UK Square Kilometre Array Regional Centre (UKSRC) Update

Ian Collier, Jeremy Yates

GridPP 48, Ambleside

1st September 2022

SKAO REGIONAL CENTRE NETWORK

The SKAO Project in a Nutshell

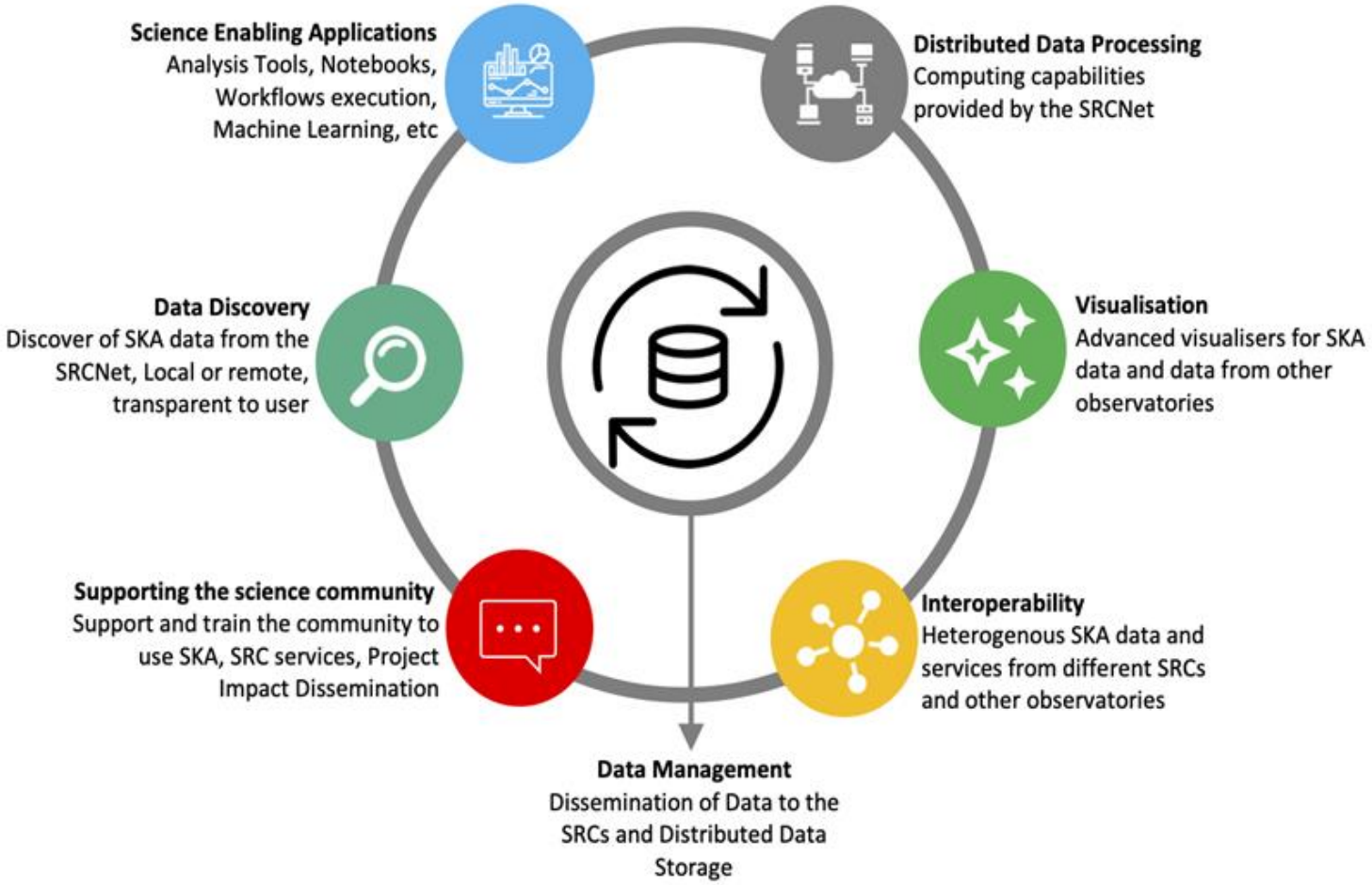
SKA Scientific Timeline

The Pre SRCNet deployment is timeline is above.

- We are in the prototyping phase
- In 2023-24 the Implementation and deployment phase will begin
- The target is the AA0.5 Deployment and testing phase, which allows testing of the Telescope-SRCNet functions to be tested.

Activity	**SKA-LOW**		**SKA-MID**	
	Date	**# of stations**	**Date**	**# of dishes**
Start of Construction	Jul 2021		Jul 2021	
Start of major contracts	Aug 2021		Aug 2021	
Finish of Array Assembly 0.5 (AA0.5)	Feb 2024	6	Mar 2024	4
Finish of AA1	Feb 2025	18	Feb 2025	8
Finish of AA2	Feb 2026	64	Dec 2025	64
Finish of AA3	Feb 2027	256	Jun 2026	133
Finish of AA4	Nov 2027	512	Jun 2027	197
Operations Readiness Review	Jan 2028		Dec 2027	
End of Construction	July 2029		July 2029	

What is SRCNet



- SRCNet is the network of SKA Resource Centers (SRCs)
- Can think of SRCNet as logically like WLCG
  - UKSRC is analagous to GridPP

- Individual SKA Resource Center (SRC) nodes in member countries
- Provides all elements in the beginning-to-end lifecycle of SKA science projects
- Compute, Storage, science data archive
- The interface by which the community access and interact with SKA data and undertake SKA science

SRCNet

- An operating SRCNet and UKSRC Node is needed by 2025
  - Needs 10-20% of the capacity and 80% of the functionality of what is needed by 2028-29 for full SKA telescope operations.
- Networking infrastructure has to be in place during 2022-23 to enable the SRCNet to be developed.

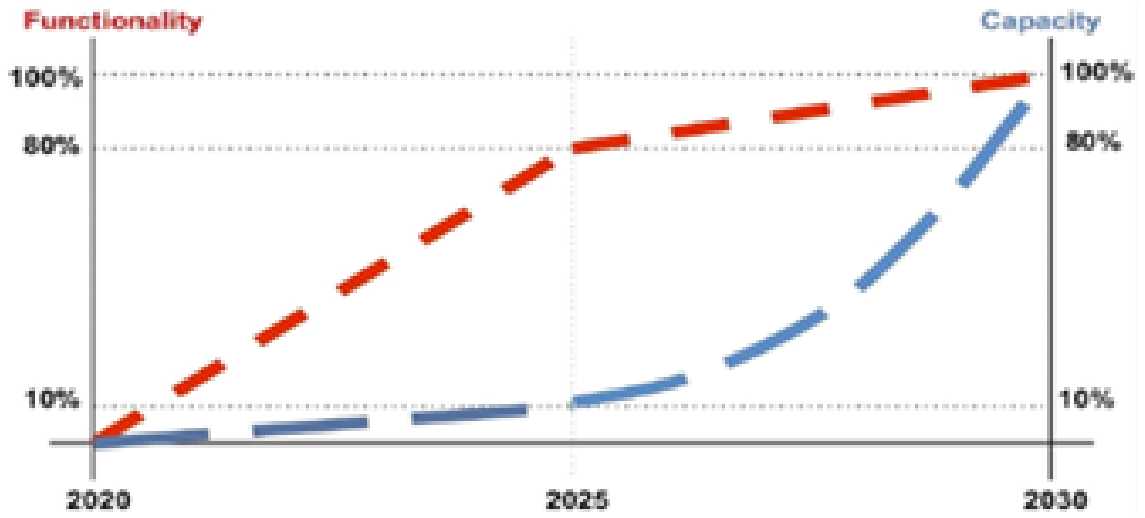


Figure 1.2: The Ramp up of the SRC Network from 2020-2030. The red dashed line denotes functionality, and the blue dashed line dendes capacity. This growth profile of the SRCNet required to support the deployment schedule of the SKA and deliver the scientific data products.

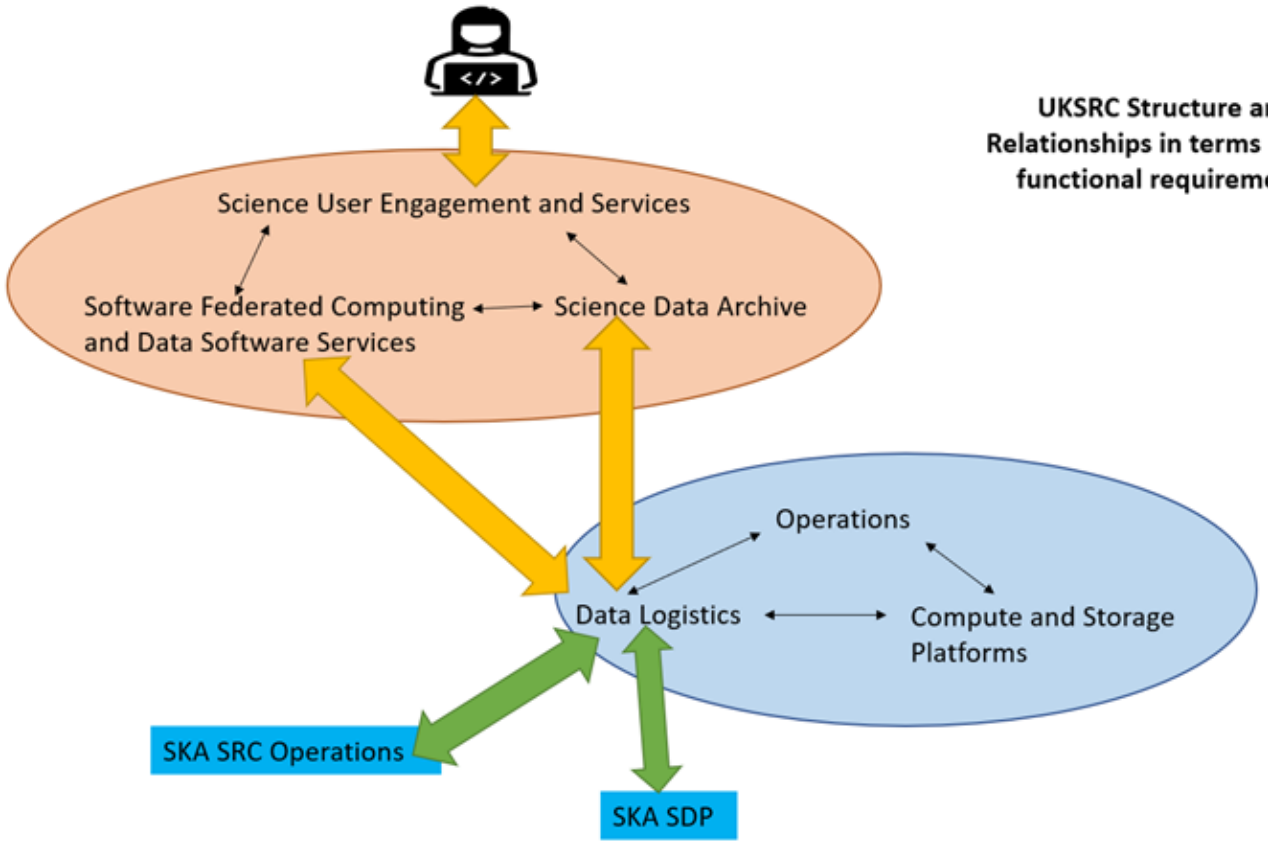
SRCNet

- The SRCNet programme will prototype, design, implement and test the following areas:
  - SRCNet Design and Architecture
  - Data Logistics
  - Operations
  - Federated Computing
  - Data Software Services
  - Science Data Archive
  - Compute and Storage Platforms
  - Science User Engagement

- End goal is functioning and capable SRCs and global; SRCNet

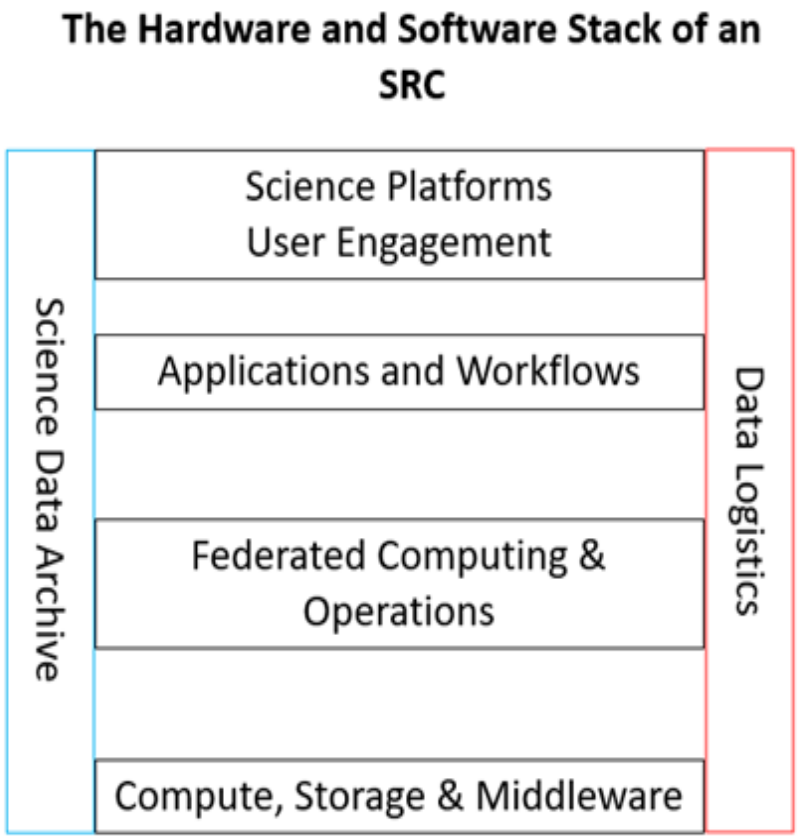
SKAO REGIONAL CENTRE NETWORK

(UKSRC)



- Maximise impact of UK investments in SKAO
- Lead internationally in the development of the global SRCNet - ensuring the success of SKAO-SRCNet
- Support UK leadership across SKA science, and related data intensive supercomputing, big data etc. over the expected 50-year lifetime of the SKA
- Ensure UK researchers are engaged with, lead and gain competitive advantage from discoveries by providing a world leading UKSRC

The Hardware and Enabling Software Stack



- There are four distinct horizontal layers in the stack. Each stack has to function in concert with the stack above/below it to deliver the service the user/application/middleware requires. This requires a strong set of interfaces (APIs) between each layer to ensure good performance and reliability
- There are two distinct vertical layers that interact with each other and the four horizontal layers. These vertical layers are key to the data transport, data distribution and data archiving functions of the SRCNet and proto-SRC
    - The interfaces (APIs) of these layers need to be able to interface with all the layers in the proto-SRC Stack as well as with the SRCNet SRCs
- The Data Logistics Layer (data transport and Networking) is the Key Functionality to ensure the SRCNet and local SRCs can function. The software must be common to all SRCs
- The Science Data Archive layer allows all data products to be stored, catalogued and curated (e.g. addition of Meta Data). This layer must be composed of software common to all SRCs
- The Compute, Storage and Middleware enable the data and applications to be stored and computed and so generating the advanced data products for scientific analysis.
- The Federated Computing and Operations Layer allows access to the computational resources and the required data (which may be distributed among several sites), so allowing workflows and their constituent applications to be to be run.
- The Applications and Workflow Layer contains the software required to perform processing, analysis, modelling, curational functions on the data.
- The Science Platform and User Engagement Layer provides the portal by which Research Projects and users engage with the SKAO data and generate scientific data and analyses for publication and dissemination.

Data, WLCG, GridPP

- Dave mentioned GridPP support for SKA use of LHC technology

- Global data movement will be familiar to us in many ways
  - Rucio, FTS3 & xrootd
    - (They do need to address metadata)
- Some important differences
    - Many data sets MUCH larger – up to multiple TBs
    - Telescope arrays are somewhat remote with relatively limited bandwidth
    - Science groups much smaller – global movement of data coordinated by SRCNet, not VOs

- GridPP expertise will be invaluable

- Need to build up dedicated UKSRC data management capability at sites like RAL/Manchester
  - Started with the introduction to xrootd Rose mentioned
  - More will follow

Data Volumes and movement

- Current model
  - 100gb/s from each of SKA MID and SKA LOW
    - With no overhead could mean 350PB/year from each → ~700PB/year total
  - All data will go to at least two regional (national) SRCs → ~1.4EB/year
  - Realistically probably closer to 800PB/year without faster connections
  - UKSRC (in practice RAL) *might* receive 1/6th of that → more than 130PB/year
    - Goes to tape and distributed around UK cloud sites for analysis
      - A lot of work on how to do this
      - The optimization work Duncan described will be vital
    - RAL will be using existing technology (Antares & ECHO) developed for GridPP

- Global data movement
  - Prototyped in ESCAPE project
  - So far focused on rucio, FTS, xrootd
    - Which are familiar to us
  - Opportunity to leverage GridPP expertise in developing UKSRC capabilities

SRCNet protypes

- Prototyping work began in earnest earlier this year
- Agile teams focused on different areas

- In UK participation funded by bridging grants
  - Bid for 3 year UKSRC project submitted in mid August
- UK participating in several prototype teams
  - Leading AAI team
  - Preparing to lead work developing policy framework
    - Drawing upon WLCG/EOSC/IRIS experience
    - Funded through EOSC Future

UKSRC Architecture

- A multi-cloud: Big OpenStack clouds for data processing and analyses at RAL and Cambridge
  - smaller ones elsewhere
- RAL as bulk data hub – the UK Connection to the Telescopes
  - Data Archive layer that talks to bulk data
  - Mechanism and network for staging data to clouds
- Data Management and the UKSRC Data Lake
- Data Transport Service
- Authorisation and Authentication
- Development activities at SCD, Manchester & Cambridge

UKSRC and IRIS

- UKSRC is an IRIS Project.
- Provider Partners will be SCD, DiRAC and GridPP
- It will have its own capital line from 2023 that will be administered and scrutinized by the IRIS RSAP
- Covering hardware, enabling software and digital assets
  - There is a separate line for Networking costs
- There is huge scope for collaboration with GridPP and the LCG
  - What can we do in common in terms of core infrastructure?
  - RUCIO, Indigo-IAM, Job provisioning, shared hardware, Openstack, HelpDesk

Outline of UKSRC three year project

- Aim to deliver 80% of capability, 20% of capacity
- Developing
  - Underlying (openstack) compute capabilities
  - Data movement/management infrastructure
  - UK part of Science data archive
  - UK part of science analysis capabilities
- Distinct capital funding for hardware and some digital assets through IRIS
- Also supporting UK leadership in global SRCNet effort
- ~30FTE/year and a £6M Capital are needed to achieve these goals