SKA Regional Centres Network (SRCNet): Addressing Large-Scale Data Challenges

Jesús Salgado
SKA Regional Centres Network Architect

SKA1-low – the SKA's low-frequency instrument

The Square Kilometre Array (SKA) is a next-generation radio astronomy facility that will revolutionise our understanding of the Universe. It will have a uniquely distributed character: **one** observatory operating **two** telescopes on **three** continents. Construction of the SKA will be phased and work is currently focused on the first phase named SKA1, corresponding to a fraction of the full SKA. SKA1 will include two instruments – SKA1-mid and SKA1-low – observing the Universe at different frequencies.

SKA SQUARE KILOMETRE ARRAY

Location: Australia

Frequency range: **50 MHz** to **350 MHz**

~131,000 antennas spread between **500 stations**

Total collecting area: **0.4km²**

Maximum distance between stations: **>65km**

SKA1-low

Data transfer rate: **7.2 Terabits** per second

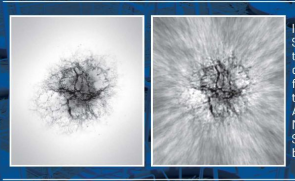


Image quality of SKA1-low (left) versus the best current facility operating in the same frequency range, the LOw Frequency ARray (LOFAR), in the Netherlands (right). SKA-low's resolution will be similar to LOFAR.

Compared to LOFAR Netherlands, the current best similar instrument in the world

25% better resolution

8x more sensitive

135x the survey speed

www.skatelescope.org @SKA_telescope SKAtelescope ska_telescope Square Kilometre Array ska-organisation

SKA1-mid – the SKA's mid-frequency instrument

The Square Kilometre Array (SKA) is a next-generation radio astronomy facility that will revolutionise our understanding of the Universe. It will have a uniquely distributed character: **one** observatory operating **two** telescopes on **three** continents. Construction of the SKA will be phased and work is currently focused on the first phase named SKA1, corresponding to a fraction of the full SKA. SKA1 will include two instruments – SKA1-mid and SKA1-low – observing the Universe at different frequencies.

SKA SQUARE KILOMETRE ARRAY

Location: South Africa

Frequency range: **350 MHz** to **15.3 GHz** with a goal of 24 GHz

197 dishes (including 64 MeerKAT dishes)

Total collecting area: **33,000m²**

or **126 tennis courts**

Maximum distance between dishes: **150km**

SKA1-mid

Data transfer rate: **8.8 Terabits** per second

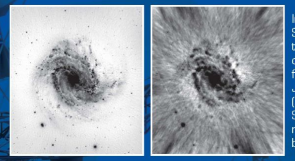


Image quality of SKA1-mid (left) versus the best current facility operating in the same frequency range, the Jansky Very Large Array (JVLA) in the United States (right). SKA-mid's resolution will be 4x better than JVLA.

Compared to the JVLA, the current best similar instrument in the world:

4x the resolution

5x more sensitive

60x the survey speed

www.skatelescope.org @SKA_telescope SKAtelescope ska_telescope Square Kilometre Array ska-organisation

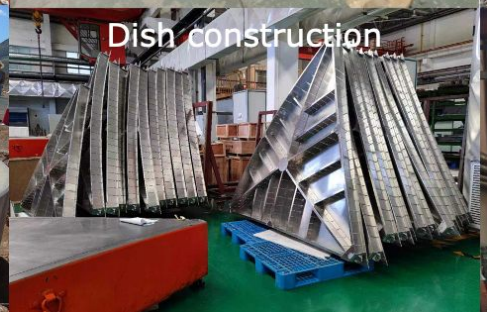

SKAO Science Working Groups
Cosmology
Cradle of Life
Epoch of Reionization
Extragalactic Continuum
Extragalactic Spectral Line
Gravitational Waves
High Energy Cosmic Particles
HI Galaxy Science
Magnetism
Our Galaxy
Pulsars
Solar, Heliospheric & Ionospheric Physics
Transients
VLBI

Staged Delivery Strategy

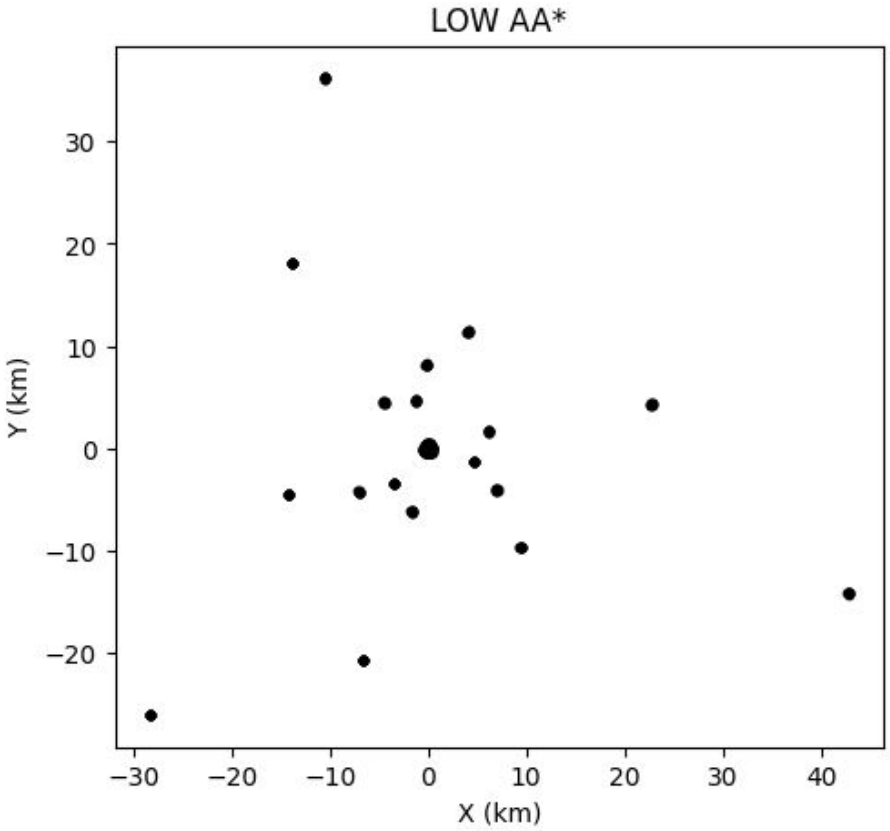
Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 May	2025 Nov
AA2	**64 dishes** **64 stations**	**2027 Apr**	**2026 Dec**
AA*	144 dishes 307 stations	2028 Jan	2028 Mar
Operations Readiness Review		2028 Apr	2028 Aug
AA4	197 dishes 512 stations	TBD	TBD

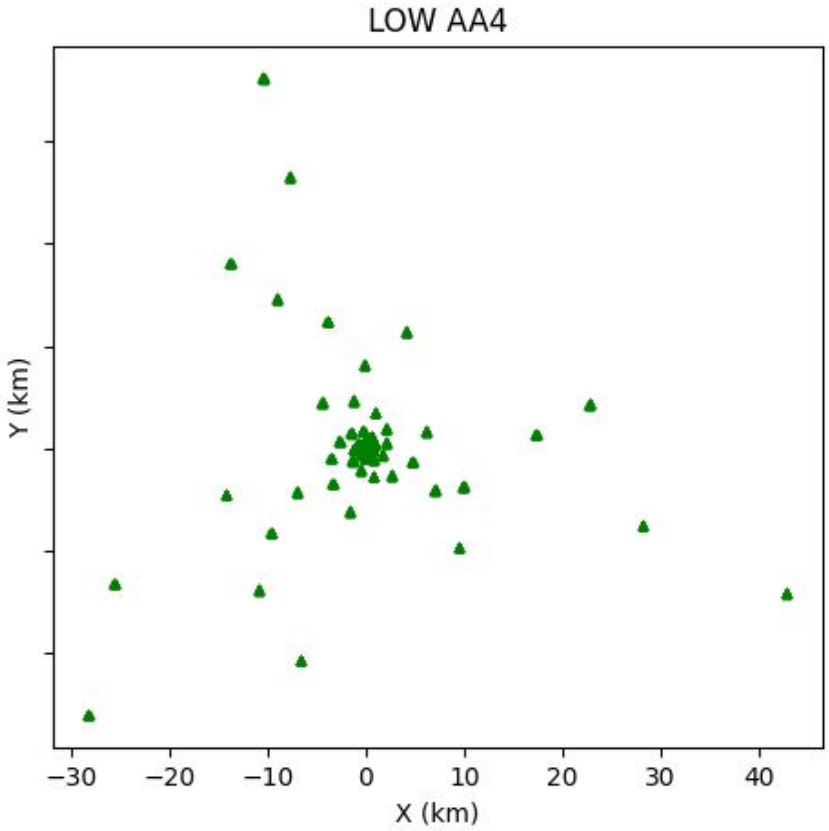
SKA-Low in country developments

SKA-Mid in country developments

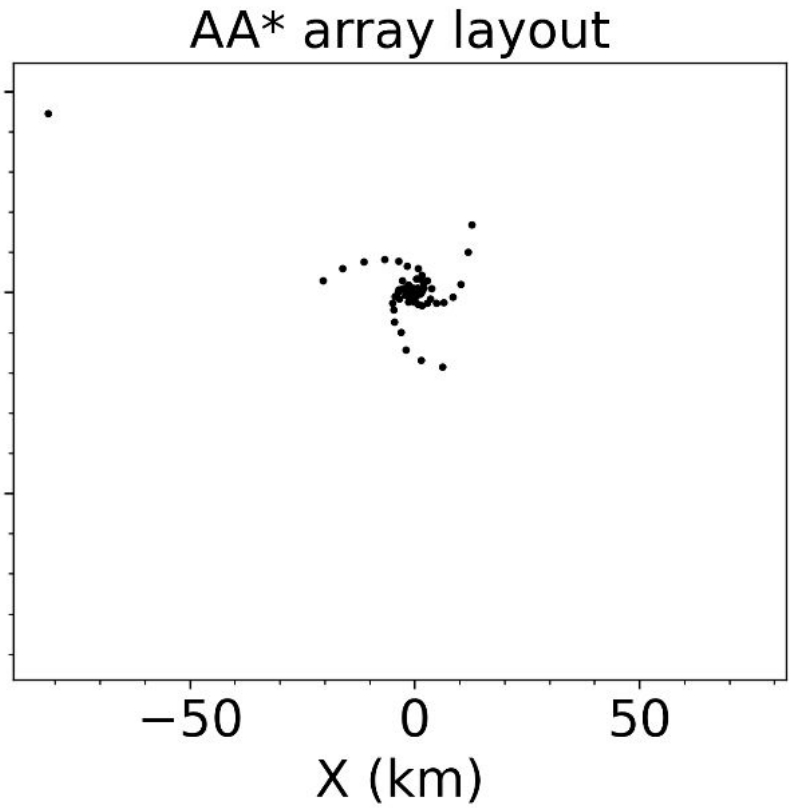


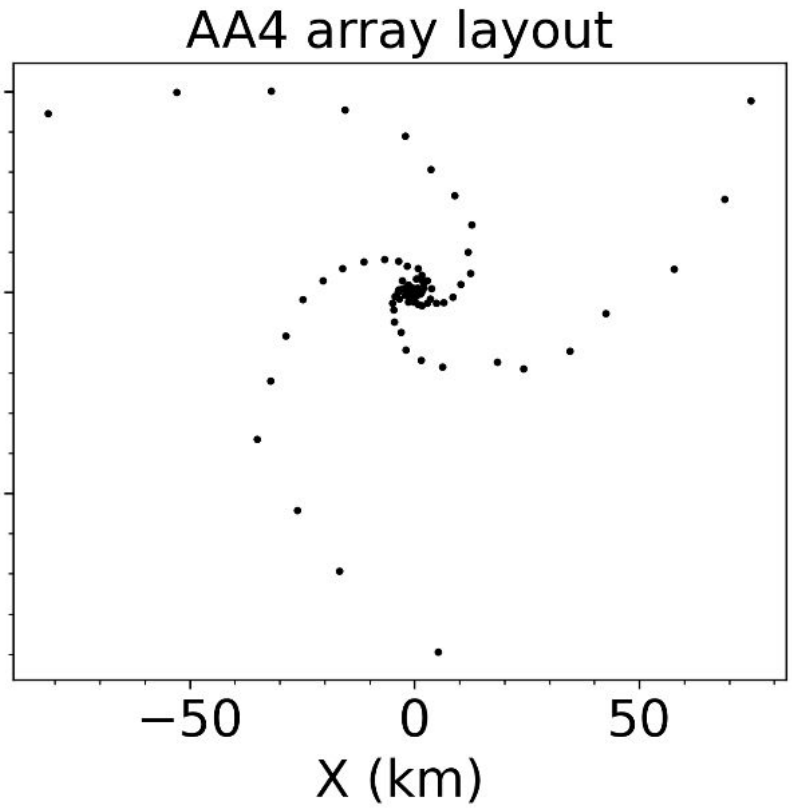
Staged delivery examples: LOW (example only)



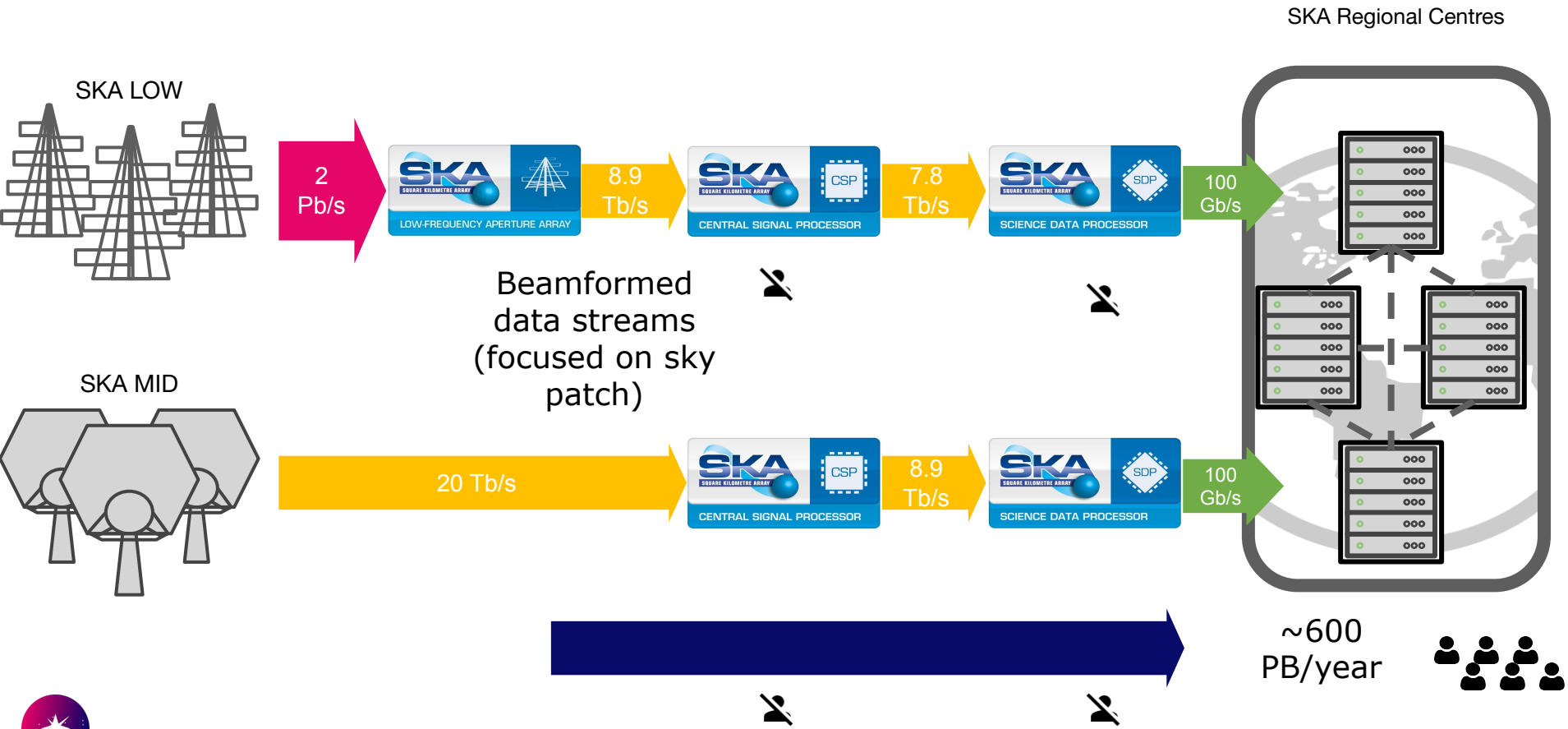


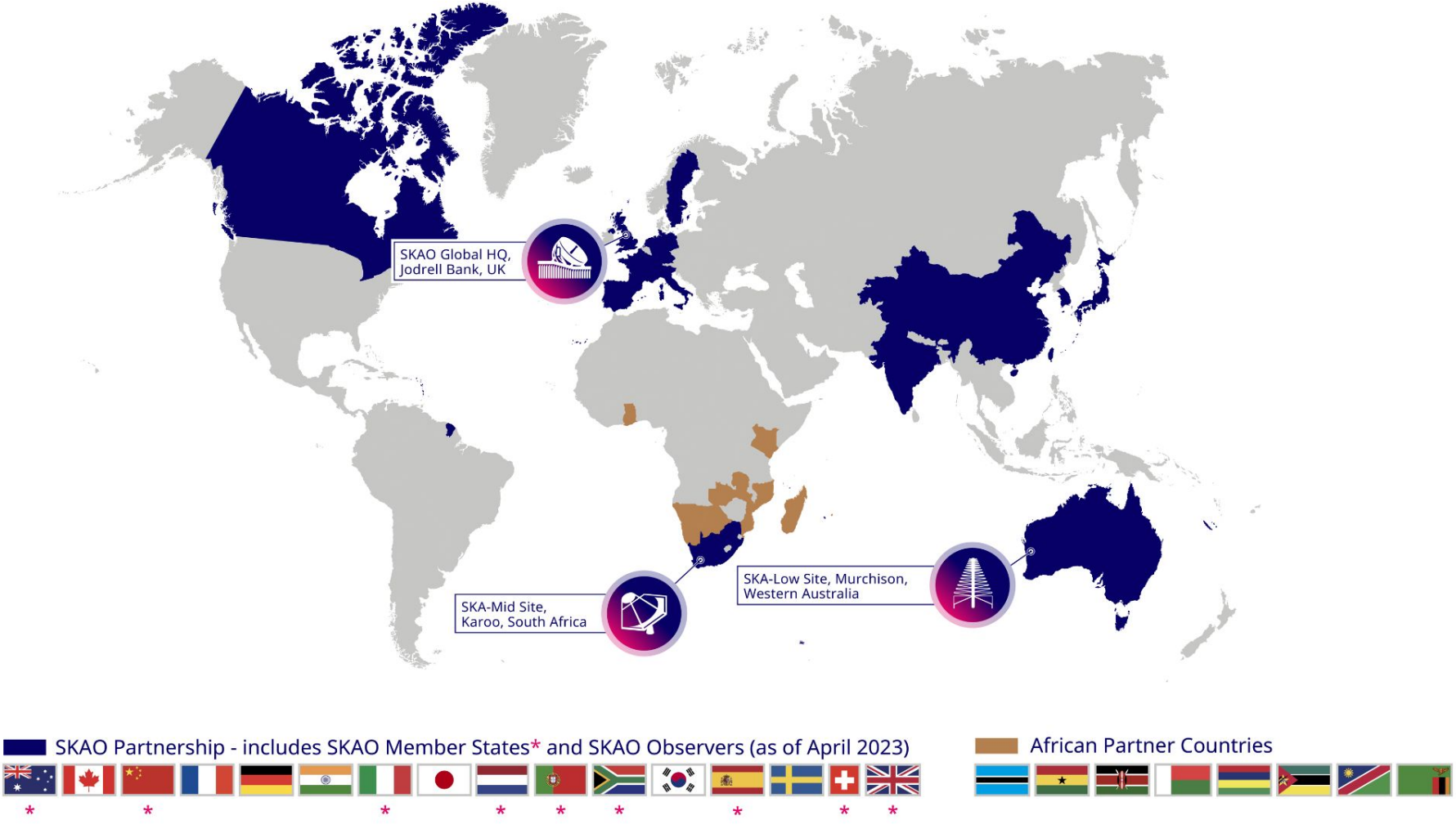
Staged delivery examples: MID (example only)



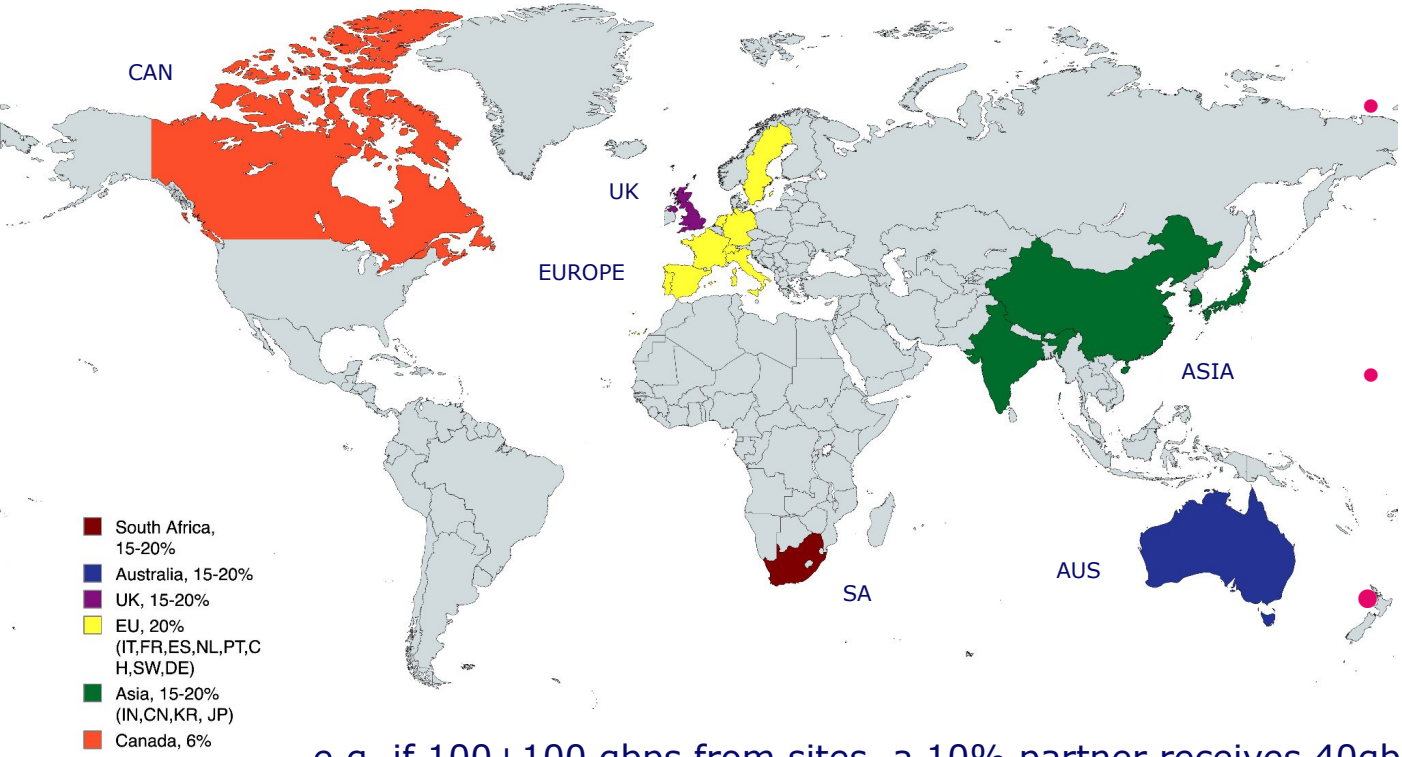


SKA Data Flow (rates for AA4)




SKAO Global HQ,
Jodrell Bank, UK
SKA-Mid Site,
Karoo, South Africa
SKA-Low Site, Murchison,
Western Australia
SKAO Partnership - includes SKAO Member States* and SKAO Observers (as of April 2023)
African Partner Countries

SKA Regional Centre Broad Distribution: Fair Share, AA4 data rates

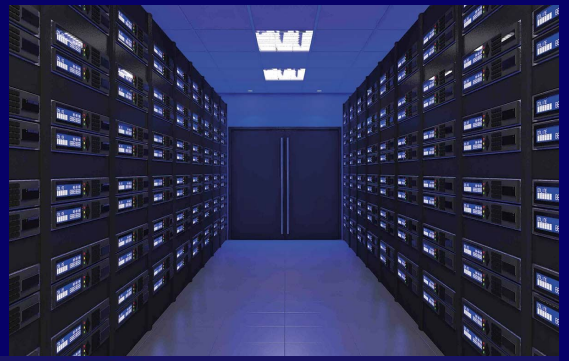


- Roughly, 6 global zones of equivalent size (Canada smaller)
- **Distribute two base copies** of each data product to different countries, and perhaps insist to different regions
- Average incoming rate per (20%) region not more than 2x40 Gbit/s = 80Gbit/s (~2x12 Gbit/s for Canada)
- **Max 100 Gbit/s out of SA and AUS**

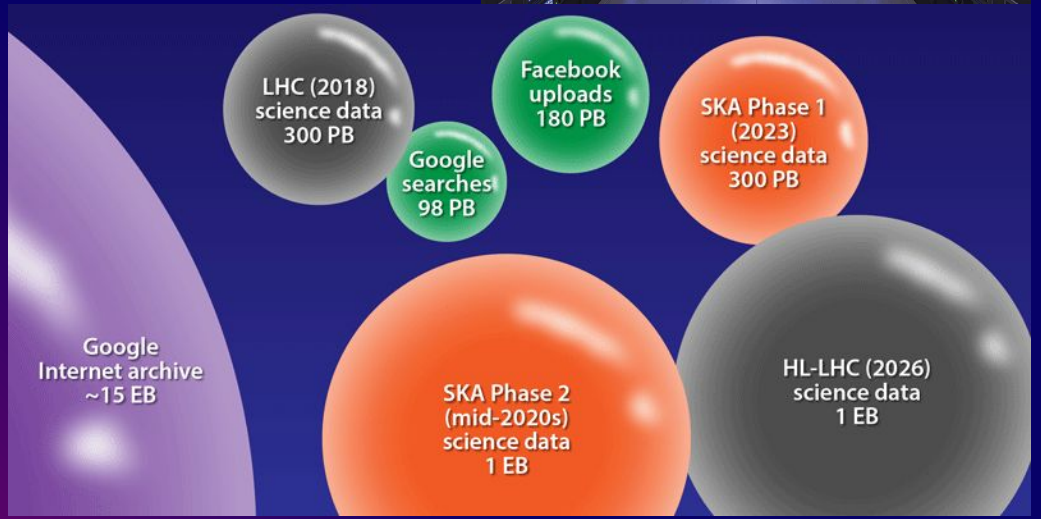
e.g. if 100+100 gbps from sites, a 10% partner receives 40gbps data (400 TBytes per day, 140 PBytes per year)

SRCNet in Numbers

- ~ 600 PB/year of Scientific Data
- 16 countries involved
- Up to 100 FTEs during development phase



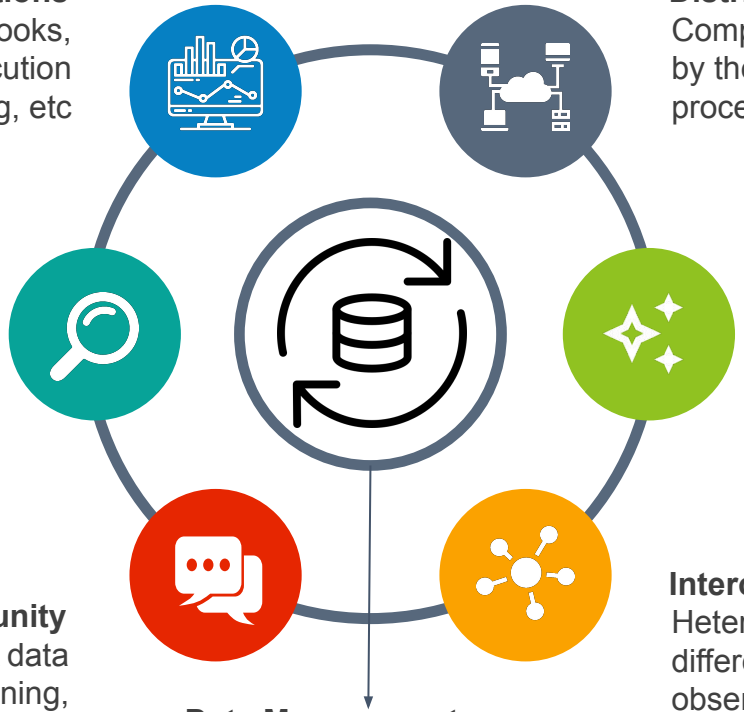
- Collaboration agreements with CERN, GEANT, CTAO
- Collaborations with CNRS, Vera Rubin and others



SKA Regional Centre Capabilities Blueprint

Science Enabling Applications
Analysis Tools, Notebooks, Workflows execution Machine Learning, etc

Distributed Data Processing
Computing capabilities provided by the SRCNet to allow data processing



Data Discovery
Discovery of SKA data from the SRCNet, local or remote, transparently to the user

Visualization
Advanced visualizers for SKA data and data from other observatories

Support to Science Community
Support community on SKA data use, SRC services use, Training, Project Impact Dissemination

Interoperability
Heterogeneous SKA data from different SRCs and other observatories

Data Management
Dissemination of Data to SRCs and Distributed Data Storage

SRC Network global capabilities

1 Data Management

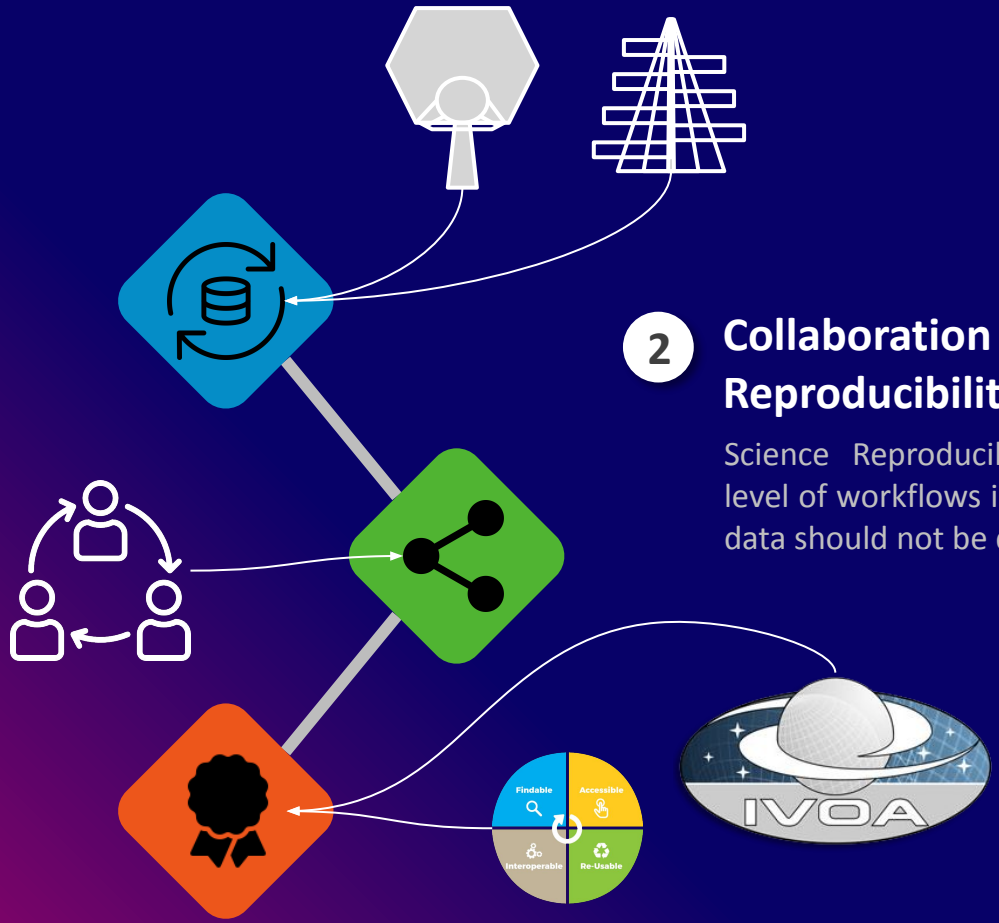
Avoid unnecessary duplication and transfers
Roughly 5-10 million dollars per year in new data, for one copy

2 Collaboration and Reproducibility

Science Reproducibility at the level of workflows is essential as data should not be downloaded

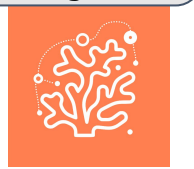
3 Use of Standards

Build SKA science archive around FAIR and IVOA standards



The *current SRCNet* SAFe teams

Spain, Switzerland, Sweden, Portugal...



Coral

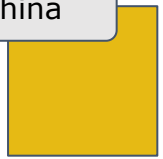
UK, Authentication

Purple

Australia, Japan, Korea

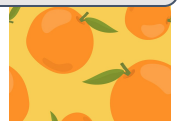
Lavender

China

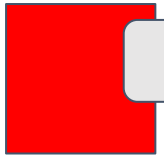


Gold

Visualisation, Italy, CDS,...



Orange



Canada + India

Red

SKAO

Magenta

UK SRC

Teal(s)

Gateway, Netherlands

Tangerine

Janneke de Boer
Robert Perry

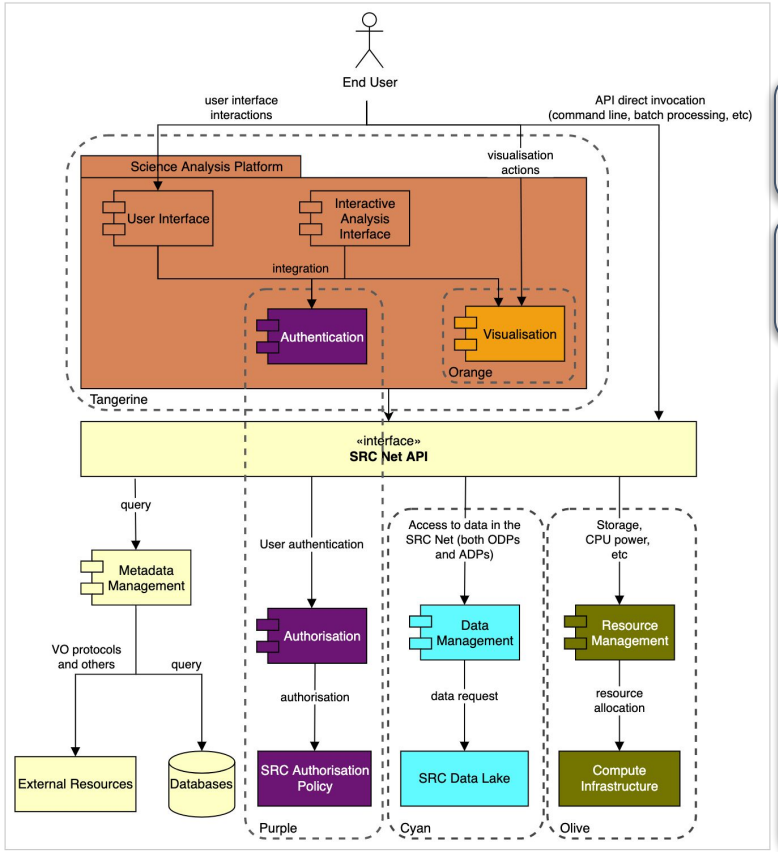
Rosie Bolton
Program Manager

Jesus Salgado
Software Architect

Jeremy Coles
Release Train Engineer

Since June 2022 we have been working as a team-of-teams
Engagement from across most SKA countries
About 70 contributors - 25 persons-worth of effort

Defining an architecture



Users

Science Platform

Interface (API layer)

Metadata query - Science Data Discovery

Authentication Who? Permissions?

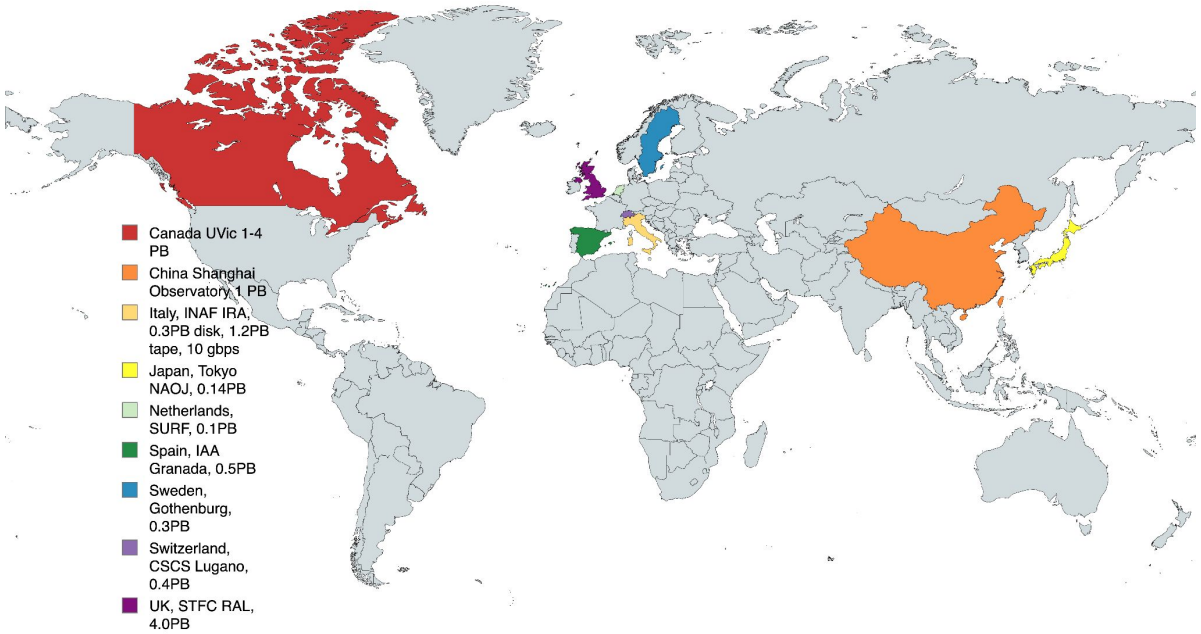
Data Logistics Globally distributed storage sites

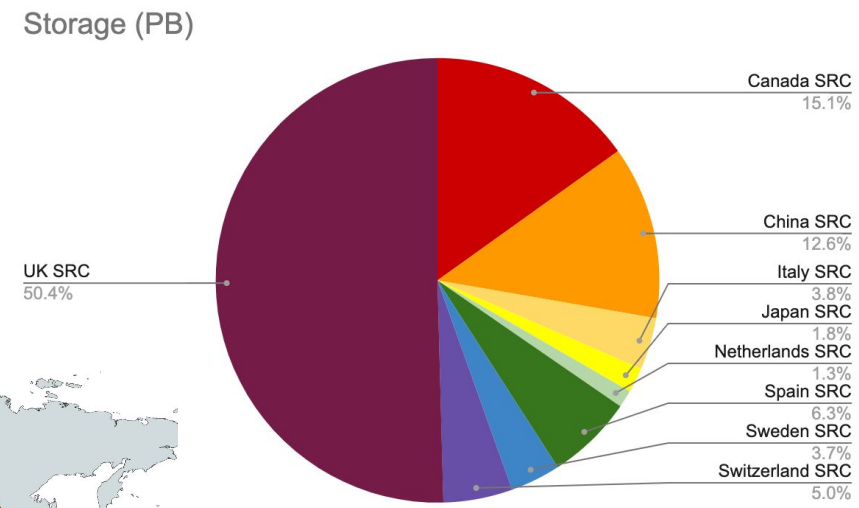
Compute Resource Management Work sharing

SRCNet0.1 intended sites

January 2025

8 PBytes total storage offered for SRCNet0.1 (c.f stated target of 20 PB)

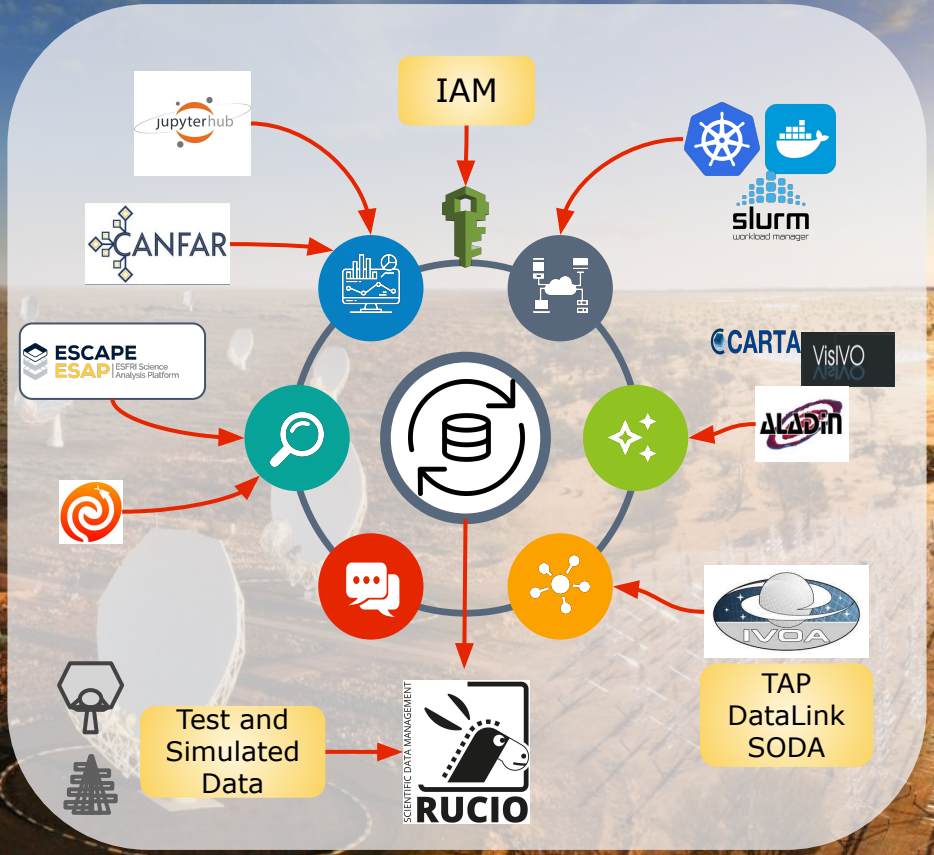




WLCG experience at some sites (Canada, Netherlands, Sweden, Switzerland, UK)

Some sites quite new and teams will learn by being involved

SRCNet v0.1 Software Stack



IVOA - SRCNet connections

Consider a Planned, Co-located data scenario

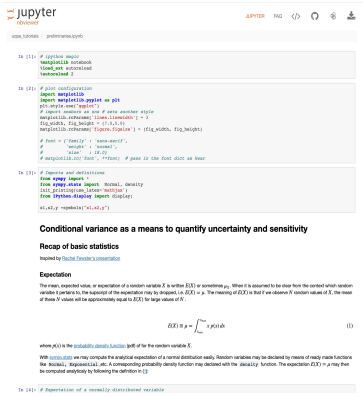
Plan data placement ahead of time

Allocation for compute resources made for projects user group at target SRC

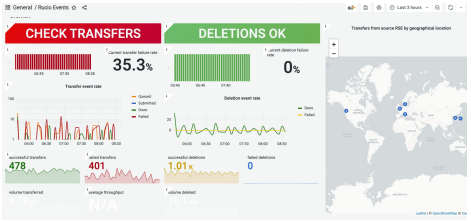
SKA SQUARE KILOMETRE ARRAY — SDP — SCIENCE DATA PROCESSOR

100 Gb/s

SKA Data products generated



Products transferred to target SRC



from our live Rucio instance

posed by actors!

Users can be allocated to different SRCs based on data location, HW mapping to compute needs; to even out demand, or other reasons

Type 2: Distributed data scenario, server side actions

Moving compute to the data

Server-side actions

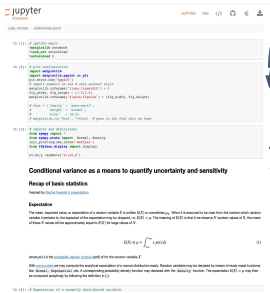
Likely public data

Not SKA Programme Users

Hard to predict

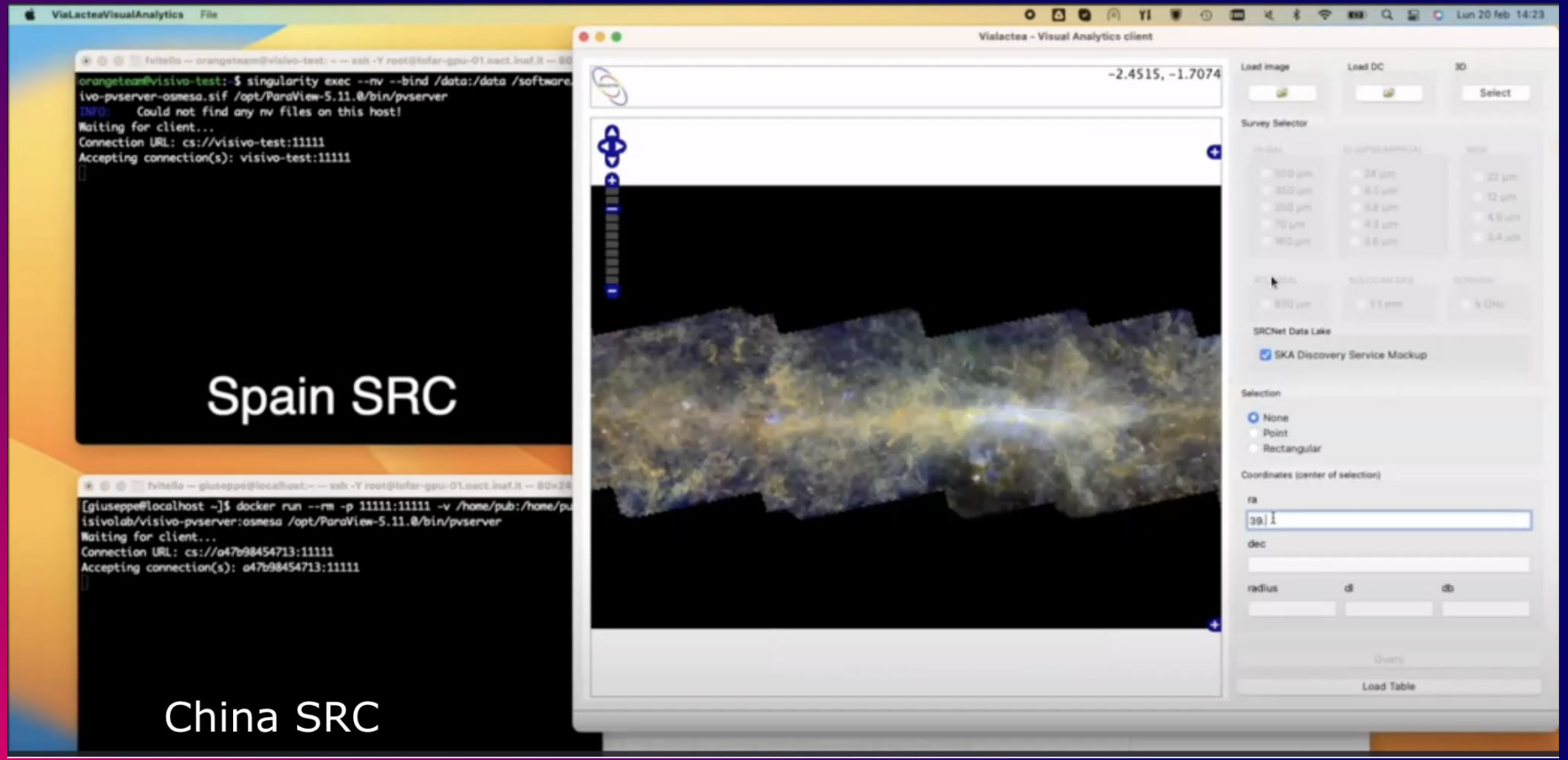
Single workflow running at primary site launching multiple secondary server-side actions at sites holding data

e.g. return image statistics for small area around source location (e.g. re-do source parameter analysis using location from a catalogue), just return the parameters



Imagine main workflow running in (e.g.) Aus

Remote Data Visualisation using SODA

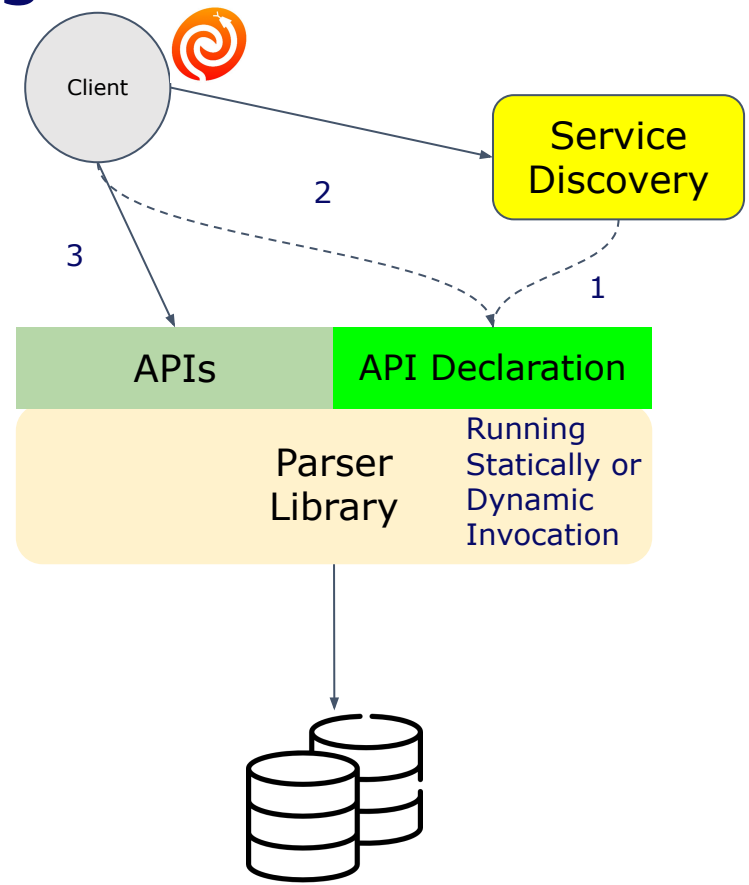


Some possible SODA “extensions” - API standard

Data Type	Operation	Input	Output
Any Type	Get Stream	ID	Input Stream
Data Cube	Cut-out	ra, dec, size, resolution	Data Cube
Data Cube	Get Spectra	ra, dec, size	Spectrum
Data Cube	Get Time Series	ra, dec, size	Time Series
Data Cube	Get Slice	wavelength	Image
Image	Change Resolution	ra, dec, size, resolution	Image (FITS to HiPS)
Image	Source Extraction	ID, algorithm params	Source Catalogue
Spectrum native	Convert to VO	ID	Spectrum VO
Source Catalogue	Similar Source	Source ID	Source Catalogue

Remote Data Parsing Operations

- Protocol close to IVOA SODA including other operations
- Operations to be included will be discussed and agreed due to scientific priority and feasibility (extension of current SODA services)
- Possible use of OpenAPI (this is under discussion at IVOA level)



Execution Broker (Execution Planner)

IVOA possible support

- IVOA data model (and formats) for Radio-Astronomy
  - Including support to complex data sets (hierarchical structure)
- IVOA provides discovery and access protocols for most of the astronomical data
  - Standards, Integration with scripting languages, Easy publication and collaboration environments
- Many astronomical use cases are enabled thanks to IVOA standards

PROMOTE

- Possible “interoperable science platform” new phase with:
  - Federated Authentication Protocols
  - Improved data access (“bulk”)
  - Remote operations
  - (Simplified) federated execution
    - Execution broker
    - Software discovery and
    - Software characterisation

PROMOTE

NEW?

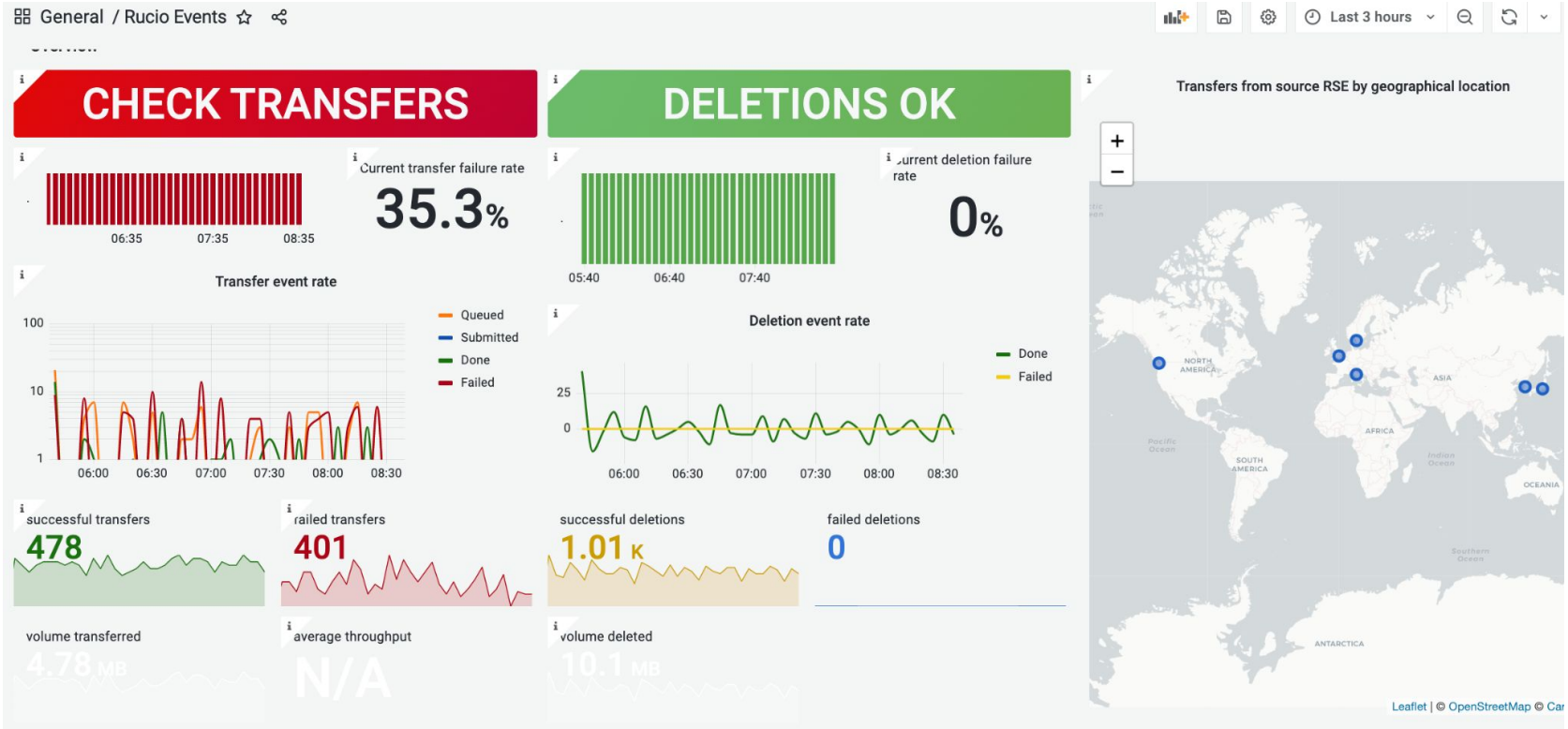
EXTEND

NEW API

STANDARD

Thanks for your attention

(Aside) We are prototyping Rucio (continuing our ESCAPE work) for this data placement scenario



SKA-Mid in South Africa

- 197 fully steerable dishes, including the existing MeerKAT dishes
- Freq range covering 350 MHz to 15.4 GHz
- Max baseline is 150 kms (0.22” at 1.7 GHz, 13.4mas at 15 GHz)
- Located in the Karoo, South Africa
- Central Signal Processor, Science Data Processor in Cape Town