2 and the SRCNet Scientific Use Cases in Section 3.

The presented use cases are intended to provide a reference for the kinds of activities that users are expecting to conduct within the SRCNet, in order to inform thinking and provide guidance to the SRC ART. This broad set of use cases incorporates several efforts in a similar vein, in an effort to define one extensive set of use cases. In particular, nearly all the SRCNet User Stories [RD1; associated user stories are given in square brackets throughout the document] stemming from a survey of the Science Working Groups (SWGs) conducted by the SRC Steering Committee (SRCSC) WG6, as well as work done by the SRC ART teams (e.g. Orange and Coral), have been folded into this set.

While this input has been invaluable to the delivery of a broad set of use cases, we expect to release future versions of this document, following feedback from the SWGs and the wider user community. Through this feedback, we hope to uncover a truly representative set of use cases, filling in gaps in the current set (e.g. simulations which are not incorporated in the current set but noted in a number of SRCNet User Stories[1]).

Throughout the document, we consider both the scientific user and SKAO and SRCNet Staff. We make no attempt to distinguish between SKAO and SRCNet Staff, but acknowledge that certain use cases may commonly apply to only one of those staff categories.

The SRCNet will be committed to providing, and abiding to, accessible and equitable tools, practices and processes as noted in the SRCNet principles to satisfy the presented use cases.

1.2 Scope of the document

This document provides a set of use cases that are intended to be representative of the kinds of activities users might want to conduct on the SRCNet. It does not:

- Attempt to address how the use cases will be implemented, e.g. the choice of software, computational requirements; or
- Describe proposal preparation and submission activities as these will not be a service provided by the SRCNet.

[1] [SRC-266, SRC-299, SRC-300, SRC-308, SRC-320, SRC-326]

1.3 Data type definitions

In general, the SKAO defines three types of SKA science data products split between two categories [AD1]:

- Observatory Data Products (ODPs):
  - Observation-level Data Products (OLDPs) are calibrated data products generated by Science Data Processor (SDP) workflows and are based on data obtained from a single execution of a scheduling block (SB).
  - Project-level Data Products (PLDPs) are calibrated data products generated by combining several related OLDPs, delivering the requirements of the PI as outlined in their original proposal.

  The Observatory is responsible for the generation of both types of ODPs, providing the workflows, software, ensuring quality assessment (QA), reproducibility and that both product types are appropriately stored and made available to users (i.e. archived). Only ODPs that have passed QA will be made available in the archive.

- Advanced Data Products (ADPs): These are the user-generated products, produced through the detailed and rigorous analysis and modelling of ODPs (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 or other facilities.

In most cases, one of the copies of a data product will be made available to the users at the SRC node where resources have been allocated for analysis. For this reason, there will be no need to download data products; these can simply be "accessed" or "retrieved" from the SRC node, reducing the burden on storage and network resources. In some instances, however, users will need to download some final products (e.g. catalogues, images, plots, spectra) for use in publications. Large data products can be linked to publications through the use of DOIs.

2 SRCNet Operational Use Cases

The SRCNet Operational Use Cases are written from the perspective of both the end user (i.e. scientist) and SKAO or SRCNet Staff.

2.1 Interfacing with the data archive

As a user, I want to be able to:

- Browse the entire SKA data collection so that I can find data products which are suitable for my science goals.

- Use quick-look tools so that I can display the data products in the archive without having to download them.
- Access public data from the archive, or proprietary data if I am a project member, including user-generated ADPs, so that I can exploit the data for my own research [SRC-275].
- Inspect (on the fly) and retrieve data products available at a certain coordinate or for a long list of positions [SRC-255, SRC-256, SRC-257, SRC-258].
- Make my data products and associated workflows public to maximise scientific return and avoid duplication of effort.

2.2 Data inspection

As a user, I want to be able to:

- Browse the QA information associated with ODPs [SRC-251, SRC-293, SRC-368].
- Track the workflow history of a data product so that I can understand how it was generated by the Observatory and/or users.

As SKAO/SRCNet Staff, I want to be able to:

- Optionally transfer a dataset that fails QA to the SRCNet so that I can investigate the cause of the failure.
- Inspect metadata, provenance, and available workflows associated with user-generated ADPs to assess their robustness (connected to [SRC-262]).

2.3 Management of user account

As a user, I want to be able to:

- Create an account so that I can access public data from the archive as well as compute resources (whether via processing proposal or not).
- Edit my account details so that I can keep my personal information up-to-date.
- Add new team members to my project via a Helpdesk ticket so that they can access my data, code and/or compute resources.

As SKAO/SRCNet Staff, I want to be able to:

- Configure and grant access to compute resources (e.g. node limits, queue, storage) for users or projects.
- Monitor usage at the project level to ensure it is in line with appropriate policies, and take appropriate action if not [SRC-240].

2.4 Management of compute resources

As a user, I want to be able to:

- Access a temporary project/personal space (e.g. scratch) within the SRCNet where I can store and process my data.
- Track my processing time and data storage so that I can manage my compute resources.
- Submit a processing proposal so I can either request additional compute resources or secure compute for an archival project.
- Estimate the data size and processing time needed for my project before submitting a processing proposal.
- Access GPUs so that I can run computationally expensive procedures such as rotation measure (RM) synthesis and RM CLEAN on accelerated hardware [SRC-304, SRC-310].
- Exploit HPC systems so that I can process big and complex data and/or reduce time to solution and/or have access to large and fast storage devices [SRC-269].
- Check the dates of any planned maintenance affecting data access and processing so that I can plan my work accordingly.

As SKAO/SRCNet Staff, I want to be able to:

- Monitor total SRCNet account usage per user and per user group, aggregated across all SRCNet resources, currently and historically, so that I can check if a user, user group or project is still within its allocated usage and take appropriate action as required [SRC-246].
- Ensure that the needs of science projects map efficiently to SRCNet resources so that science teams have access to appropriate SRCNet resources when they need them. This may mean that the most appropriate/efficient resources are granted on the basis of a TBD prioritisation, such as total resource required or project ranking [SRC-230].
- Assign or change compute resource allocations from one SRC in response to over- or under-use at individual SRCs, or to changes in SRC availability [SRC-239].
- Keep track of when the SRC sites have planned downtime and reschedule resources if required [SRC-234].
- Define benchmarks and benchmarking tools so that they can be used for capacity/pledge measurements [SRC-229].
- Monitor pledged and used resources per SRC site, allocated SRC project and nation, to track usage (e.g. quarterly, yearly and all historical data), so that they may be used for future plans accordingly and shared with colleagues outside the SRCNet ecosystem [SRC-242].

- Verify that resources and services that have been pledged have been made available to the SRCNet so that I can ensure that the accreditation awarded to the SRC nodes is upheld and that pledges are honoured [SRC-233].

2.5 Data visualisation

As a user, I want to be able to:

- Visualise and analyse various types of data products interactively [SRC-270], including image, catalogue and beamformed data. Tasks that a user might want to perform in a standard visualisation package are listed in sections 2b, 2c, 3b, 3c, 4b and 4c of Table 1[2].
- Create publication quality plots (e.g. edit font sizes, tick sizes, grid lines, axis labels, overlay the restoring beam if applicable, export images/plots with sufficiently high resolution) and export the plotting code (e.g. python, matplotlib) to ensure that my plots are reproducible.
- Customise various settings in the GUI (e.g. the default layout for a new session, the default colour map for rendering a raster image), record these settings and apply them to new sessions with the visualiser.
- Control the visualisation software with commands, as an alternative to the GUI, so that I can automate workflows.
- Collaborate with a distributed team on the same visualisation instance.

2.6 Software access and data processing

As a user, I want to be able to:

- Query and pull Observatory/SRCNet software available so that I can run my own data processing.
- Run data processing jobs and retrieve their outputs/logs.
- Monitor the progress of my processing jobs (e.g. status, used resources) so that I can manage the data reduction and identify any issues with the processing.
- Access standard astronomy packages and libraries/tools, so that I can perform suitable data analysis. It could be helpful to be able to do this via software containers [SRC-261, SRC-263].
- Access machine learning libraries/tools so that I can run a machine learning analysis [SRC-268]. It could be helpful to be able to do this via software containers

[2] Note that we do not expect a visualisation package to be used for visualising raw data products (i.e. uv data, voltages), nor simply for plotting x versus y.

- Build my own container, or add software to an existing container, and run my own custom scripts to post-process data [SRC-254, SRC-267, SRC-305].
- Deliver my own custom software (e.g. on containers) to make it part of the official Observatory software so that it can be pulled by other users.
- Interact with the software development team at the SRC through a Helpdesk in order to get updated versions of software onto the SRC platform or develop missing functionalities [SRC-272].
- Access correct and complete sphinx/doxygen style documentation so that I can use/adapt existing software [SRC-319].
- Participate in Science Data Challenges so that I can train and test new analysis methods [SRC-303, SRC-309, SRC-321].

As SKAO/SRCNet Staff, I want to be able to:
- Deploy new software (or specific software versions) so that users can find the necessary tools to process their data [SRC-264].
- Coordinate the deployment and management of software across the SRCNet to ensure that available software is consistent across the whole network [SRC-248].
- Facilitate/support the adaptation of software tools so that they can efficiently cope with SKA data products [SRC-265].
- Publish standard information on software tools in order to ensure that they are FAIR compliant and the developers are credited [SRC-271].

2.7 User support

As a user, I want to be able to:
- Contact Observatory or SRC Support Staff via a Helpdesk so that I can obtain science and operational support, report bugs and make suggestions to improve the services provided.
- Access a community support forum so that I can ask the user community for scientific and/or technical advice.

As SKAO/SRC Staff, I want to be able to:
- Easily see the details and status of all Helpdesk tickets coming from the SRCNet, per SRC Site and per ticket type, so that I can check if issues are being handled adequately, and get help from colleagues with relevant expertise [SRC-247].

2.8 Data management across the SRCNet

As SKAO/SRCNet Staff, I want to be able to:

- Monitor the global network system health so that I can see how well different components of the global network are working and request appropriate remedial actions as required [SRC-232, SRC-235].
- Use data management tools to ensure that data access policies are being adhered to [SRC-226].
- Have the ability to flag data products for deletion, either after the request of a user (if currently available in the archive) or after identifying a violation of resource usage [SRC-241].
- Monitor the progress of data product delivery from the SDPs into the SRCNet [SRC-231].
- Re-allocate and divert data products to alternative SRC storage as required [SRC-227].
- Manage data transfers via third-party copy (direct transfer from site A to site B delegated by an external service at site C) so that I can avoid having a central site, or my location acting as a data ingest and egress node [SRC-238].
- Extract data products from the SRCNet (here, specifically, updated global sky models or CLEAN-component model representations of astronomical sources) and upload them to the SDP systems. This functionality could be needed to update the catalogues or source component lists used within the SDP pipelines in a timely manner once a new Global Sky Model (GSM), source model etc. is approved [SRC-249].
- Identify which SRC sites host replicas of data products or data collections so that I can check the distribution and number of replicas per data product/collection, or the co-location of data products within a project [SRC-236].
- Plan data product replica placement at the time of SKAO or SRCNet project allocation, taking into account the (potential) SKAO observing schedule and future availability of SRC compute resources. This is needed to ensure that replicas of data products within each project are co-located and placed at SRC sites including the site determined to be most-likely to support the project, reducing the need to move data around to support user access [SRC-243].
- Understand and control the Quality of Service (QoS) being provided at each SRC storage site and manage data product replication accordingly so that I can ensure appropriate overall QoS is being provided to each data product or data collection [SRC-237].
- Plan replicas based on rules so that short-term network outages do not require any intervention in terms of replica management, i.e. rules governing the placement of data products are eventually satisfied, with all necessary re-tries being handled automatically [SRC-228].

- Monitor and understand the popularity of data products or data collections so that I can ensure that appropriate overall QoS is being provided to popular (or unpopular) data products/collections [SRC-244].
- Keep track of the existence of corrupt replicas, and check if the system has corrected corrupt or missing replicas, with results per storage site and per replica, so that I can be confident that all the copies I have are "good". This will allow me to investigate if one storage site is performing worse than others, one file type is more likely to to be identified as corrupt, and there are patterns that could help find improvements [SRC-245].

3 SRCNet Scientific Use Cases

The SRCNet Scientific Use Cases are broadly categorised into 5 sections based on broad categories of possible OLDP types: interfacing with uv data, image (cube) data, catalogue data, beamformed data and raw antenna voltage data. Each of these sections is subdivided into data access, visualisation and analysis[3]. A summary list of all the functionality is presented below, while the full use cases are given in Table 1.

Summary list of functionality:

- Interfacing with uv data:
  - Access raw uv data, calibrated uv data and uv grids
  - Access the sky model and calibration solutions that have been applied to a dataset
  - Access observation metadata
  - Visualise uv data and calibration solutions
  - Calibrate (if appropriate) and image uv data
  - Subtract a sky model from calibrated uv data
  - Calculate the power spectrum from a subset of the calibrated uv data

- Interfacing with image (cube) data:
  - Access continuum image and spectral line cubes
  - Query multi-wavelength images online and compare them with local images loaded into the visualiser
  - Render raster images, contours and vector fields
  - Explore images interactively (display frequency channels, polarisation quantities etc.)

[3] There is not always a clear-cut division between data visualisation and data analysis but we have attempted to place the use cases into the most appropriate sections.

    - Display images side by side, and optionally, match them spatially and/or spectrally
    - View the image header
    - Visualise data in virtual reality
    - Generate postage stamps
    - Generate large-scale mosaics
    - Create regions for use within the package or another tool
    - Generate cubelets
    - Perform 2D Gaussian fitting
    - Generate spectral index/curvature maps
    - Generate moment maps
    - Generate position-velocity maps
    - Generate RM synthesis cubes
    - Extract spatial profiles
    - Extract spectra - from region or pixel
    - Fit a power law or any general functional form to an SED
    - Fit emission or absorption lines in a spectrum and overlay known spectral lines
    - Extract spectra from a Stokes hypercube and perform Stokes QU fitting
    - Extract time series from multiple images at different epochs and study time periodicities
    - Compute image statistics
    - Perform mathematical operations on image cubes and profiles derived from image cubes
    - Run continuum or spectral line source finding
    - Calculate image QA metrics
    - Classify sources based on their radio morphology using deep learning algorithms

- Interfacing with catalogue data:
    - Select and retrieve a subset of sources from large catalogues, including the Local Sky Model (LSM) Catalogue and the Imaging Transient Source Catalogue
    - Access the Science Alerts Catalogue produced by SDP
    - Overlay a local or remote source catalogue on an image
    - Extract flux densities/spectra at catalogue positions
    - Filter and edit a catalogue within the visualiser
    - Create 2D scatter and histogram plots to explore catalogue data
    - Perform cross-matching
    - Classify sources based on their tabulated properties

- Interfacing with beamformed data:
  - Access sieved pulsar and transient candidates, transient buffer data and pulsar timing solutions
  - Display single-pulse candidates
  - Display pulsar/transient candidates found through period searches
  - Display velocity and acceleration curves for newly discovered binary pulsars
  - Query external pulsar catalogues
  - Query pulsar observations in the SKA archive
  - Reduce, process and visualise Stokes filterbank data
  - Analyse pulsar data including de-dispersion and folding
  - Extract pulsar spectra from polarisation-calibrated profiles and calculate RMs
  - Fit Time of Arrivals (ToAs) with a pulse timing model and inspect residuals
  - Analyse the noise properties of a pulsar

- Interfacing with raw voltage data:
  - Access raw antenna and particle-detector voltage data
  - Display information about voltage dump requests
  - Reconstruct the properties of a cosmic ray from the raw antenna and particle-detector voltage data
  - Analyse the raw antenna voltage data with near-field interferometry

Table 1: SRCNet Scientific Use Cases broadly categorised into 5 sections: (1) interfacing with uv data; (2) image data; (3) catalogue data; (4) beamformed data and (5) raw antenna voltage data.

No.	SRCNet Scientific Use Case	Description	Reference
1	Interfacing with uv data		
1a	uv data access		
1a1	Accessing raw uv data	Access the raw uv data for a small set of observations, with the option of averaging the data in time and/or frequency, in order to refine any observational or pipeline configuration parameters before the bulk of the observations are performed. (Note: SKAO expects raw uv data to be delivered for a very small number of projects)	SRC-295 SRC-316 SRC-350

1a2	Accessing calibrated uv data	Access the calibrated uv data, with the option of averaging the data in time and/or frequency, in order to further refine the calibration, re-image the data, or analyse the data in the uv plane. (Note: SKAO expects calibrated uv data to be delivered for a small number of projects)	[AD1] SRC-280 SRC-298 SRC-307 SRC-328 SRC-350 SRC-358 SRC-379
1a3	Accessing uv grids	Access the calibrated uv data gridded at the spatial and frequency resolution required by the experiment as well as the accumulated weights for each uv cell in order to re-reduce/re-image the data.	[AD1] SRC-250 SRC-274 SRC-280 SRC-341 SRC-345 SRC-362 SRC-375 SRC-384
1a4	Retrieve a sky model and calibration solutions	Retrieve the sky model and calibration solutions that have been applied to a dataset so that they can be used in post-processing.	SRC-351 SRC-378
1a5	Accessing observation metadata	A user would want to access information about the quality of the uv data, e.g.: - Missing antennas/stations - Percentage of flagged data - Distance of A-team sources from pointing centre - Atmospheric seeing conditions - Ionospheric conditions The standard data products made available by the Observatory will already have data quality control performed.	SRC-276 SRC-283
1a6	Accessing metadata at the station/dipole level	Access metadata at the station/dipole level, including humidity, temperature, ionospheric information, quality statistics, maintenance activity, and data required to calculate station beam models (e.g. detailed information about failed or non-working dipoles).	SRC-352
1b	uv data visualisation	(not necessarily expected or required to be part of a standard visualisation package)	
1b1	Visualising uv data	Visualise the uv data (or a subset thereof). Quantities that the user may want to visualise include: - amplitude versus time/frequency	SRC-277 SRC-278 SRC-292 SRC-311

		- phase versus time/frequency - amplitude versus uv-distance - rms versus uv-distance - weight versus time - uv-gridded visibilities as either amplitude / phase or real / imaginary - accumulated uv-gridded weights Allow colour-coding by source, scan, antenna (or antenna subset), baseline, spectral window etc.	SRC-315 SRC-340
1b2	Visualising calibration solutions	Plot the calibration solutions (delays, phases, amplitudes, fringe rates, correction for atmospheric absorption etc.) to inspect the quality of the calibrated uv data in detail.	SRC-378
1c	uv data analysis		
1c1	Calibrating uv data	It should be possible to calibrate the data using pipelines that are functionally identical to the standard processing that would occur within SDP to generate equivalent data products (in this case calibrated visibilities).	SRC-295 SRC-298 SRC-316 SRC-350
1c2	Imaging gridded uv data	A user would want to image the uv-gridded, calibrated uv data. It should be possible to image the data using pipelines that are functionally identical to the standard processing that would occur within SDP to generate equivalent data products (in this case image cubes). (For example, they may want to apply a weighting function that allows for an approximately constant PSF across all frequency channels (as may be needed for e.g. spectral work), adjust the deconvolution parameters, use smaller frequency channels (e.g. to reduce depolarisation), or sub-divide the data by time.)	SRC-252 SRC-274 SRC-295 SRC-298 SRC-307
1c3	Initial EoR data quality assessment	A user would want to calculate the power spectrum of 21 cm fluctuations from the calibrated uv data for a single track, providing sufficient SRC computing resources are available, so that they can perform an initial assessment of the quality of the dataset.	SRC-355
1c4	Subtracting a sky model from calibrated uv	A user would want to subtract a specified sky model from the calibrated uv data. For example, they may want to image a small field-of-view that only contains the source of interest. The model to	SRC-301 SRC-313

	data	be subtracted could be a subset of sources from the LSM catalogue corresponding to their observation/project. Alternatively, they may want to visualise the residuals to assess the goodness of fit of source models to the data.	
1c5	Changing the phase centre	Specify a phase centre and rotate the visibilities to this phase centre to image a particular source or calibrate the data in a particular direction.	SRC-312
2	Interfacing with image data		
2a	Image cube access		
2a1	Access continuum image cubes	Access a subset of (full-Stokes) continuum image cubes satisfying certain criteria (e.g. sky coverage, frequency range, time range, and possibly polarisation properties). This should include the CLEANed restored image (multiple Taylor terms if applicable), CLEAN component image, residual image, dirty image, PSF map, and if available, the RMS noise and background maps.	[AD1] SRC-279 SRC-280 SRC-290 SRC-296 SRC-317 SRC-322 SRC-329 SRC-359 SRC-360 SRC-373 SRC-383
2a2	Access spectral line cubes	Access a subset of (full-Stokes) spectral line cubes, before or after continuum subtraction, satisfying certain criteria (e.g. sky coverage, frequency range, time range, and possibly polarisation properties). This should include the CLEANed restored image, CLEAN component image, residual image, dirty image, PSF map, and if available, the RMS noise and background maps.	[AD1] SRC-279 SRC-296 SRC-329 SRC-347 SRC-359 SRC-360 SRC-374
2a3	Accessing local imaging data from VLBI observations	For VLBI observations, access image cubes from the same sub-array from which the VLBI beams were formed to support the VLBI calibration, as well as to complement the science data return by providing images (of the VLBI beam size) with different bandwidths, angular and spectral resolutions.	SRC-382
2b	Image cube visualisation	Efficient visualisation of large image cubes (hundreds of GB to a few tens of TB)	SRC-256 SRC-292

			SRC-306 SRC-344
2b1	Raster rendering	Display a raster image. Various options should be available to control how the raster image is rendered, including the colour map, scaling function, clip level and bias/contrast.	
2b2	Contour rendering	Generate a contour image. Various options should be available to control how the contours are rendered, including the contour levels, smoothing function and contour styling.	
2b3	Vector field rendering	Overlay a vector field (e.g. linear polarisation, magnetic field) or a scalar field (e.g. temperature) on an image. Allow the user to control the intensity to vector length mapping, angle offset etc.	
2b4	Comparing images	Compare images by overlaying contours on a raster image, displaying a semi-transparent image over another image, or blinking.	SRC-286
2b5	Exploring an image	Change the view of an image by zooming in and out, and panning. Allow the user to switch WCS (Galactic, ICRS, J2000, B1950 etc.) and intensity units (Jy/beam, K, Jy/sr etc.).	
2b6	Displaying cursor information	When browsing an image, display pixel information at the cursor position, including the RA and Dec coordinates, the pixel value and the frequency/velocity/polarisation parameter (if applicable).	
2b7	Multi-panel view	Visualise different images side by side.	
2b8	Matching images spatially and spectrally	Match different images in world coordinates spatially and/or spectrally. This would enable display operations (e.g. zooming, panning, reporting the pixel value at the cursor position) and mathematical operations (e.g. measuring the rms noise in a certain region of sky) to be performed on them simultaneously.	
2b9	Forming a Stokes hypercube	Combine individual Stokes image cubes on the fly to form a Stokes hypercube.	

2b10	Displaying frequency channels	Step through a series of image cube slices after defining a specific range of velocities/frequencies for a particular source. Users should be able to jump to specific positions in the cube.	
2b11	Displaying polarisation quantities	For Stokes cubes, step through the polarisation components and additional computed components (i.e. polarised intensity, fractional polarisation and position angle). The latter should be calculated on the fly.	SRC-282 SRC-297 SRC-318 SRC-333 SRC-339 SRC-365
2b12	Viewing the image header	View a brief summary of the image properties as well as the full image header.	
2b13	Displaying station-level images from the near-field	Display station-level images from the near-field made from raw antenna voltage data, as generated using code [RD2] from the High Energy Cosmic Particles (HECP) SWG. It should be possible to overlay aerial photos (for example) on the images to identify radio frequency sources. (The requirements for displaying station-level images from the far-field should be the same as those for standard interferometric images.)	
2b14	Online image query	Access multi-wavelength images and images from other radio telescopes registered in the VO.	SRC-259 SRC-284 SRC-348
2b15	Simultaneous viewing of local and remote data	A user may want to load local data into a visualisation window, and compare them with remote data.	
2b16	Visualising data in virtual reality	Visualise (and potentially analyse/manipulate) images with a VR headset, in particular 3D data products such as image cubes.	
2c	Image cube analysis		
2c1	Generating image cutouts	A user would want to generate their own sized image of a patch of sky. A user would want multiple postage stamps of various sources from one or more (full-Stokes)	SRC-281 SRC-287 SRC-291 SRC-297 SRC-302 SRC-318

		images selected from a reference catalogue, with the option of stacking these. The sky position/object name and region size would need to be provided as a minimum. Additional flags could be provided, such as frequency/velocity range, time range, and polarisation properties. A user would want a coverage map to identify the constituent pointings.	SRC-349 SRC-391
2c2	Mosaicking	Combine multiple pointings into a larger (full-Stokes) image, potentially covering thousands of square degrees, taking into account the pixel weights. One of the outputs from the mosaicking should be a weight map, or equivalently a noise map. This could be part of 2a1, 2a2, 2a3 and 2c1, where the user simply gives a position and region size, mosaicking is performed automatically if required, and the user is informed.	SRC-281 SRC-297 SRC-324 SRC-332 SRC-339 SRC-364 SRC-371
2c3	Combining multi-epoch data in image domain	Regrid and co-add together images at multiple epochs to improve the sensitivity. RMS noise maps or weight maps could be provided to weight the images. The software should be capable of dealing with various astrometric projections.	
2c4	Generating cubelets	Extract subsections (in sky area, frequency and polarisation) of an image cube for detailed investigations.	SRC-290 SRC-357
2c5	Extracting regions	Create regions (point, line, rectangle, ellipse, polygon etc.) on an image cube for use within the package. Image planes should be included in the region selection. Allow the user to label regions and modify their appearance (colour, line thickness etc.). A user may want to export regions as a region file for analysis with another tool, or import region files from another tool.	
2c6	Computing image statistics	Compute statistics (sum, mean, RMS, maximum etc.) in an image cube (or subregion of an image cube in sky area, frequency and polarisation). There should be the option of querying multiple regions, and printing the results to screen or in a text file.	

2c7	Generating image histograms	Visualise image data within a selected region as a histogram.	
2c8	2D Gaussian fitting	Perform image 2D Gaussian fitting to extract the position, flux density and size of a source. It should be possible to fit multiple sources and print the fitted parameters to screen or a text file.	
2c9	Measuring distances	Measure a geodesic distance between two locations on an image.	
2c10	Algebraic operations on images	Perform algebraic operations on each pixel of one or more images. If the image consists of three or more axes, it should be possible to specify the image planes to be used in the calculations. For a single image, this could be useful for a change of scale (e.g. change of units) or to obtain some parameters (e.g. map the excitation temperatures of a line assuming a value for the opacity). With multiple images, one could obtain, for example, continuum-subtracted image cubes, spectral index maps, isotope abundance maps (comparing the line emission of different isotopologues), or temperatures (comparing maps of different transitions of the same molecule).	
2c11	Complex operations on images	Perform non-algebraic/complex operations on one or more images. Complex operations include smoothing, convolution, interpolation, re-binning, calculating the Fourier Transform, and collapsing along a particular axis.	
2c12	Generating spectral index and curvature maps	From a Stokes I image cube, generate a spectral index or spectral curvature (taken to include the fitting of any general functional form) map. The user may only want to plot the spectral index/curvature if the error is below a certain threshold, and overlay the flux contours on these maps to better interpret them. Additional Stokes I image cubes from other frequency bands or instruments may be included in the spectral fitting if required, after convolving them to a common resolution.	SRC-281 SRC-282 SRC-297 SRC-331 SRC-363
2c13	Generating moment maps	Derive a moment map of a region/slice after defining a specific line. The moment maps should include: - Mean value of the spectrum	SRC-253 SRC-281 SRC-333 SRC-339

		- Integrated value of the spectrum - Intensity weighted coordinate (traditionally used to get velocity fields) - Intensity weighted dispersion of the coordinate (traditionally used to get velocity dispersion) - Median value of the spectrum - Median coordinate - Standard deviation about the mean of the spectrum - RMS of the spectrum - Absolute mean deviation of the spectrum - Maximum value of the spectrum - Coordinate of the maximum value of the spectrum - Minimum value of the spectrum - Coordinate of the minimum value of the spectrum	SRC-344 SRC-365
2c14	Generating position-velocity maps	Derive a position-velocity map along a cut (e.g. the major axis of a galaxy) in an image cube, for a specific line. Allow the user to define the cut interactively or via an imported region file.	SRC-281 SRC-333 SRC-339 SRC-344
2c15	Generating RM synthesis cubes	Perform RM synthesis from a Stokes QU image cube. Software examples: rm-synthesis, RM-Tools, cuFFS, pyrmsynth	SRC-282 SRC-297 SRC-318 SRC-339 SRC-389
2c16	Extracting spatial profiles	Display a 1D cut across RA, Dec, or a user-specified direction (e.g. along the axis of a radio galaxy). The spatial profile extractor should be capable of supporting line and polyline regions.	SRC-315
2c17	Extracting spectra	Extract and display a spectrum from an image cube at the cursor position (single pixel), or over a user defined region. The user may want to overlay multiple spectra from different image cubes. For spectral line data, allow the user to: - switch between displaying the frequency, velocity and channel number on the x axis. - convert the units for the intensity (Jy/beam, K), frequency (GHz, MHz) and velocity (km/s, m/s). - extract spectra for a list of objects/positions, pre-process the spectra (rebinning/rest frame conversion) and stack them together with an uncertainty analysis.	SRC-281 SRC-334 SRC-339 SRC-342 SRC-344 SRC-366

2c18	Spectral index and curvature fitting	Fit a power law or higher-order polynomial to an SED. Return fitted parameters, errors, and goodness-of-fit. Overlay the fitted model on the SED.	
2c19	Fitting emission/absorption lines in a spectrum	Fit emission or absorption lines in a spectrum and derive properties (frequency, width, etc.). A user may want to use more complex line models than a Gaussian, e.g. double-horned profile, polynomial fit, hyperfine transitions and multiple velocity components for a single spectral line.	SRC-344
2c20	Querying databases for astronomical spectroscopy	Query an external database (e.g. Splatalogue) to obtain the frequencies or velocities of known spectral lines.	SRC-344
2c21	Overlaying lines on a spectrum	Overlay known spectral lines on a selected region of a spectrum based on an external database, with the ability to adjust for local velocity or redshift.	SRC-344
2c22	Extracting spectra from a Stokes hypercube	From a Stokes hypercube, extract and display basic polarisation quantities (including polarised intensity, polarisation fraction and position angle) as a function of frequency at the cursor position (single pixel) or over a region of interest.	
2c23	Stokes QU-fitting analysis	Model the Stokes parameters Q and U using wavelength-dependent analytical models available in the literature, to study the nature and structure of the magnetised plasma.	
2c24	Extracting time series	Where images exist at different epochs, extract and display the time series of flux density for an extracted source to check for variability or perform a time series analysis. This extraction could be performed at a single frequency, or over a range of frequencies to increase the SNR.	SRC-282 SRC-335 SRC-346 SRC-367
2c25	Studying time periodicities	From a time series, study time periodicities through a Fourier domain analysis.	
2c26	Generating dynamic spectra	From multiple image cubes at different epochs, extract and display the dynamic spectrum (i.e. frequency/velocity on one axis, and time on the other) over a region of interest.	
2c27	Collapsing dynamic spectra	Collapse a dynamic spectrum along the time or frequency axis to create a spectrum or a time series, respectively.	

2c28	Profile smoothing	Apply smoothing (e.g. Boxcar, Gaussian) to spatial profiles, spectral profiles, time series and dynamic spectra to increase the SNR.	
2c29	Algebraic operations on profiles derived from image cubes	Perform algebraic operations on one or more profiles derived from image cubes. For example, co-add data to increase the SNR or apply a multiplication factor to the intensity scale.	
2c30	Computing statistics for profiles derived from image cubes	Compute statistics for profiles derived from image cubes with the option of querying different subsets of the data, and printing the results to screen or in a text file.	
2c31	Continuum source finding	Run a source finder on a continuum image cube and generate a source catalogue. Performed using classical methods (e.g. PyBDSF, Aegean, Duchamp, Blobcat) or deep learning algorithms such as convolutional neural networks.	SRC-325 SRC-356
2c32	Spectral line source finding	Run spectral line (emission and absorption) source finding to detect and characterise lines in spectral line cubes. Performed using classical methods (e.g. SoFiA, Duchamp) or machine learning algorithms.	SRC-253 SRC-343
2c33	Image data quality control	A user would want to inspect QA reports to independently verify the quality of images (astrometry, flux scale, RMS, dynamic range, polarisation fluxes and angles, etc.) generated by the Observatory.	SRC-323 SRC-336 SRC-356 SRC-368 SRC-386 SRC-387 SRC-388
2c34	Source classification from images	Classify sources based on their appearance in images using deep learning algorithms (e.g. convolutional neural networks). This could be in combination with use case 3c3 which classifies sources based on catalogue data extracted from a source finder.	
3	Interfacing with catalogue data		
3a	Catalogue access		

3a1	Retrieving a list of sources	A user would want to select a subset of sources from catalogues generated by the SDP, including the LSM Catalogue and Imaging Transient Source Catalogue, based on various criteria (e.g. search radius, frequency range, time range). For a simple cone search, the user could select a catalogue, and specify the required RA, Dec and radius, via a GUI. For more complex searches, the user could create an SQL/ADQL command either from scratch or from an example. Running the command would produce a list of sources within the window. This list could be retrieved in various formats (txt, html, csv, etc). Large source lists may need to have their rows/columns truncated while displayed depending on server speeds.	[AD1] SRC-279 SRC-281 SRC-290 SRC-294 SRC-297 SRC-318 SRC-323 SRC-330 SRC-339 SRC-348 SRC-353 SRC-361 SRC-390
3a2	Investigating source attributes through a schema browser	A user would want to check various attributes associated with source table data. A source may have hundreds of associated columns in the database detailing technical information, flags, attributes. The user will need access to a schema browser to understand which columns are relevant for them: data type, bit length, format, units, description, column ids, etc. For example, pulling all columns during an SQL query may be highly impractical in terms of data sizes and workflow efficacy, so the user may want to check which ones are important for their needs.	
3a3	Accessing the Science Alerts Catalogue	Access the Science Alerts Catalogue produced and communicated by the SDP, which provides a searchable and retrievable record of past alerts.	[AD1] SRC-392
3b	Catalogue visualisation	Efficient visualisation of large catalogues (few GBs)	SRC-292 SRC-306
3b1	Catalogue filtering	Apply a filter to a source catalogue, such as value range or string match.	
3b2	Catalogue image overlay	Overlay a known source catalogue on an image cube, or a source catalogue extracted from the image itself.	

3b3	Catalogue 2D scatter plot	Render a 2D scatter plot from two numeric columns of a catalogue.	
3b4	Catalogue histogram plot	Render a histogram from one numeric column of a catalogue.	
3b5	Linked catalogue visualisation	The catalogue table, image overlay, histogram and 2D scatter plot should be interlinked, meaning that if a set of sources are highlighted in one place, this should trigger source highlights in other places.	
3b6	Online catalogue query	Query multi-wavelength catalogues and catalogues from other radio telescopes registered in the VO in a selected region of an image, and return a list of sources present in the region.	SRC-259 SRC-284 SRC-348
3c	Catalogue analysis		
3c1	Computing column statistics for catalogues	Compute statistics for the values in each of the columns of a catalogue. There should be the option of querying different row subsets, and printing the results to screen or in a text file.	
3c2	Editing source catalogues within the visualiser	Display a catalogue on an image or sky projection and edit a version of that catalogue within the visualiser. In this way, users can create enhanced versions of catalogues or filter catalogues from within the visualiser. They could perform manual or automated operations on source catalogue data including: - adding tags to sources (e.g. sourceness tags, morphological tags, physical class labels, quality tags) - adding/removing sources from the catalogue - correcting source parameters or contours - filtering sources by position, assigned tag, source parameters, etc.	
3c3	Source classification from catalogues	Classify sources based on their tabulated properties. This could be in combination with use case 2c34 which classifies sources directly from images (deep learning).	

3c4	Cross-matching	Allow for spatial matching of multiple catalogues within a definable search radius (such as in TOPCAT).	SRC-259 SRC-273 SRC-285 SRC-348
3c5	Extracting flux densities at catalogue positions	Extract flux densities or spectra from an image cube at the positions listed in a catalogue.	
3c6	2D visualisation of catalogues	Run a dimension reduction algorithm such as Uniform Manifold Approximation and Projection (UMAP) to visualise large catalogues in 2D before running more complex machine learning algorithms.	SRC-343
3c7	Running cosmology pipelines	Run cosmology pipelines on source catalogues, e.g. calculate two- or three-point correlation function. (to be expanded into more detailed use cases by Cosmology SWG and other members of the community)	SRC-327
4	Interfacing with beamformed data		
4a	Beamformed data access		
4a1	Access sieved pulsar and transient candidates	Access the output data from the pulsar search section: - for each pulsar/transient candidate, a data cube folded and de-dispersed at the best dispersion measure, period and period derivative determined from the search - a single ranked list of non-imaging pulsar/transient candidates from each SB - for those transients or pulsars deemed of sufficient interest based on a set of parameters, the associated “filterbank” data - a set of diagnostics/heuristics including metadata associated with the SB and observation	[AD1] SRC-381
4a2	Access transient buffer data	Access raw voltage data for pulsar/transient candidates when the transient buffer is triggered.	[AD1] SRC-376

4a3	Access pulsar timing solutions	For each detected pulsar, retrieve the output data from the pulsar timing section, including: - the original input data (channelised time-series of complex voltages) as well as averaged versions of these data products (either averaged in polarisation, frequency, or time) in PSRFITS format - the arrival time of the pulse - the residuals from the current best-fit timing model for the pulsar	[AD1] SRC-377
4b	Beamformed data visualisation		
4b1	Displaying single-pulse candidates	Display the dynamic spectrum for any beam and polarisation parameter, and other diagnostic plots such as DM/SNR versus time to inspect the nature of a single-pulse candidate. (See e.g. diagnostic plots used by MeerTRAP and CHIME/FRB.) Dynamic spectrum plots are also applicable to solar observations.	
4b2	Displaying pulsar/transient candidates found through period searches	Display diagnostic plots such as folded pulse profiles, SNR versus DM, frequency/time versus phase and period derivative versus period, to inspect pulsar/transient candidates found through period searches. (See e.g. folded pulsar timing results displayed by Thousand-Pulsar-Array (TPA).)	
4b3	Displaying velocity and acceleration curves for newly discovered binary pulsars	Extract initial information from the folded pulse profiles of a newly discovered binary pulsar, including velocity and acceleration curves, after correcting for relativistic binary effects. (See circorbit and Dracula for examples of interactive diagrams for visualising binary pulsars.)	
4b4	Querying external pulsar catalogues	Access external pulsar catalogues such as PSRCat in order to retrieve relevant pulsar attributes (e.g. DM, period, flux density) and easily compare pulse profiles with published ones. This could be achieved via a quick-look tool allowing the user to display data associated with a pulsar.	

4b5	Querying pulsar observations in the SKA archive	Produce interactive displays/plots summarising the information related to all SKA observations performed on a given pulsar (number of observations, observing dates, integration times, SNRs, optimised periods, DMs, etc.), to be compared with data in external pulsar catalogues for previously known pulsars.	
4c	Beamformed data analysis		
4c1	Operations on Stokes filterbank data	Perform operations on Stokes (polarised power) filterbank data, including: - interactive flagging of bad data - averaging the data in time/frequency - extracting particular time/frequency ranges of interest - collapsing along the time or frequency axis to create a spectrum or a time series, respectively - subtracting one beam from another (for example to subtract an off-source beam from an on-source) - converting Stokes IQUV to linear polarisation, position angle and total power - calculating the Fourier Transform to form a secondary spectrum	
4c2	Pulsar data analysis	Perform RFI excision, de-disperse the data at the known dispersion measure (DM), measure and correct for Faraday rotation using the measured/catalogue rotation measure, and fold the data using different timing ephemerides. This would result in a number of output data products including de-dispersed time series, dynamic spectra and folded pulse profiles.	
4c3	Extracting pulsar spectra and calculating RMs	From polarisation-calibrated profiles, extract and display Stokes IQUV and additional computed components as a function of frequency. Calculate rotation measures using methods such as maximising the linear polarisation, and QU fitting.	
4c4	Pulsar timing	Derive pulse ToAs and their uncertainties, or if they are already available at the SRC, re-derive them using a different timing template and pulse profile template. Fit the ToAs with a pulse timing model incorporating transformation to the solar-system barycentre, pulsar rotation and spin-down and, where necessary, one of several binary models. Plot the residuals as a function of	

		MJD or orbital phase to assess the quality of the fit.	
4c5	Pulsar noise analysis	Analyse the noise properties of pulsars (e.g. dispersion and scattering effects, instrumental instabilities, rotational instabilities), as required in e.g. Pulsar Timing Array (PTA) experiments.	
5	Interfacing with raw voltage data of air shower observations		
5a	Raw voltage data access		
5a1	Accessing raw antenna and particle-detector voltage data	A user would access the raw antenna and particle-detector voltage data, or an arbitrary subset of it, to analyse with their own software.	
5a2	Accessing metadata associated with voltage dump requests	Display information on received antenna-level voltage dump requests (regardless of whether or not they are acted upon by the SKA system). Display information on antenna-level voltage dumps by trigger source (e.g. cosmic ray particle detectors), including: - number of triggers - total amount of data - plot the trigger times (number per hour/day; and cumulative number of triggers as a function of time) Allow the user to retrieve this information over a particular time period (day/week/month etc.). Select a particular trigger, and display information such as: - ground map of returned antennas - any antennas that are malfunctioning or unable to return data, and other pertinent metadata - total data volume - number of antennas triggered	
5b	Raw voltage data analysis		

5b1	Full cascade simulation	A user would run simulations of particle cascades and the associated radio emission, and fit these to the raw antenna and particle-detector voltage data to reconstruct the properties of a cosmic ray. Initial pipeline: ported version of software currently used with LOFAR	
5b2	Near-field imaging	A user would analyse the raw antenna voltage data with near-field interferometry to produce an image of the particle cascade. They would then run simulations of particle cascades and fit them to the image, or analyse the image with their own software. Near-field imaging techniques are currently still under development.	

4 References

4.1 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] SKAO Science Data Products: a Summary (SKA-TEL-SKO-0001818), Revision 01

4.2 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] SRCNet User Stories from SWGs, https://docs.google.com/spreadsheets/d/19rSGUl5L6F4t-M8geGYmRyp-vZuHUpAKYS8GFs7TA3E/edit#gid=1824017509

[RD2] Code for LOFAR single station imaging, https://github.com/lofar-astron/lofarimaging

DOCUMENT HISTORY

Revision	Date Of Issue	Engineering Change Number	Comments
01	2023-06-20	-	First released version

DOCUMENT SOFTWARE

	Package	Version	Filename
Word processor	MS Word	Office 365	SKAO-TEL-0000000-01C_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