

SKAO IRIS Activities

Rohini Joshi

r.joshi@skatelescope.org

SKA1 MID - the SKA's mid-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.



Location:
South Africa



Frequency range:
350 MHz to
14 GHz



~200 dishes
(including 64 MeerKAT dishes)



Total
collecting
area:

33,000m²



or
**126
tennis
courts**



Maximum distance
between dishes:
150km



Total raw data output:

2 terabytes
per second

62 exabytes
per year



x340,000



Enough
to fill
340,000
average laptops with
content **every day**

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



4x
the
resolution

5x
more
sensitive

60x
the survey
speed

Dish Array

SKA1 MID

SKA1 LOW - the SKA's low-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.



Location: Australia



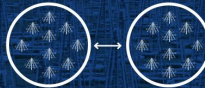
Frequency range:

50 MHz to
350 MHz

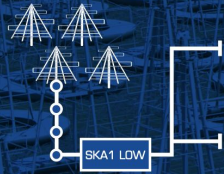


~130,000
antennas spread between
500 stations

Total collecting area:
0.4km²



Maximum distance between stations:
>65km



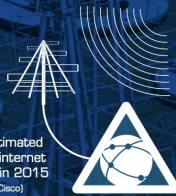
Total raw data output:

157 terabytes
per second
4.9 zettabytes
per year



Enough to fill up
35,000 DVDs
every second

the estimated
global internet
traffic in 2015
(source: Cisco)



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



25%
better
resolution

8x
more
sensitive

135x
the survey
speed

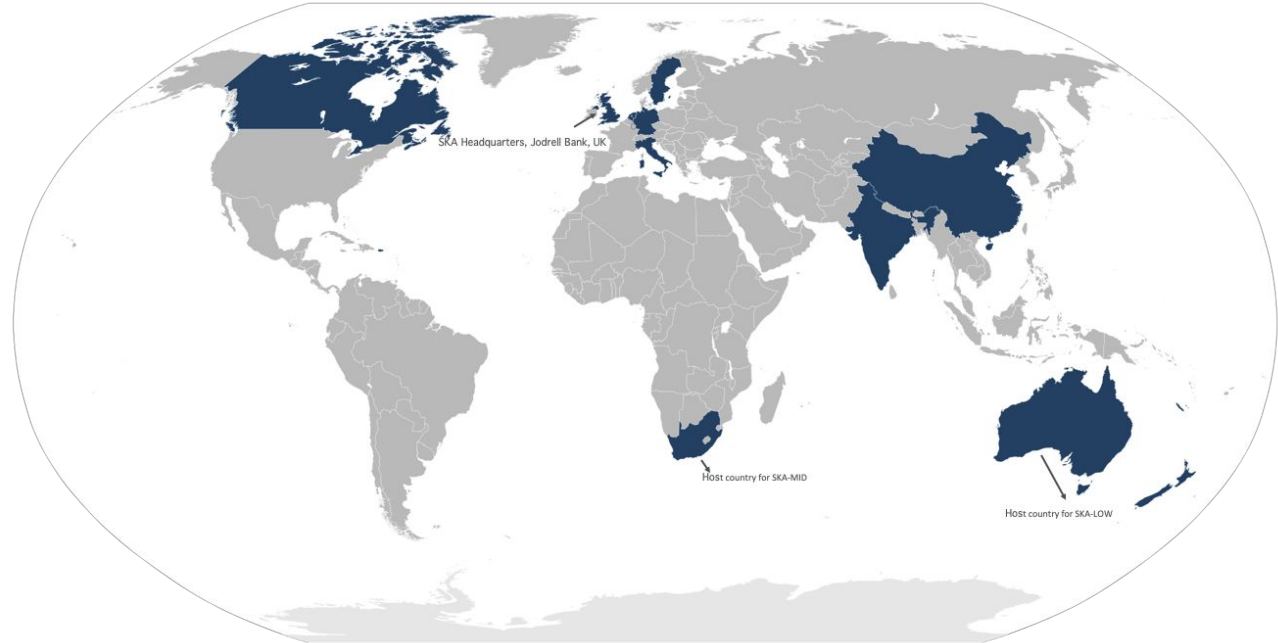
Aperture Array

SKA1 LOW

SKA Regional Centres

Science Data Processors at both the SKA sites will collect, process and churn out science data products that will be pushed out to one or more regional centres around the world

SKA Regional centres will provide a platform for data access, data distribution, post-processing, archival storage, and software development.



Our Team

SKA operates a scaled agile software development framework (SAFe), we are a part of the Data Processing Agile Release Train

ESCAPEES team: James Collinson, Rob Barnsley, Alex Clarke, Rosie Bolton

SRC related methods and technologies that aligns with responsibilities to ESCAPE



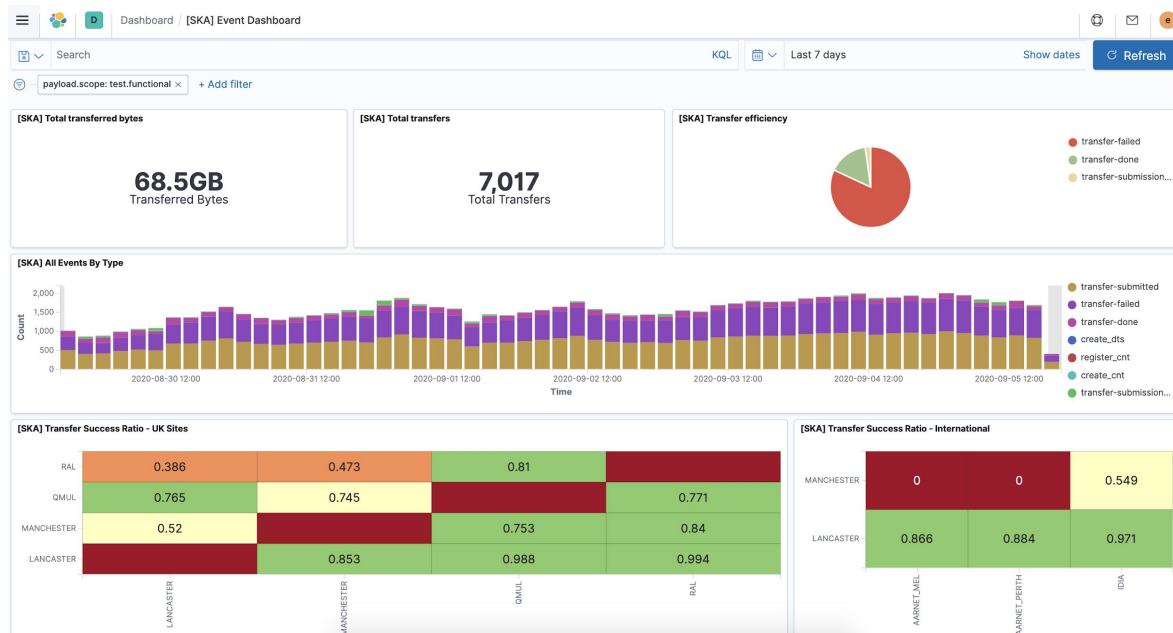
Continued updates

Rucio instance

- multi-VO upgrade
- Kibana dashboard
- Test playground
- Australian RSEs

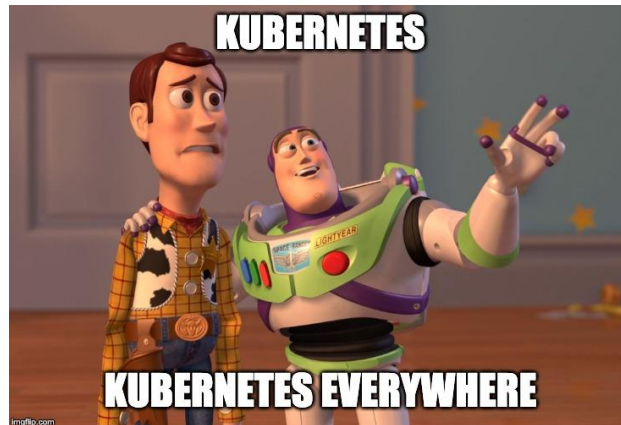
Grid workflows, high memory node utilisation and MeerKAT pipeline work continued by Anna Scaife's group

STFC Cloud usage ...



STFC Cloud usage 1/3

- Kubernetes!
- Kubernetes for hosting JupyterHub IRIS/ESCAPE
<http://130.246.212.44/iris>
<http://130.246.212.44/escape>
Deployment documented [here](#)
- For the most part, we don't know what we're doing!
Happy to speak with other communities deploying similar things. User liaison meetings are great!
- Built an astronomy specific JupyterHub environments ie, RASCIL.
- Looking forward to LBaaS, and Cinder volumes



Img Credit

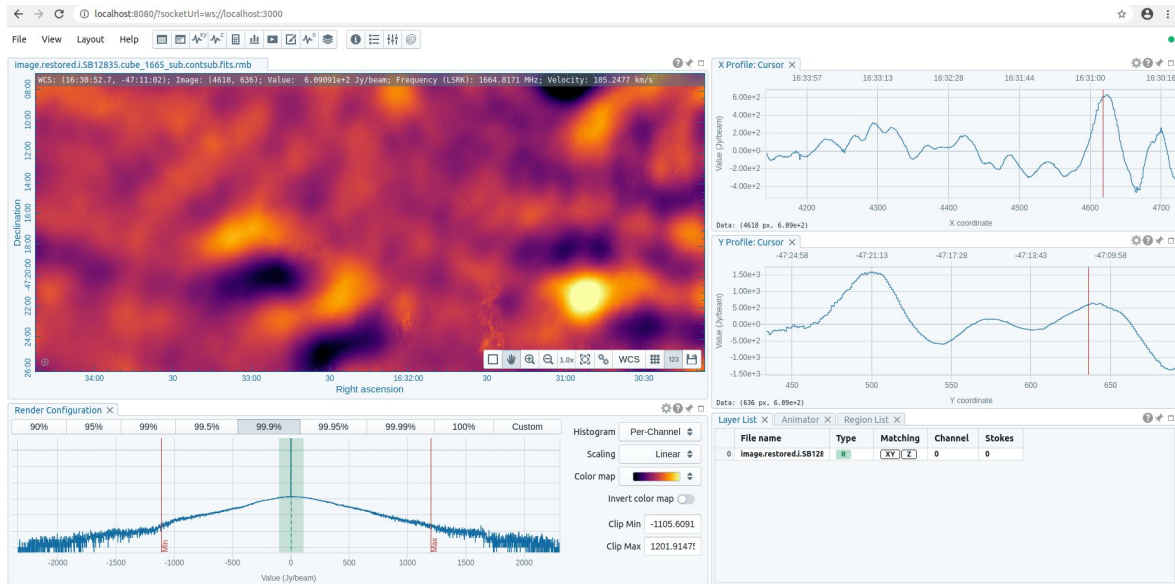
STFC Cloud Usage 2/3

SKA Science machine

- Similar specs to the in-house science server hosted at SKA
- Made available to the Science Team (via ssh pub key)

ASKAP analysis

- Visualising ASKAP data to determine suitability
- 'Desktop' installation of the CARTA viewer vx.x
<https://carta.readthedocs.io/en/latest/>



Pilot field of the 1665 MHz OH line taken for the GASKAP survey
Image Credit: Shari Breen, Rob Barnsley

SDC1: Continuum emission

- Datasets of continuum emission images, simulating observations for SKA MID Bands 1, 2 and 5
- 32k x 32k Images populated by star forming galaxies (SFGs) and active galactic nuclei (AGN)
- Challenge description: To find and characterise sources
- Results now submitted for publication in MNRAS
- Data hosted on ownCloud. Submissions, support, troubleshooting handled over email
- Challenge was live Nov 2018 - May 2019
- <https://astronomers.skatelescope.org/ska-science-data-challenge-1/>

MNRAS 000, 000–000 (2020)

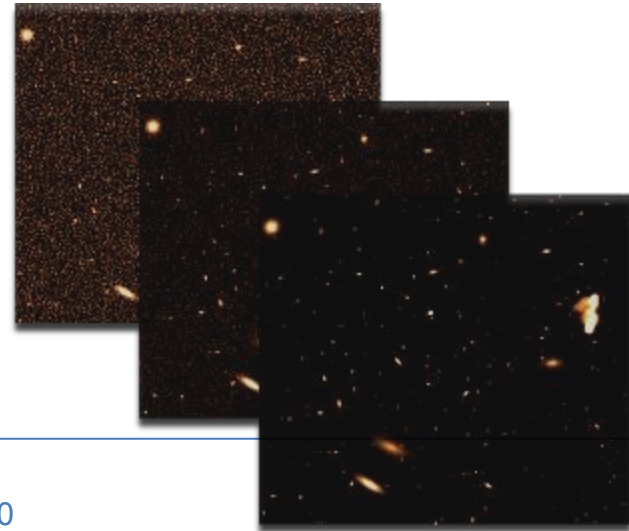
Preprint 30 May 2020

Compiled using MNRAS L^AT_EX style file v3.0

Square Kilometre Array Science Data Challenge 1: analysis and results

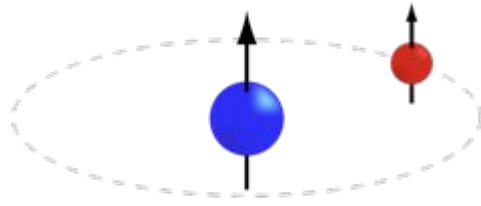
A. Bonaldi^{1,2*}, T. An³, M. Brüggen⁴, S. Burkutean⁵, B. Coelho⁶, H. Goodarzi⁷, P. Hartley¹, P. K. Sandhu⁸, C. Wu⁹, L. Yu¹⁰, M. H. Zhoolideh Haghighi⁷, S. Antón^{11,6}, Z. Bagheri^{7,12}, D. Barbosa⁶, J. P. Barraca^{6,13}, D. Bartashevich⁶, M. Bergano⁶, M. Bonato⁵, J. Brand⁵, F. de Gasperin⁴, A. Giannetti⁵, R. Dodson⁹, P. Jain⁸, S. Jaiswal³, B. Lao³, B. Liu¹⁰, E. Liuzzo⁵, Y. Lu³, V. Lukic⁴, D. Maia¹⁴, N. Marchili⁵, M. Massardi⁵, P. Mohan³, J. B. Morgado¹⁴, M. Panwar⁸, Prabhakar⁸, V. A. R. M. Ribeiro^{6,15}, K. L. J. Rygl⁵, V. Sabz Ali⁷, E. Saremi⁷, E. Schisano¹⁶, S. Sheikhezami^{17,7}, A. Vafaei Sadr¹⁸, A. Wong¹⁹, O. I. Wong^{9,21,20}

Affiliations are at the end of the paper

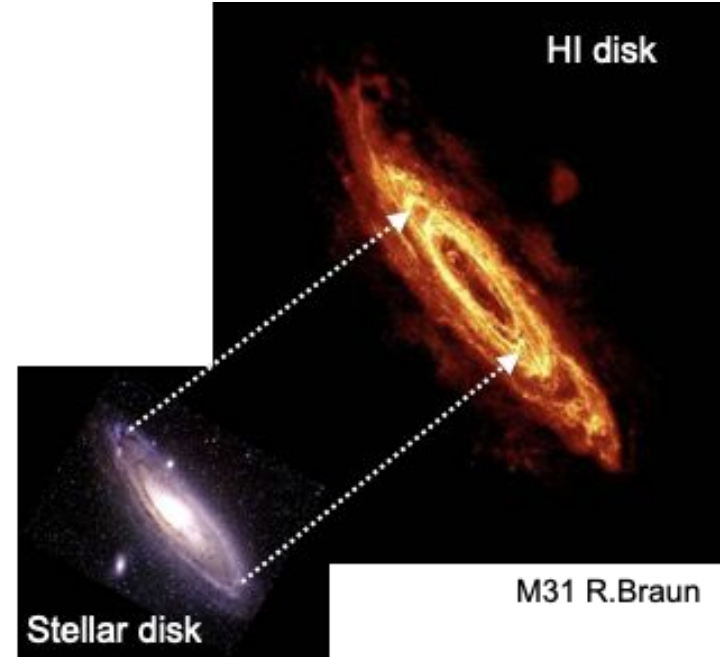


SDC2: Neutral hydrogen (HI)

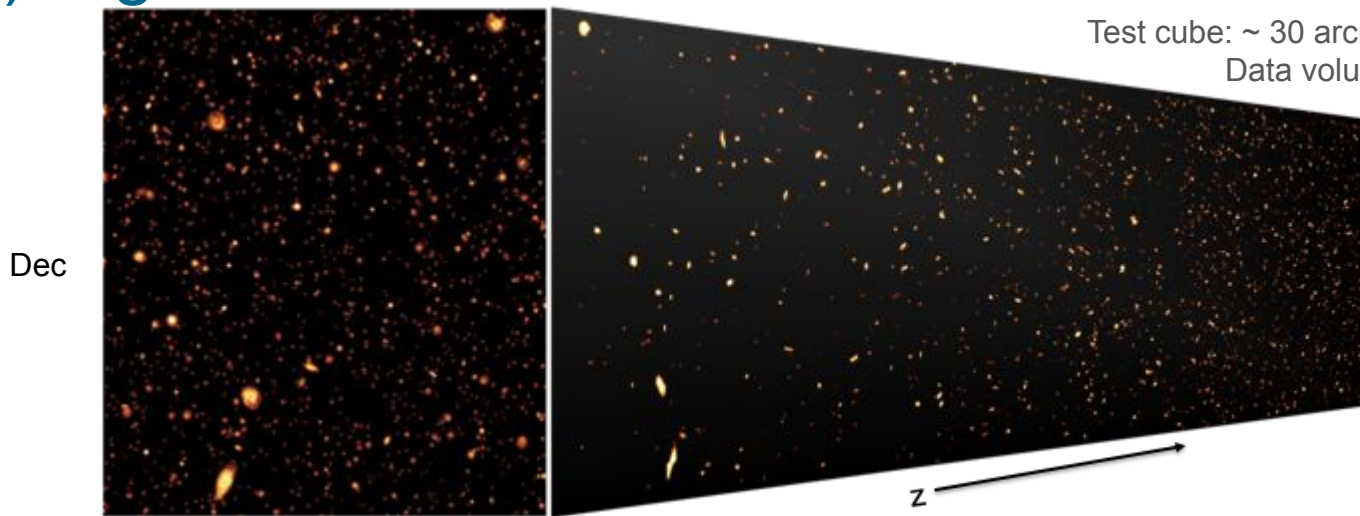
- Neutral hydrogen (HI) found in star forming galaxies and in tidal interactions in the intergalactic medium
- HI emits radiation during a 'spin flip' of its single electron



- HI spectral line at $\nu_0 = 1.42$ GHz ($\lambda_0=21$ cm) redshifted to lower frequencies
- SKA will trace HI from the current time, all the way back to the formation of the first galaxies during 'Cosmic Dawn'



(Fairly) Big data



Test cube: ~ 30 arcmin across
Data volume: 16 GB

RA

Integration time = 2000h
Spatial resolution = 5 arcsec
Frequency resolution = 115 kHz
RMS per channel 13-18 μ Jy

FoV = 20 square degrees

Frequency = 950 MHz – 1350 MHz ($z = 0$ to 0.5)

Data volume = 1TB

SDC2 access and processing

- The primary challenge is source finding and characterisation
- Webpage for overview of the challenge, resources available, submission templates, discussion forum, facilitate internal and external collaboration <http://sdc2.astronomers.skatelescope.org>, EoI
- We will be arranging access to compute facilities for participants to solve the data challenge
- Preparing the scientific community to deal not only with SKA-like data, but also future SKA/SRC practices
- Exploring the functionalities of current proto-SRCs

STFC Cloud Usage 3/3

Solving SDC1 - classifying sources in radio images

- SDC1 solution will likely form the basis for the more advanced pipeline required for SDC2 and will help test SDC2 platform hosts
- Demonstrating best-practice with respect to making our solution open, portable, and efficient.
- <https://gitlab.com/ska-telescope/sdc/sdc1-solution>

SDC2 simulation testing

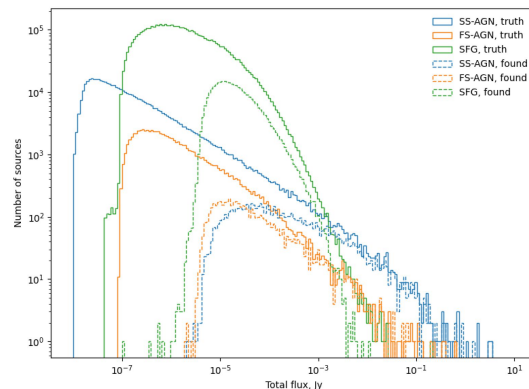
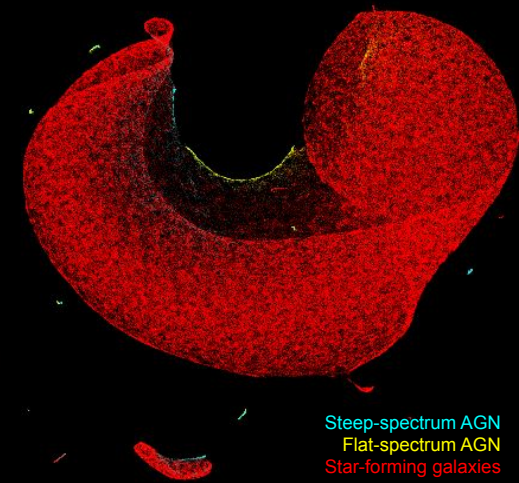
- Building and testing the portability of the SDC2 simulation pipeline
- Given the size of the input data cube, will be best to create the cube at site rather than transfer

Applied a Random Forest algorithm to classify sources based on features output by a source finder.

Visualised the results in a 2D space using dimension reduction algorithm called UMAP.

Compare results to the truth catalogue to assess performance

Visualisation of our classification scheme in a 2D feature space



Finding AGN is hard because they are the minority class

SDC2 host platform requirements

- We expect ~30 multi-user teams with the challenge live for 6 months - 1 year
- Each team will likely require RAM-heavy interactive resources: at least 8 cores, 60-90 GB RAM, 100 GB+ disk with shared access to the input cube.
- Interested facilities include the Shanghai Astronomical Observatory (SHAO) HPC, EngageSKA (Portugal), INAF ICT, SKA France
- Working to establish format of IRIS resources offered and considering the operational responsibility of the same
- Looking into resource management and how to ensure fair consumption of resources among teams

In progress...

Next steps

SDC2

- Solidifying platform requirements, user models, timelines for SDC2
- Initiating conversations with STFC Cloud
- SDCSS
- Accounting dashboard consolidation

Rucio

- Deploy an ElasticSearch/Influx DB database to populate ESCAPE datalake monitoring dashboard with Rucio events
- Custom environment to interact with data in the ESCAPE data lake
- Rucio real-data transfers from IDIA to Manchester using the multi-VO instance

Provision IRIS resource for SDC

Collinson, James ESCAPEES

Identified SPO-792

Create a dashboard capable of displaying Rucio events

Collinson, James ESCAPEES

Identified SPO-784

Reproduce and develop a JupyterHub custom environment for Rucio

Collinson, James ESCAPEES

Identified SPO-786

Develop JupyterHub prototype for SDC use case

Collinson, James ESCAPEES

Identified SPO-787

Thank you!