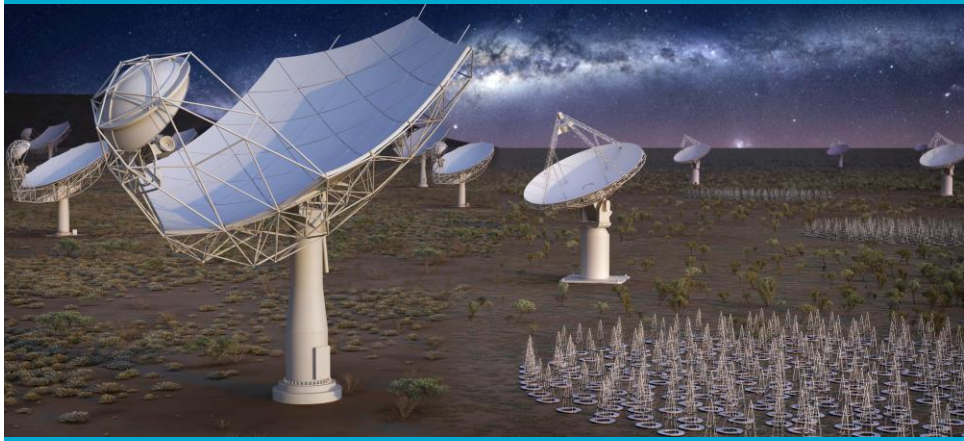
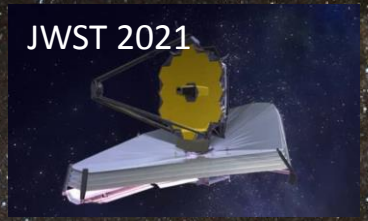
The Square Kilometre Array Computing

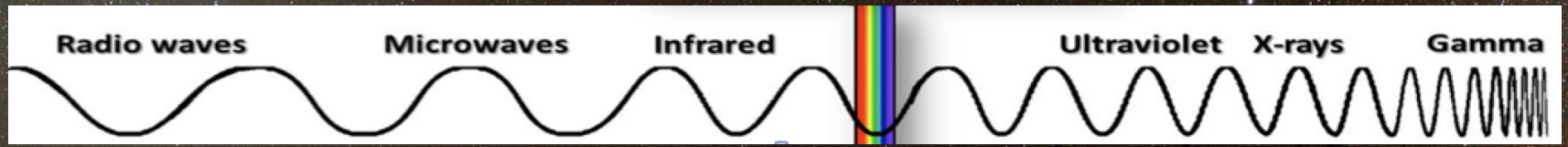
From SDP to SRC

Minh Huynh | 4 Nov 2019
Senior Data Scientist and Research Astronomer



21st Century Telescopes and Observatories





Square Kilometre Array:

Billion-dollar+ project to build the world's largest Radio Observatory.

Two ground breaking radio interferometers, 50 year life span

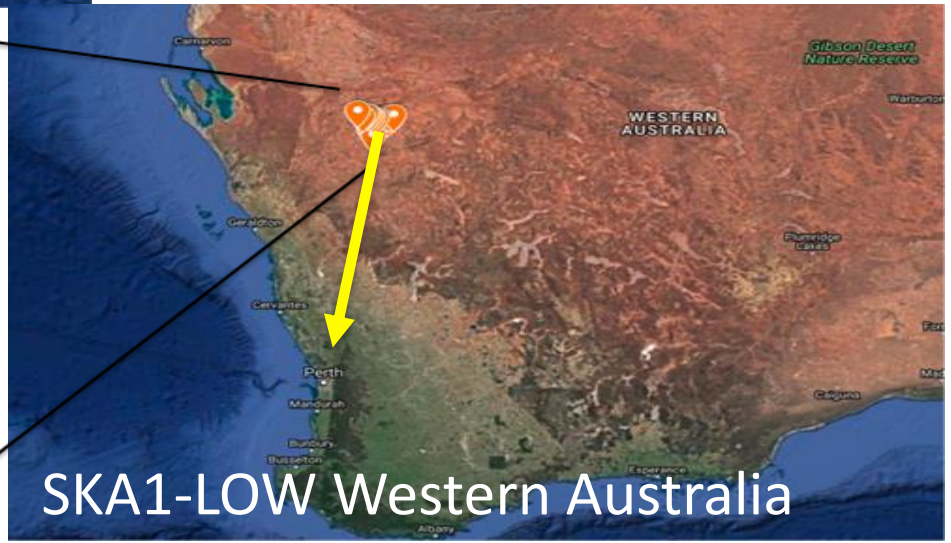
SKA: A Global Project

SKA Observatory convention signed by : AUS, CHN, ITA, NED, POR, RSA, GBR

Other SKA Organization members: CAN, ESP, IND, SWE, FRA, GER, NZL. Interest from: JPN, KOR, SUI

WHERE will SKA be built?

SKA1-MID South Africa



SKA1-LOW Western Australia

SKA1-Mid, South Africa

133 SKA1 dishes (15m) + 64 MeerKAT (13.5m) dishes

Densely populated core (~ 2-km diam) + 3 log-spiral arms, 150-km baselines

0.35 → 15 GHz covered in 5 bands, 65,000 frequency channels

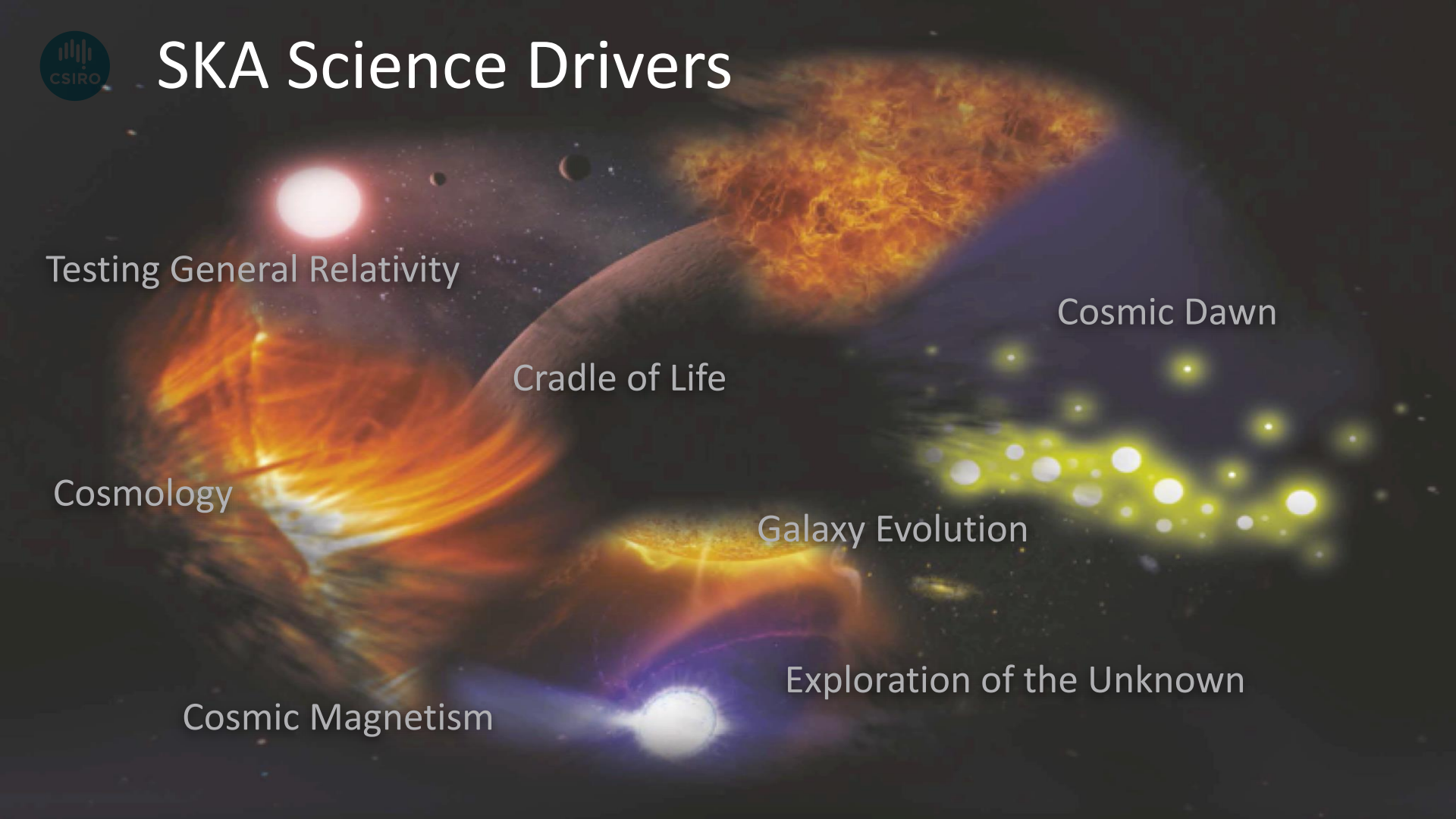
SKA1-Low, Western Australia

131,072 antennas: 512 stations each of 256 antennas

Densely populated core (~ 1-km diameter) + 3 log-spiral arms, 65 km baselines

50 → 350 MHz full instantaneous bandwidth, 65,000 frequency channels

SKA Science Drivers



SKA Status

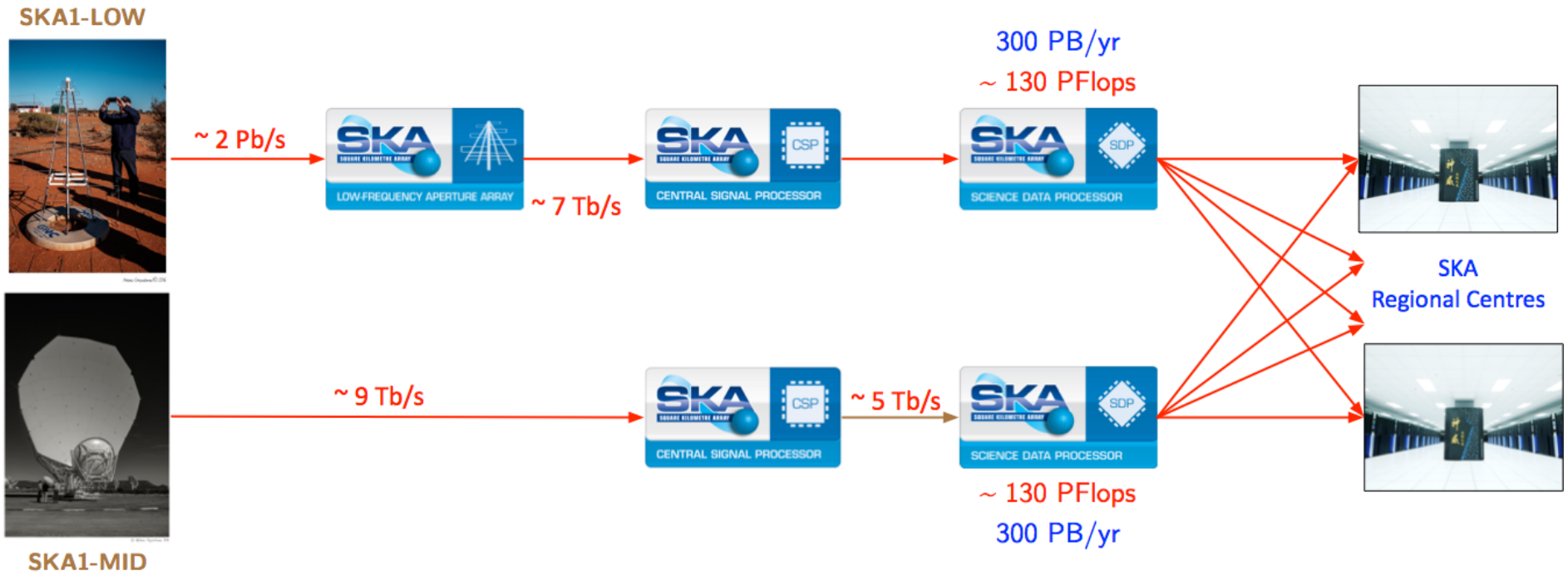
- Design consortia have completed Critical Design Reviews (2018/2019)
  - Science Data Processor CDR completed in May 2019
  - System CDR planned in Dec 2019

- Construction to begin 2021
  - Staggered array releases from 2024
  - Science verification from 2026
  - Operations from 2028

AAVS1.5 from above, populated with 48 antennas. Image credit: ICRAR/Curtin

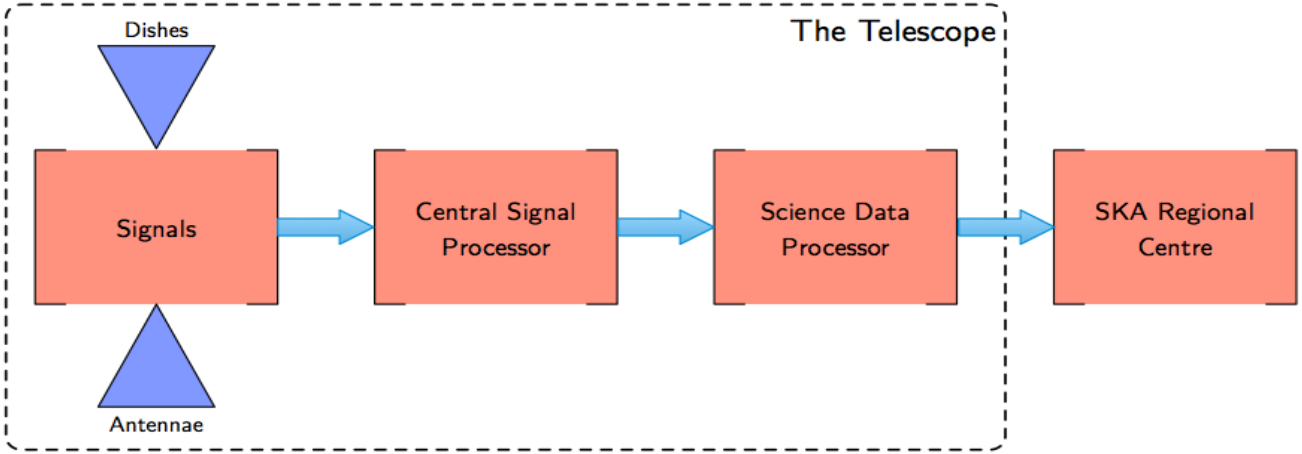
SKALA-4.1 antennas on the AAVS1.5 station. Image credit: ICRAR/Curtin

How will data be handled?



10 – 50x data rate reduction by Science Data Processor

Data Challenges



- Data Rates and volumes from central signal processor are so high we can not store the raw data from it
  - Cheaper to re-observe than store the raw data indefinitely
- The science data processor becomes a schedulable resource for the telescope for observation planning

Science Data Processor Challenges

Complexity

- Multi-axis datasets
- Iterative convergent pipelines
- Must be predictable for scheduling – but each scheduling block different
- Orchestration required for data ingest, processing, control, preservation and delivery

Scalability, Extensibility, Lifetime

- ~0.25 ExaFLOPS of compute (10% efficiency)
- HW and SW will change – refresh cycles for HW and new algorithms
- Near 365 days a year operation, 50 year lifetime
- HPC and HTC landscape will evolve

Science Data Processor Challenges

Cost

- constrained on capital cost - SKA1 is a 674M (2016) EURO project
- SDP around 10% of that (HW and SW)

Power

- Most green supercomputer ~**70 MW** / ExaFLOPS
- Current Exascale roadmap indicates ~**40 MW** / ExaFLOPS in 2023
- We need about 4 times better…

SDP headline design numbers*

Input

- ~800 GByte/s INGEST (in total), from Central Signal Processor

Temporarily store

- Data set up to 15 PBytes
- 100PBytes total distributed buffer

Process

- 250 PFLOPS total peak
  - 10% efficiency assumed

Preserve and ship

- (up to) 2 PetaByte per day of Science Data Products

*all numbers subject to change (Totals for both SKA-Low and SKA-Mid SDP)

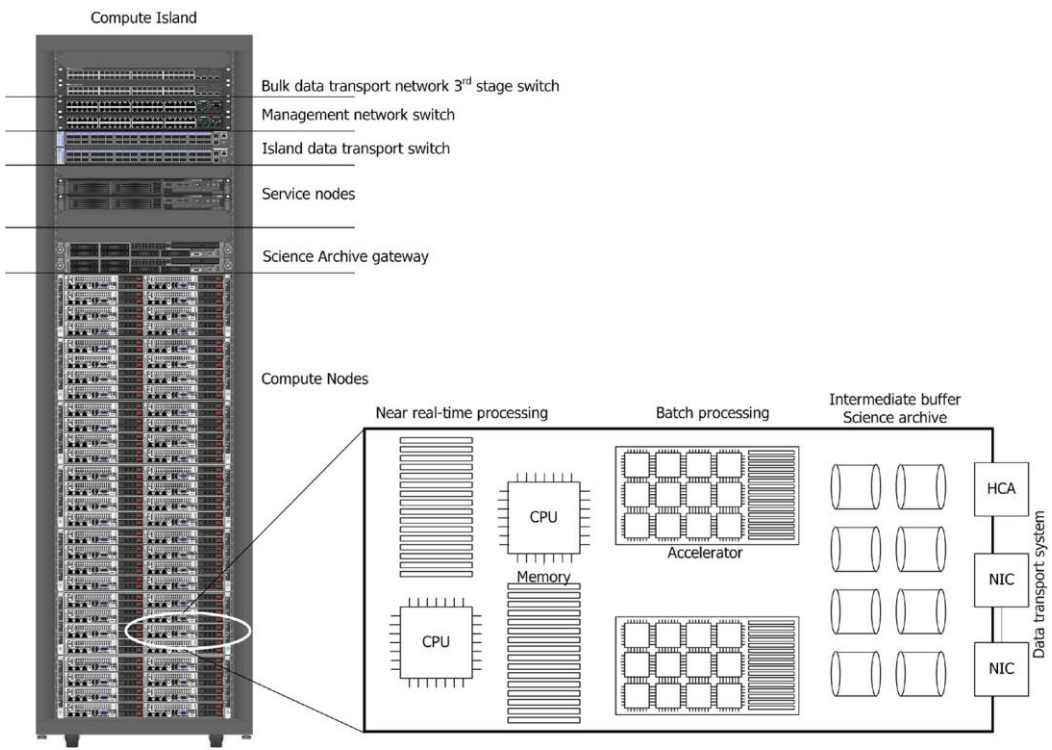
Current SDP Hardware Concept

See S. Ord's talk
Track 9 Exascale Science
Thu 14:00

Envisage a COTS based approach - an evolution of today's HPC and Cloud platforms.

Staged roll out to minimise risk

Full SKA1 SDP purchase not until 2024+



SDP Progress!

Summit showcase.

See A. Wicenec's talk
Track 3 Middleware and Distributed Computing
Tue 11:30

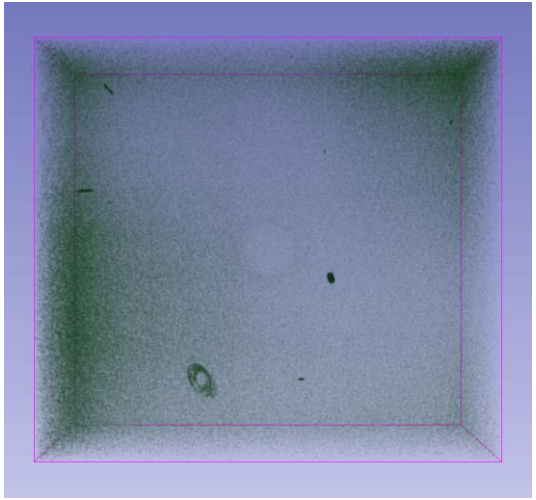
- Simulated SKA-Low telescope data
- Distributed to GPUs for processing into image cube (with DALiuGE graph engine)
- Image cube sent to Perth

Summit — Oak Ridge National Laboratory's 200 petaflop supercomputer. Number 1 in Top 500 list. Credit: Oak Ridge National Laboratory.

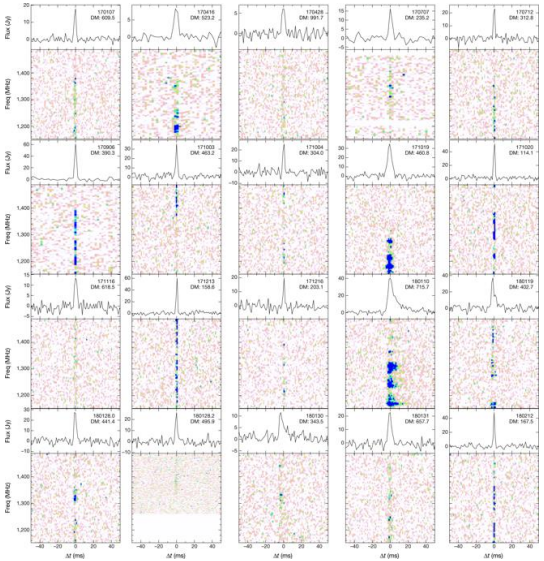
SKA SDP Data Products

Image cubes up to 1 PB

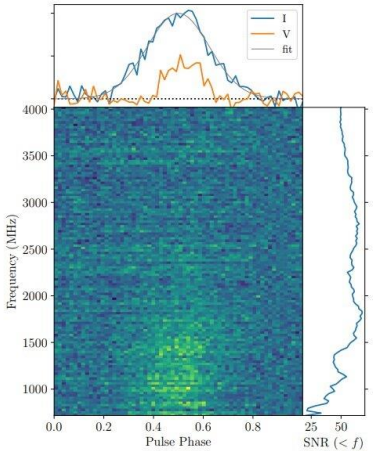
Example HI cube: ASKAP cube of Eridanus

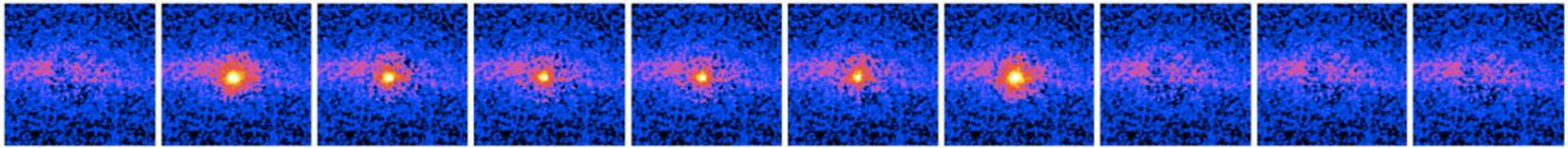


ASKAP fast radio bursts (Shannon et al. Nature 2018)

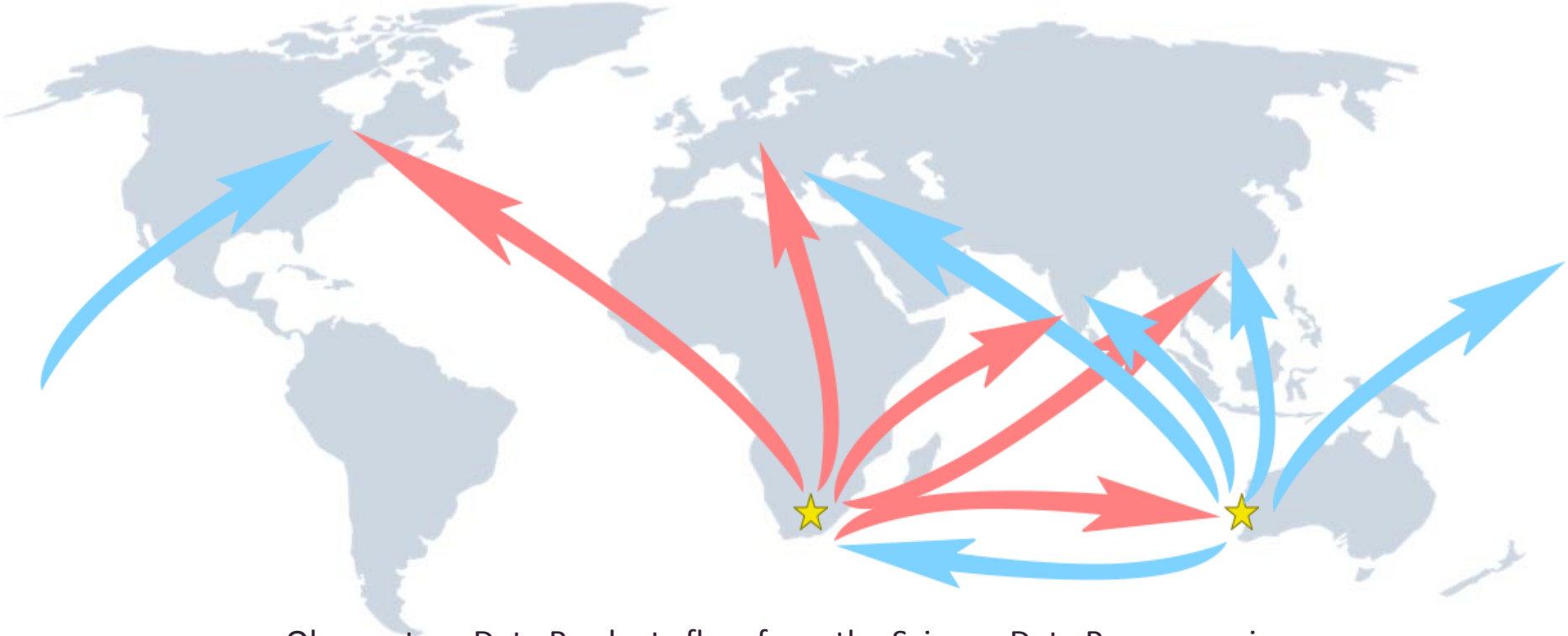


Folded pulse profile of PSR J1431−6328 (Kaplan et al. 2019)





Delivery to users with SKA Regional Centres



Observatory Data Products flow from the Science Data Processors in Perth and Cape Town to SKA Regional Centres around the globe

SKA Regional Centres

2016: SKA Board identified a shortfall between the construction cost of SKA-1 and the cost of supporting:

- Creation of advanced data products by the community,
- User support of Key Survey Projects teams as well as smaller PI projects,
- Creation and curation of a long-term SKA science archive capable of supporting archival research

Recommendation:

- a collaborative network of SKA Regional Centres (SRCs) to provide the essential functions that are not presently provided within the scope of the SKA1 project

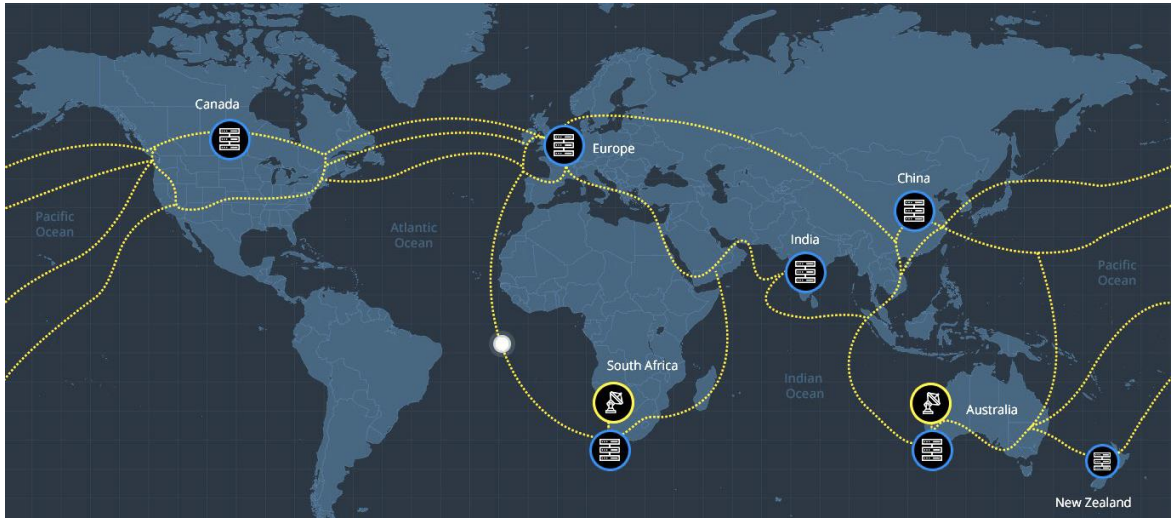
SRC Essential Functions

- **Data Flow** - delivering data from the observatory to those that have been given time to acquire it
- **Data Processing** - the resources necessary to work on the data after it is delivered
- **Data Curation** - providing a performant and persistent science archive that allows discovery and new science
- **User Support** - supporting all users with all of the things above
- **Commonality** - support a common and minimum set of tools to enable users to work at SRCs
- **Resource Management** - enable and support an interface to observatory time-allocation and operations processes to ensure maximal use of distributed SRC resources

SKA Regional Centres

- Essential for delivery of science
- Not part of SKA Observatory
  - Separately funded by individual nations
- Hope for something “like” the WLCG model
  - Transparent and location-agnostic for users
  - Resources pledged into system
  - Access given by user-linked data privileges
  - Accounting to track resource use
  - Tools developed for WLCG maybe applicable to SRCs, and lessons learnt?!
  - BUT also major differences – minimise moving of data, so Tiered model is not exactly correct

Global SRC Network



Work in progress!

- Where will the SKA science data archive be hosted?
- How will that data be transported from the sites to the SRCs?
- How can we take optimal advantage of existing infrastructure?
- What are the processing requirements and technologies to consider?
- What interfaces, tools, and techniques will users need for analysis?
- How do we setup and operate an international network of SRCs?

Global SRC Activities

Europe 2017-2023: AENEAS and ESCAPE H2020 programs, 6M€

Australia 2017-2022: ERIDANUS and AusSRC, 3M€

China 2017-2021: ERIDANUS and ChinaSRC, 5M€

South Africa 2017-2022: IDIA (Inter-University Institute for Data Intensive Astronomy), 0.7 M€

Canada 2018+: CIRADA, the Canadian Initiative for Radio Astronomy Data Analysis (Uni Toronto, 9 university partners, CADC), 7 M€

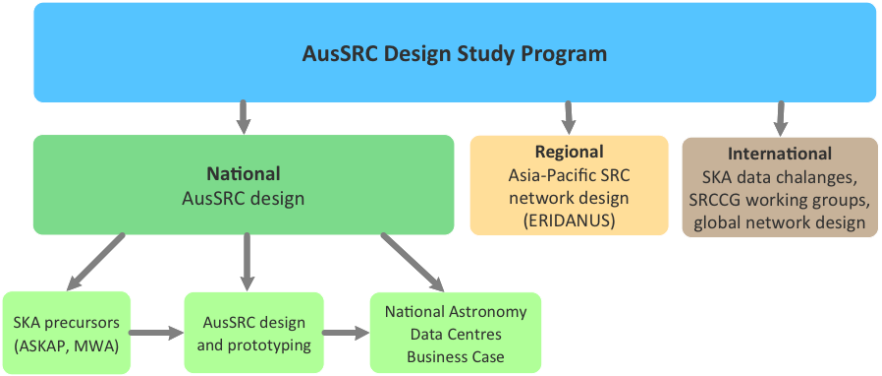
India NCRA, 6M€ requested

Total: ~28 Million Euros 2017 -2023.

Australian SKA Regional Centre

AusSRC Design Study Program:

- $4M AUD, 2019-2022
- Partners: ICRAR (UWA/Curtin), CSIRO, Pawsey
- Deliver projects for users of precursor telescopes (ASKAP and MWA): 2019 – 2022
- Define architectures and technologies relevant to an AusSRC at national and international level : 2019 – 2022
- Deliver a business case to the Federal Government for an investment in a staged deployment of an AusSRC 2023+

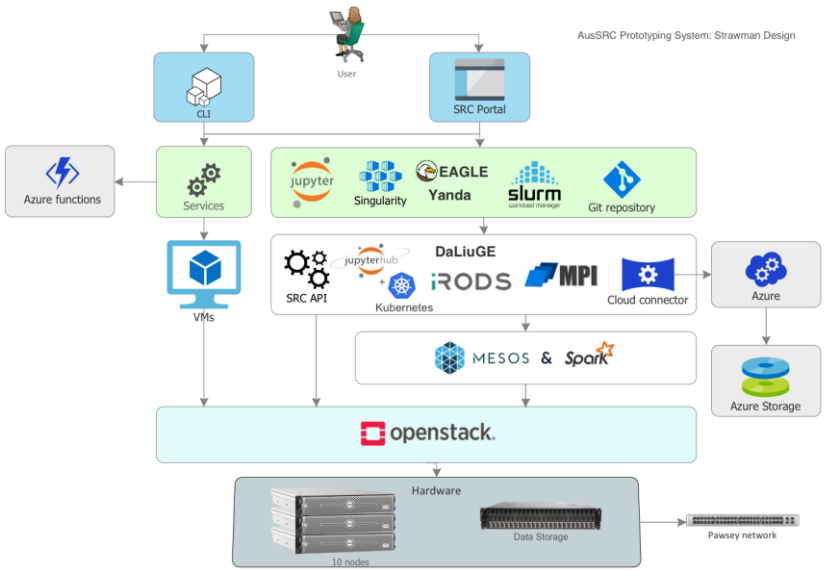


AUS SRC recent and upcoming activities

ERIDANUS
Exascale Research Infrastructure For Data In Asia-Pacific Astronomy Using The SKA

- Aus/China SKA Big Data Workshops 2017 and 2018 in Shanghai
- National community discussion -> Requirements White paper 2017
- Industry briefing events in Sydney and Perth in 2018 – more than 100 companies represented

- OzSKA and AusSRC workshops in Canberra 6-8 Nov 2019
- SKA Precursor telescope (ASKAP and MWA) support projects chosen, work to begin ASAP
- Prototyping underway – Strawman Architecture



Summary

- The SKA will be the largest radio observatory in the world
  - Two radio interferometers, in Western Australia and South Africa
  - Commissioning data in ~2024, key science to start ~2028
  - Broad science drivers, 50 year lifespan

- Data processing within the observatory ~250 PetaFLOPS: challenging power, memory bandwidth, computational efficiency.

- SKA Regional Centres are essential for delivering the science
  - Users will access, reprocess and analyse data through SRC
  - Development of SRC models, inc. AusSRC, and prototyping underway

Thank you

CSIRO Astronomy and Space Science
Minh Huynh
Senior Data Scientist and Research Astronomer

ATNF Science Team Leader

CSIRO ASKAP Science Data Archive (CASDA) Lead

+61 8 6436 8696
Minh.Huynh@csiro.au