MeerKAT Large Survey Projects (~70%)

imaging

- **LADUMA (Deep atomic hydrogen)**
- **MIGHTEE (Deep continuum imaging of the early universe)**
- **Fornax (Deep HI Survey of the Fornax cluster)**
- **MHONGOOSE (targeted nearby galaxies HI)**
- **MeerKAT Absorption Line Survey (extragalactic HI absorption)**

Time domain

- **ThunderKAT (exotic phenomena, variables and transients)**
- TRAPUM (pulsar search)
- MeerTime (pulsar Timing)

SKA Key Science themes

Indo-SA
Flagship Programme in Astronomy

Joint Exploration of the Deep Radio Sky with MeerKAT and the GMRT: The Pathway to the SKA

With Dharam Lal: Tata Institute for Fundamental Research, NCRA

Department of Science and Technology
Government of India

Final Workshop: June 2019

NRF National Research Foundation | Astronomy sub-Agency

SuperMIGHTEE

- Combined data increases sensitivity to ~1 microJy (0.5-2.7 GHz)
- Provides matched resolution (5”) at 0.5-0.8 GHz
- Ultra broad band spectra from 0.5 – 2.7 GHz for microJy radio sources
- HI science to higher redshift (and high resolution for stacking)
- Ultra-broad band polarimetry to measure Faraday complexity
  - 3.2 times the Faraday synthesis precision

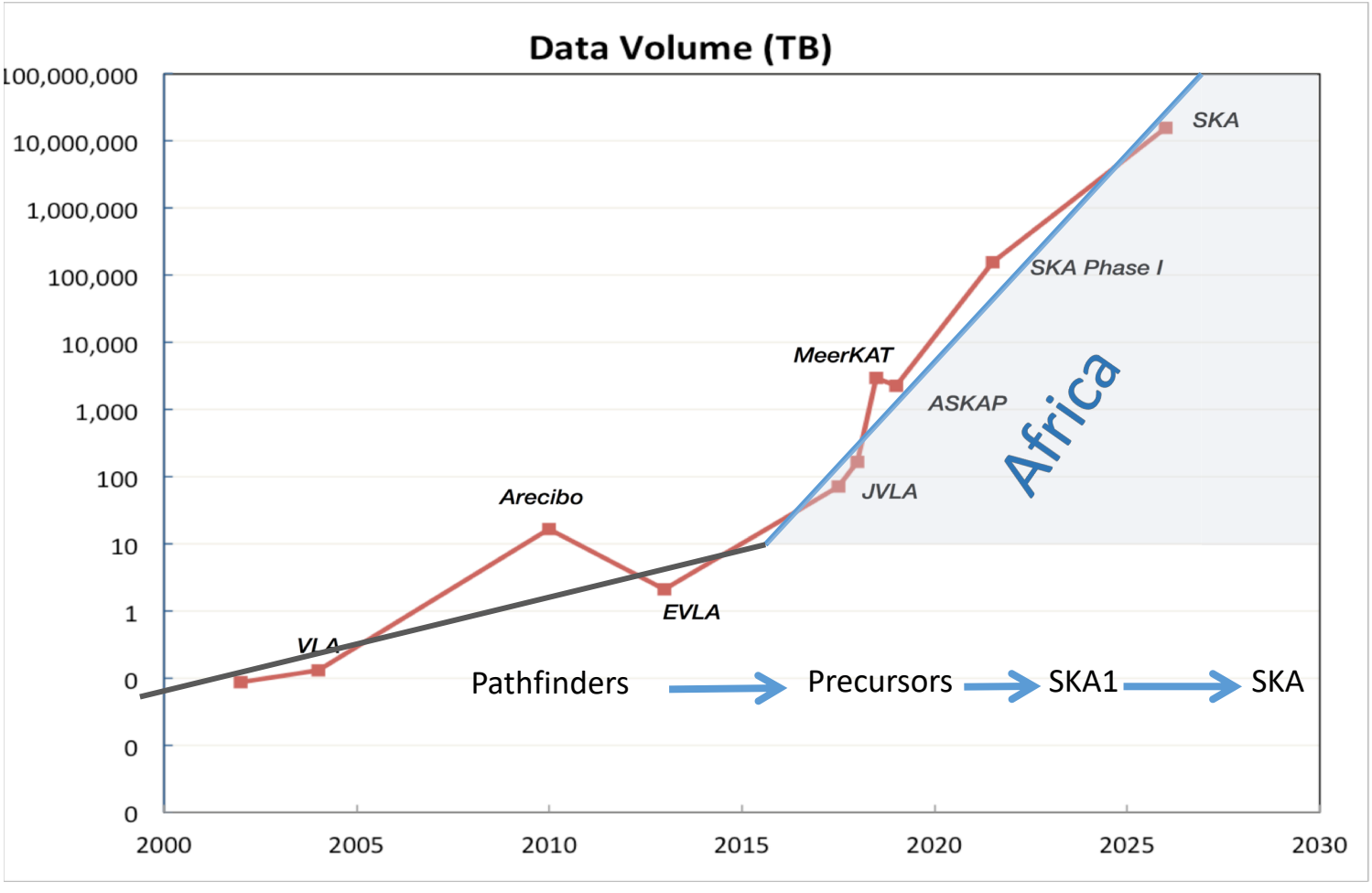
FAST-MeerKAT and Cosmic Magnetism in the SKA Era

Fast-MeerKAT and SKA Pathfinders Synergies

2018 Radio Astronomy Forum, Pingtang, China, June 2018

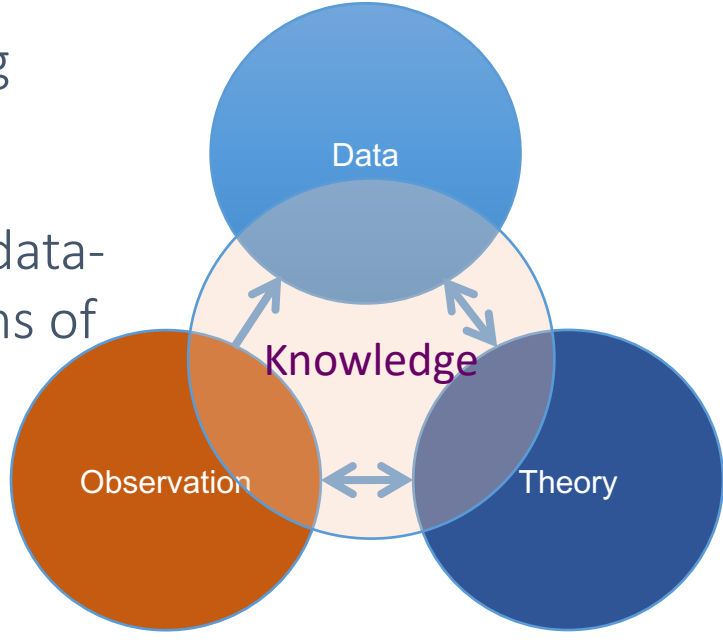
(see also proposal 13)

Growth of Data Volumes to Radio Astronomers



The Challenge: data to knowledge

- Managing exponential increase in rates and volumes. (**network, transport, data-centric architectures**)
- Empowering the end user for multi-purpose processing, analytics, data mining and exploration **(cloud, HPC)**
- Fusion with big multi-wavelength data and big simulations (**cloud, federated data**)
- Collaboration, sharing and joint execution of data-intensive projects by globally distributed teams of researchers (**cloud, e-science platforms**)



Proposal to BRICS STI Framework Programme
2nd Call 2017

BRICSKA:
Big **D**ata **R**esearch **I**nfrastructure **C**ollaboration toward the **SKA**

(Proposal 12 and 14 in package)

Project Leader:

South Africa: Russ Taylor (UWC,UCT) + 16 researchers @ 6 institutions
(UWC, UCT, Pretoria, Rhodes, Zululand, IAU/OAD)

Partners:

- China: Chenzhou Cui (Chinese Academy of Sciences, NAOC) + 14 researchers @ 6 institutions
- India: Yogesh Wadadekar (TIFR, National Centre for Radio Astrophysics) + 4 researchers @ 3 institutions
- Brazil: Luiz Gadelha (National Laboratory for Scientific Computing) + 4 researchers @ 2 institutions

BRICSKA:
Big **D**ata **R**esearch **I**nfrastructure **C**ollaboration toward the **SKA**

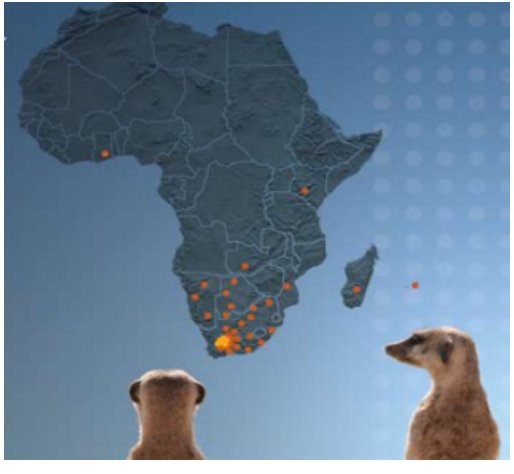
- ***Open source cloud implementation and technologies*,** including open source systems for provisioning of infrastructure, platforms, software and tools for big data processing and analytics, globally federated cloud environments, and systems for data transport and distribution and data fusion between federated cloud nodes.
- **D*evelopment of novel algorithms and HPC software for processing and analysis of big (radio) astronomy data*,** including
  - processing pipelines for raw data to science products, including development of workflows systems and research data management and reproducible science technologies.
  - Algorithms for data mining big data for information, including machine learning
  - cloud-based visual analytics of remote big data
  - science gateway and portal technologies for user access to the federated cloud environment and to cloud processing, analytics and visualisation tools.
- ***Building open source based architecture and tools to enable deployment of HPC systems*** by agile provisioning of HPC in the cloud to take advantage of parallelism, concurrency for massive radio astronomy data set processing.
- ***Human Capital Development*** through training programs and through development and delivery of data driven outreach and education projects and programs with cloud enabled platforms

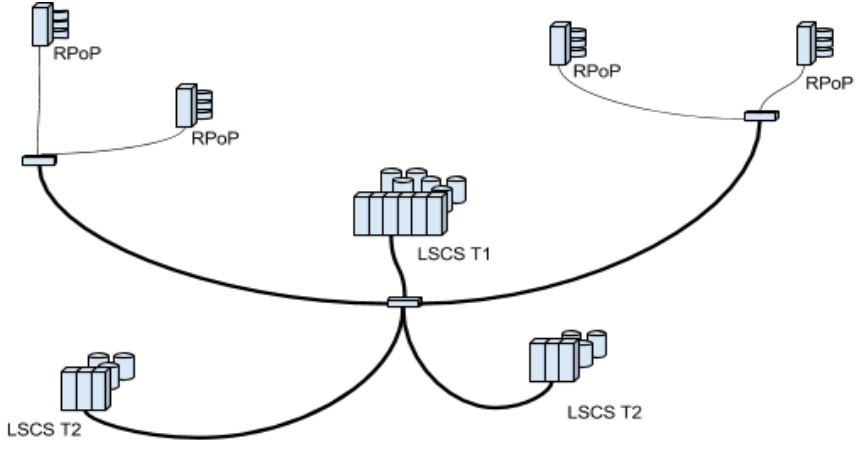
African Data Intensive Research Cloud

IST-Africa 2016 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2016
ISBN: 978-1-905824-54-0

The African Data Intensive Research Cloud

Rob SIMMONDS[1], Russ TAYLOR[2,3], Jasper HORRELL[4], Bernie FANAROFF[5], Happy SITHOLE[6], Sakkie JANSE VAN RENSBURG[7], Boeta PRETORIUS[8]





AstroCloud: Astronomy cloud of China-VO

AstroCloud: A Distributed Cloud Computing and Application Platform for Astronomy

Chen-Zhou Cui, Bo-Liang He, Chang-Hua Li, Dong-Wei Fan, Shan-Shan Li, Lin-Ying Mi, Zi-Huang Cao, Si-Si Yang, Yun-Fei Xu, Yue Chen, Zheng Li, Xu Han
Center for Information and Computing ,National Astronomical Observatories, CAS,
Beijing, China
Email: ccz@bao.ac.cn

Ce Yu, Jian Xiao, Zhi Hong, Shucheng Yin, Chen Li
School of Computer Science ,Tianjin University
Tianjin, China

Chuanjun Wang, Yufeng Fan, Jianguo Wang, Junyi Chen
Gaomeigu Observatory ,Yunnan Astronomical Observatories, CAS
Kunming, China

Hailong Zhang
Information Center ,Xinjiang Astronomical Observatory, CAS
Urumqi, China

Liang Liu, Na Gao, Zherui Yang
Information Center, Purple Mountain Observatory, CAS
Nanjing, China

Xiao Chen, Min Liu
Computing Center ,Shanghai Astronomical Observatory, CAS
Shanghai, China

Cuilan Qiao, Kangyu Du
College of Physical Science and Technology ,Center Normal University
Wuhan, China

Liying Su, Wenming Song
College of Mechanical Engineering and Applied Electronics Technology ,Beijing University of Technology
Beijing, China

Screenshot

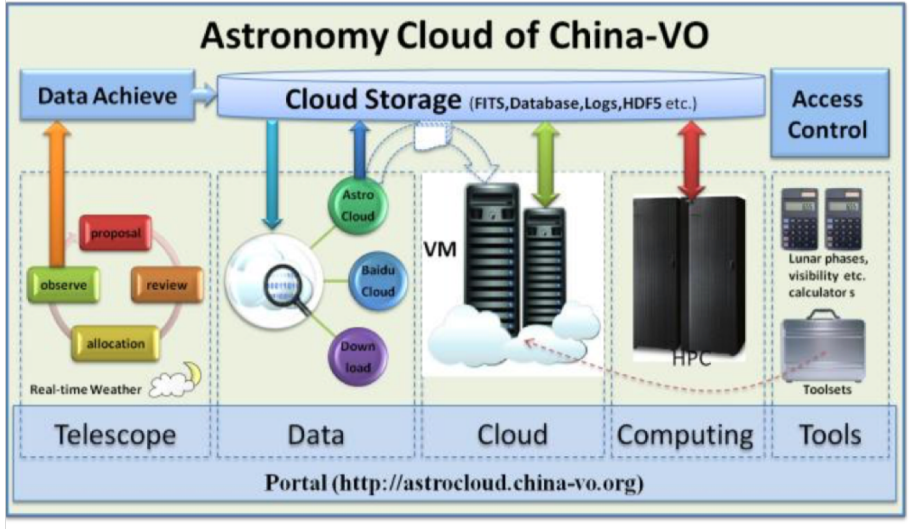
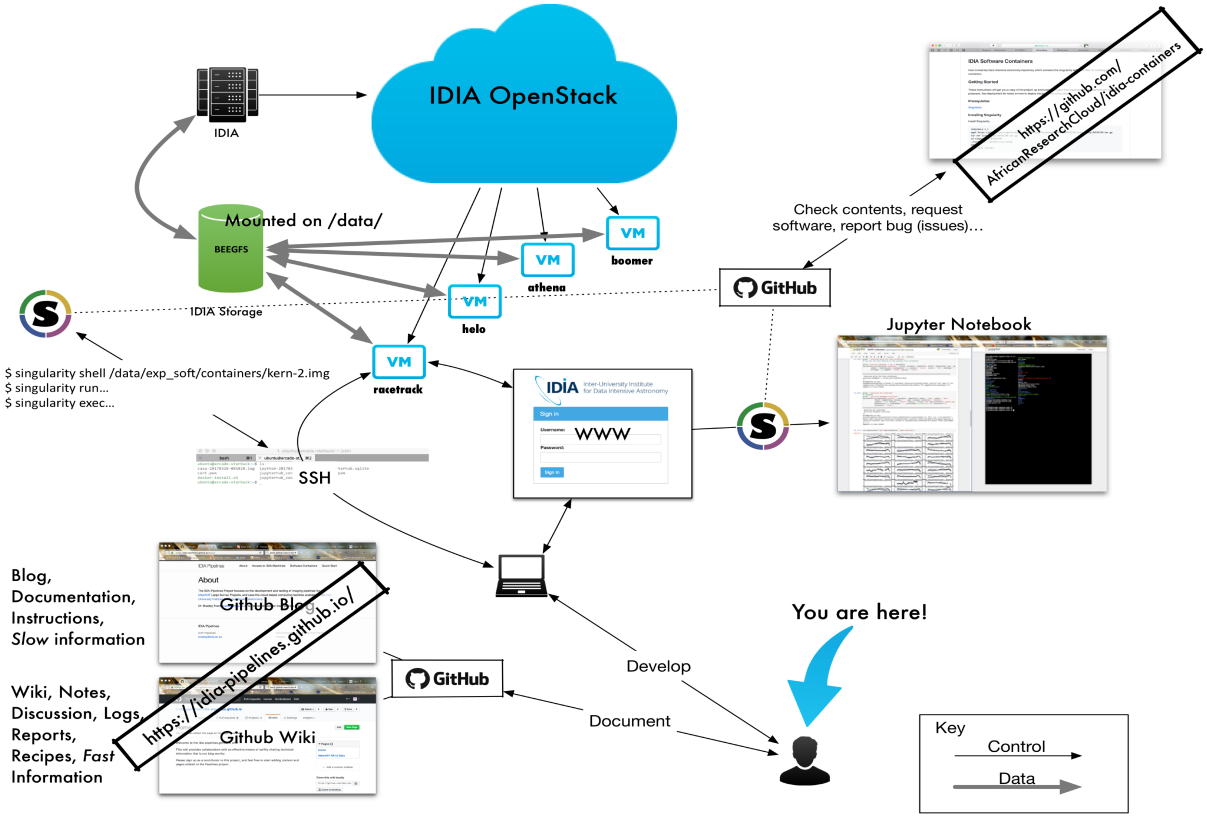
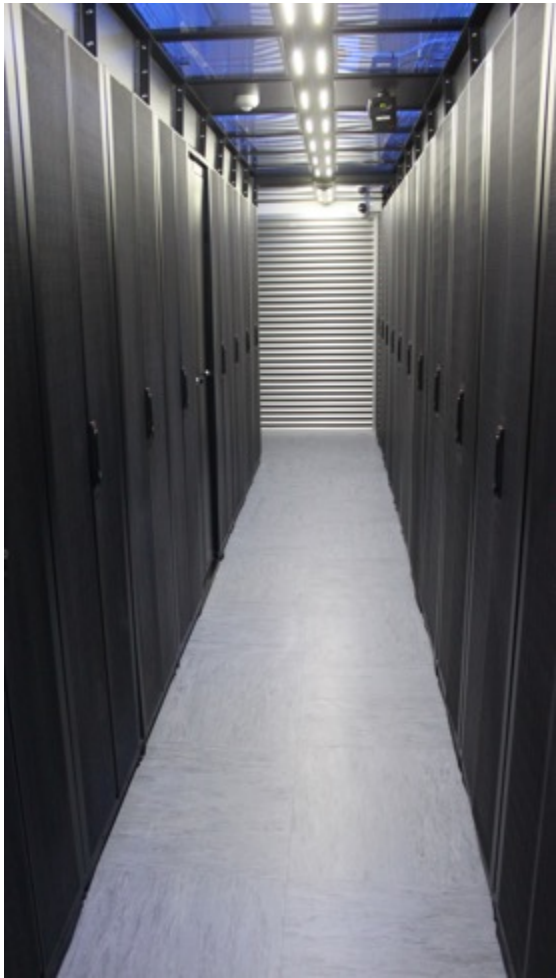


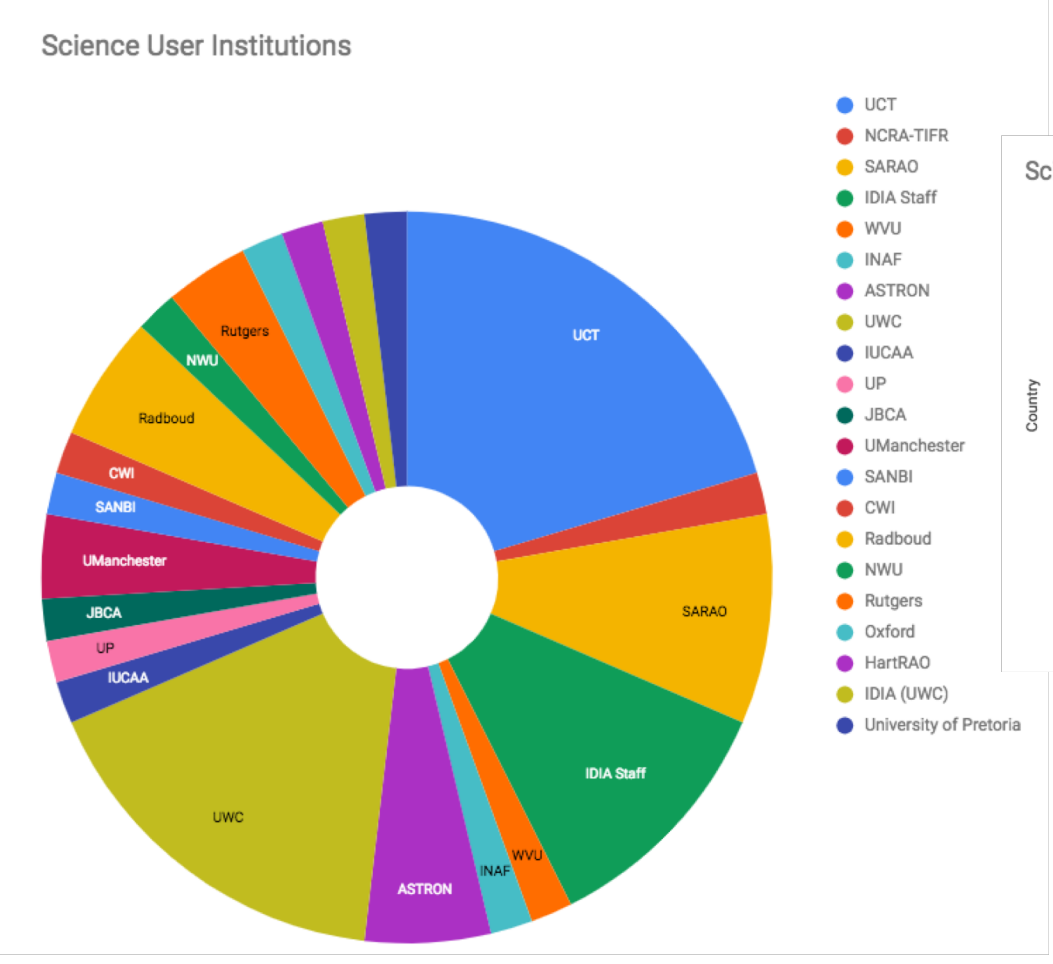
Figure 1. High level overview of the astro cloud platform

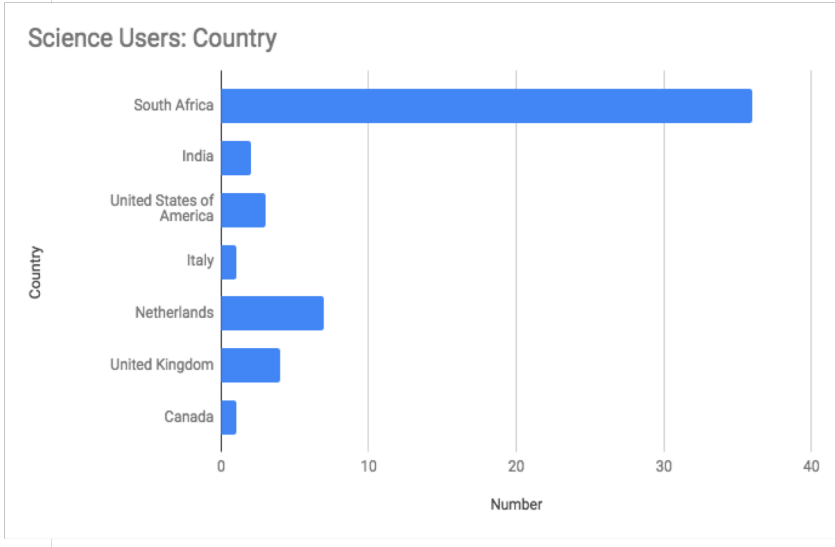
IDIA Data-Intensive Cloud: V1.0 Jan 2017





IDIA Cloud MeerKAT Science Users





ILIFU: Tier 2 Data Intensive Research Facility

Data Intensive Research Cloud V2.0

Joint investment DST/DIRISA, IDIA, Computational Biology (NIH)

- Astronomy (IDIA, SARAO)
  - Data Intensive Astronomy with priority on MeerKAT Large Survey Programs
  - Precursor SKA Regional Science Centre
- Data Intensive Bioinformatics
  - Tuberculosis Surveillance in Africa (UWC)
  - Imputation service for African human genetics (UCT)
  - Omics for Precision Medicine (SU)
- Research Data Management (CPUT)
- South African Data Intensive Research Cloud federation with T1 and T3 infrastructure

University of Cape Town · Universiteit van Kaapstad · iYunivesithi yaseKapa · Universiteit van Kaapstad

SOL PLAATJE UNIVERSITY

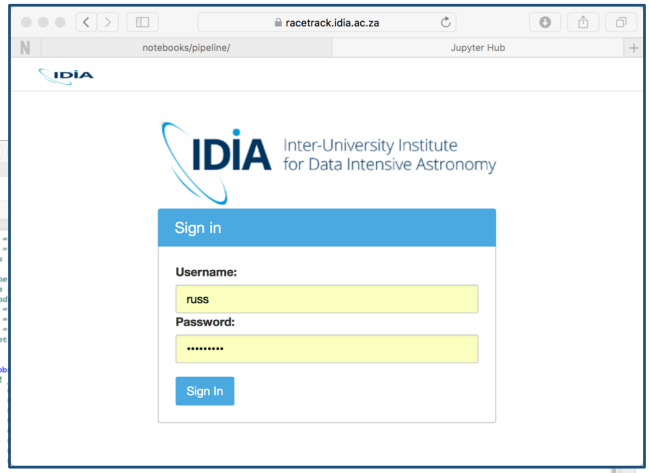
UNIVERSITY of the WESTERN CAPE

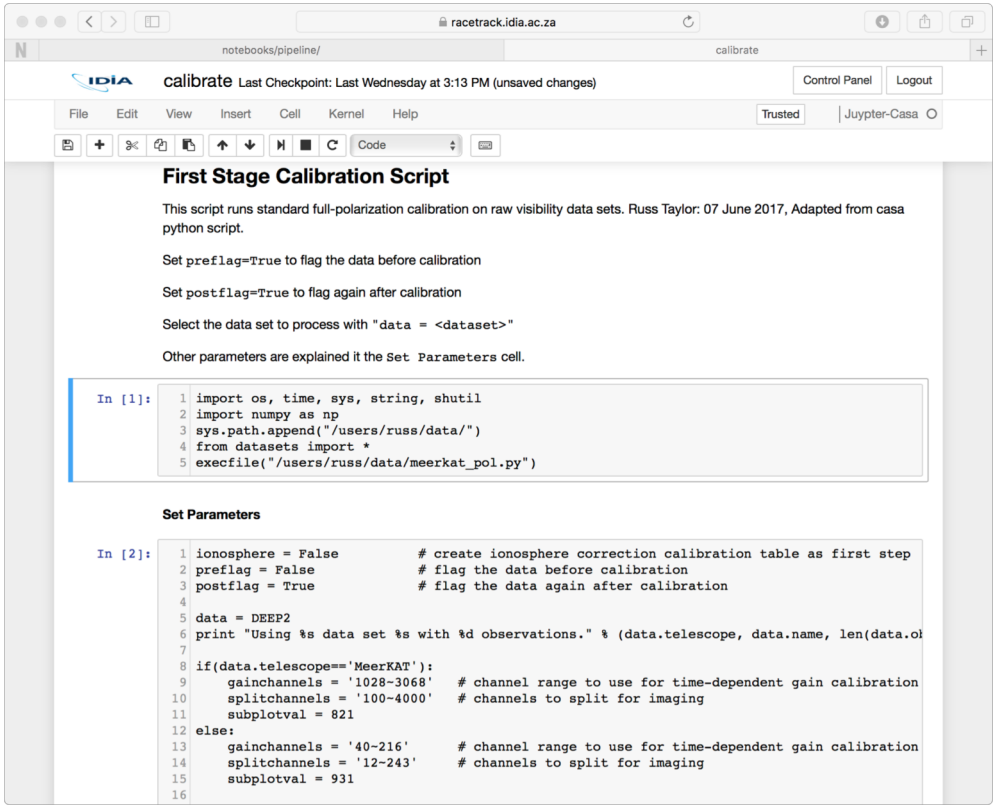
Cape Peninsula University of Technology

UNIVERSITEIT STELLENBOSCH UNIVERSITY

SKA AFRICA SQUARE KILOMETRE ARRAY

Data Intensive Astronomy Cloud

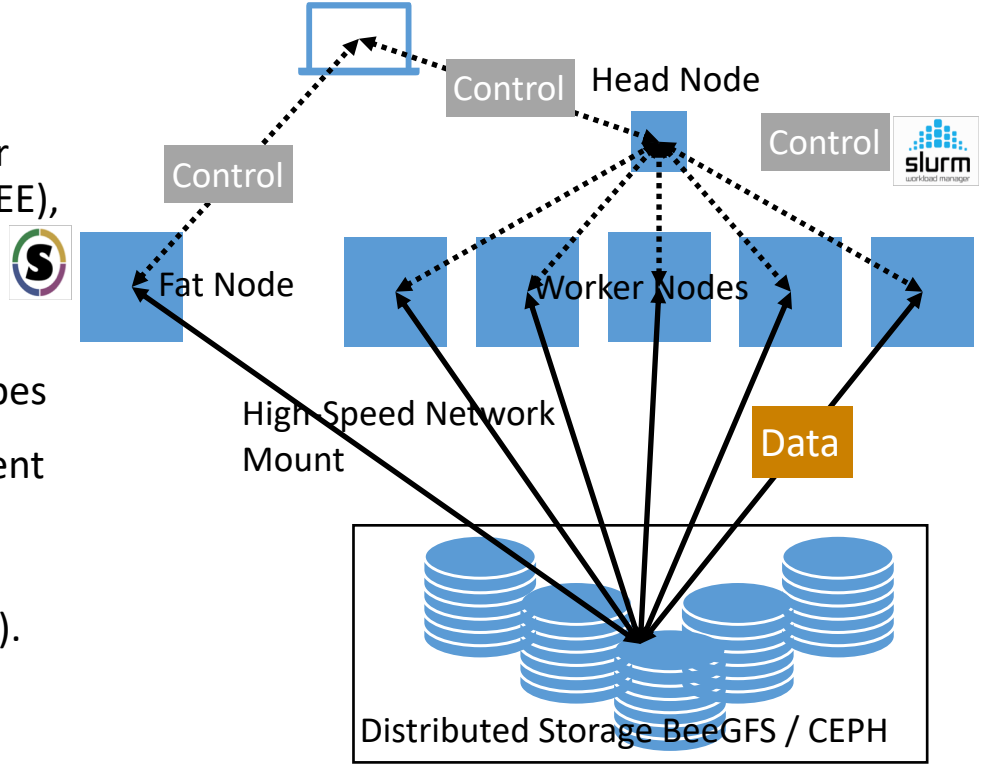


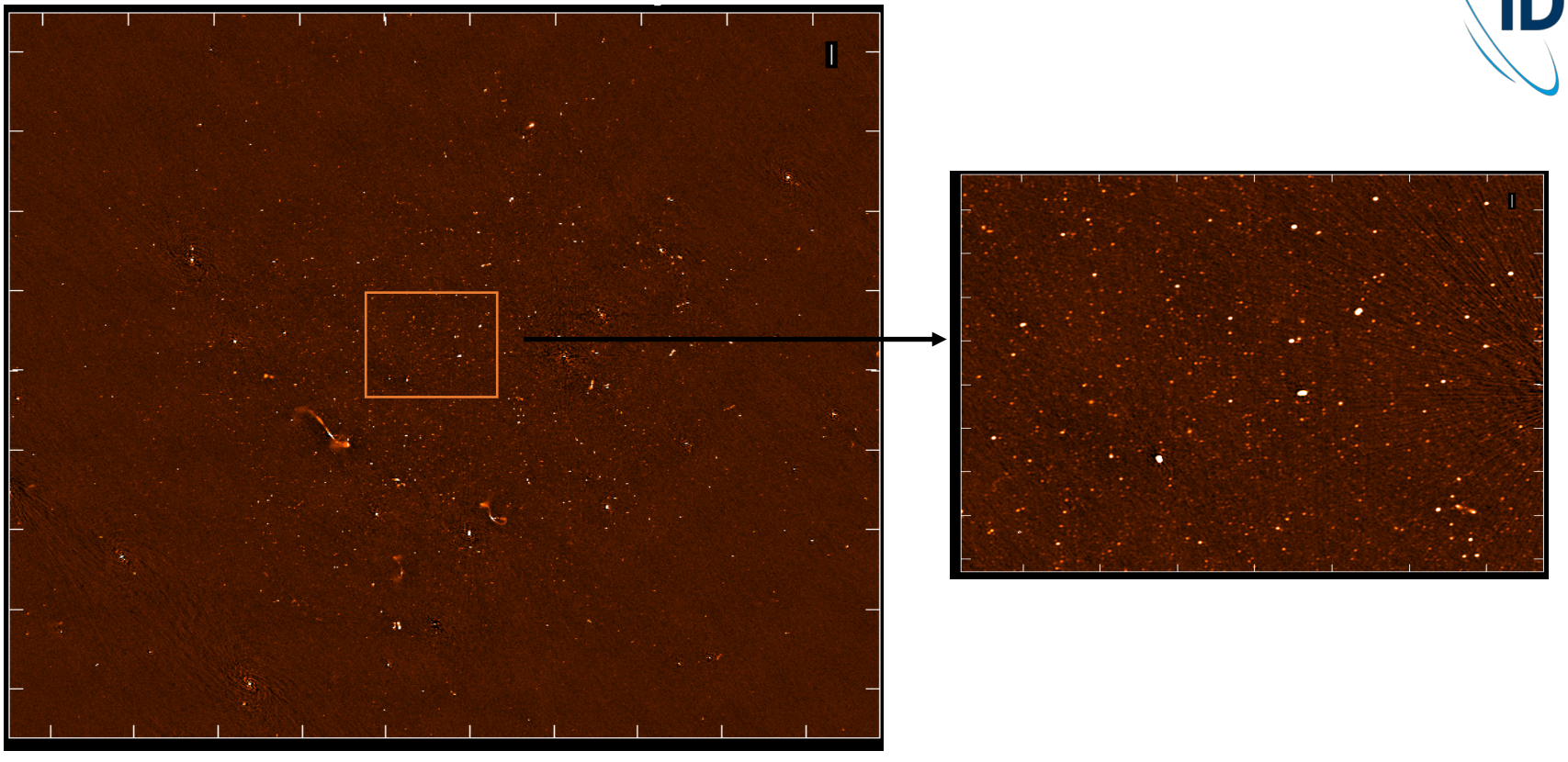


CASA in Jupyter Notebooks

IDIA Cloud Pipelines V1.0: Virtualized HPC

- `processMeerKAT`: VHPC implementation of CASA/MPICASA (SLURM).
- Core team: Bradley Frank, Srikrishna Sekhar, Jordan Collier (IDIA), plus developers.
- Involvement from MeerKAT LSPs: Russ Taylor (MIGHTEE), Ian Heywood (LADUMA, MIGHTEE), Andrew Baker & John Wu (LADUMA).
- GMRT and JVLA SKA pathfinder projects
- Full-Stokes calibration and imaging to 4D cubes
  - Parallelised (single-node/OMP) + Concurrent (Cluster/MPI).
  - 3~4x speedup (flagging included) over conventional CASA (without optimisation!).
  - Any user (group) will have access

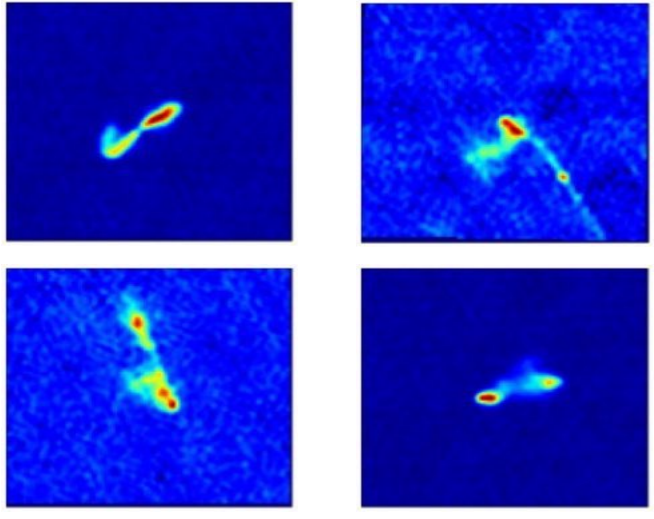




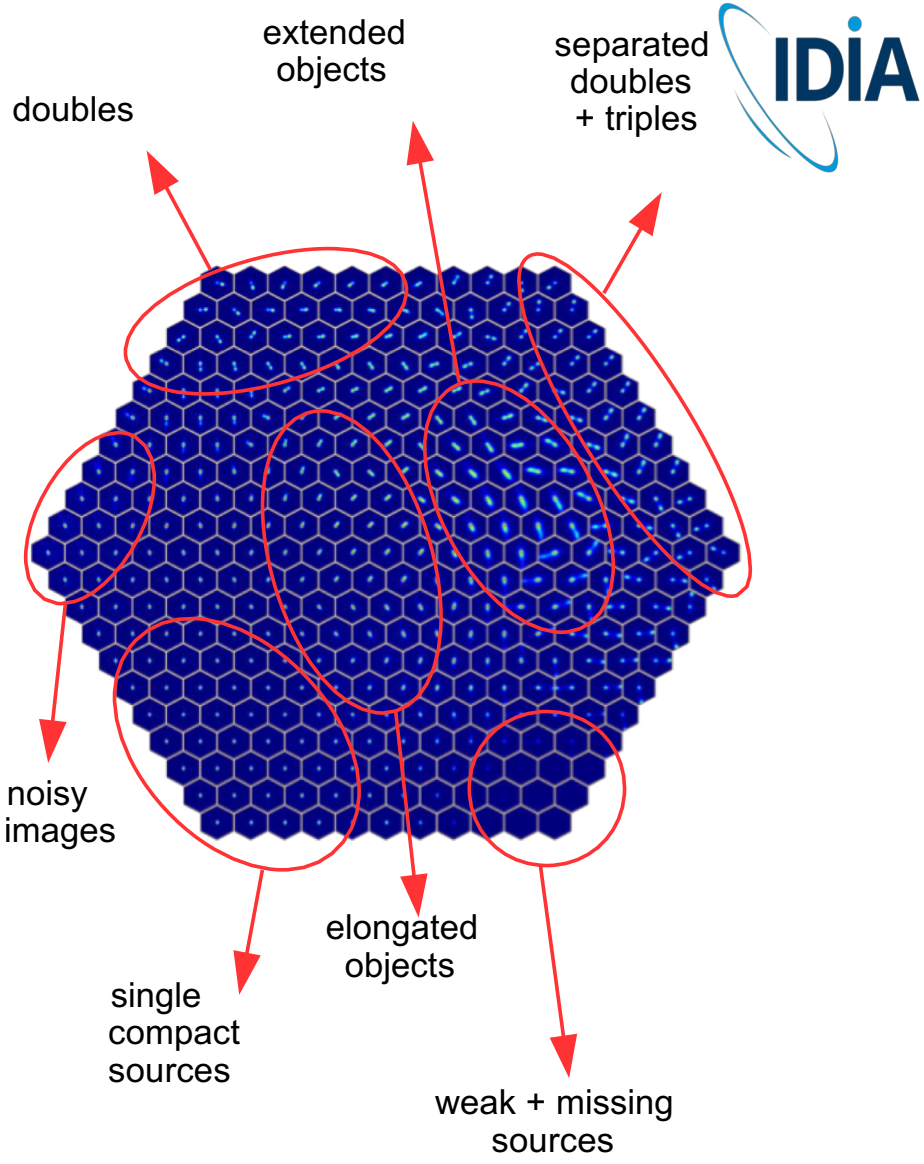
MeerKAT MIGHTEE CDFS Data.
Stokes-I Map of 55-dish 4k/**856 MHz** Data: Noise Limited!

Self-organized maps

Automated & Unsupervised
Classification & Outlier Selection



Polsterer et al. 2015



CNN for radio galaxy classification

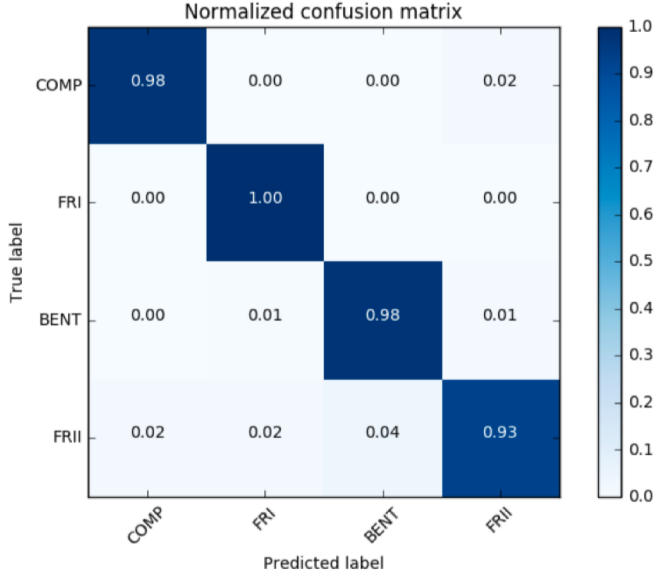


Figure 8. Normalized Confusion Matrix on our testing dataset, where the true labels at Y axis and Predicted labels at the X axis, the blue boxes at the diagonal represents the recall values .

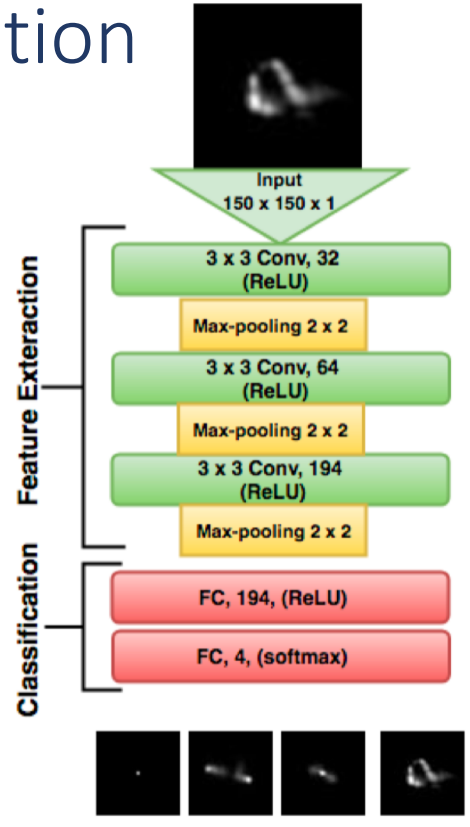
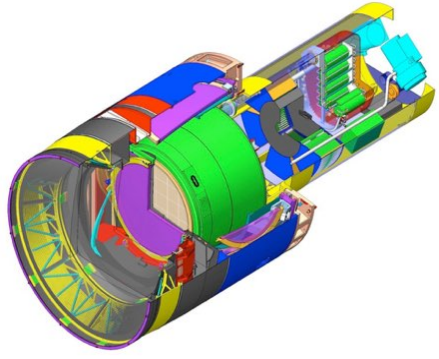


Figure 6. Deep CNN model architecture. In this Figure each convolutional layer (Conv) is followed by ReLU activation function and Max-pooling function. The number in each box represents the number of channels in the corresponding feature map.

Wathela Hassan, MSc. UCT: 2018

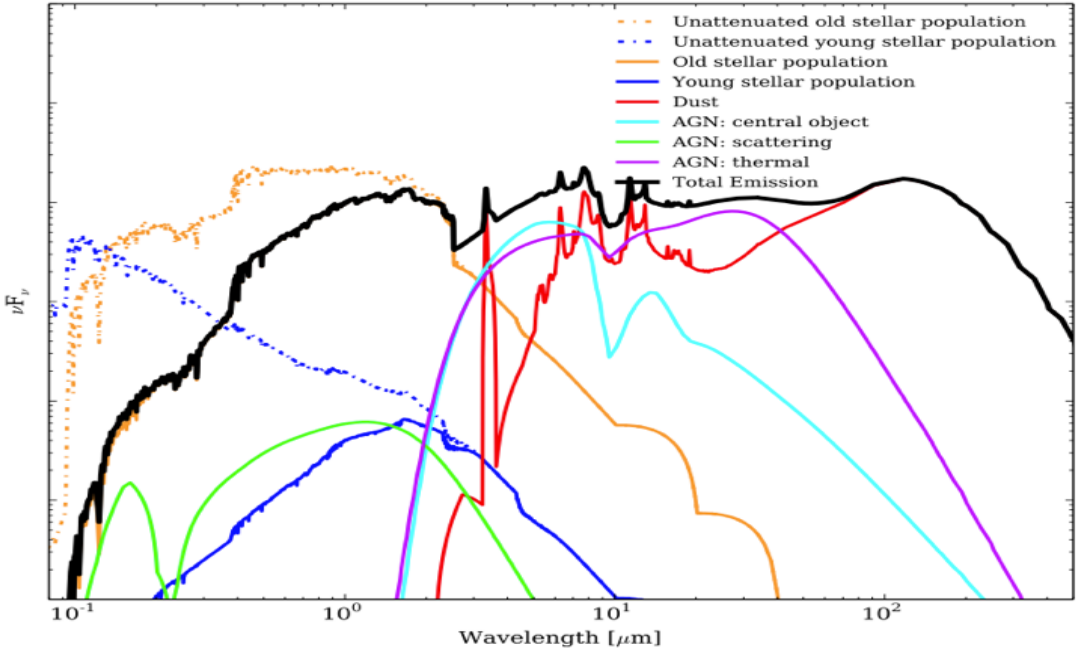
Multi-wavelength: Fusion of optical/IR and radio



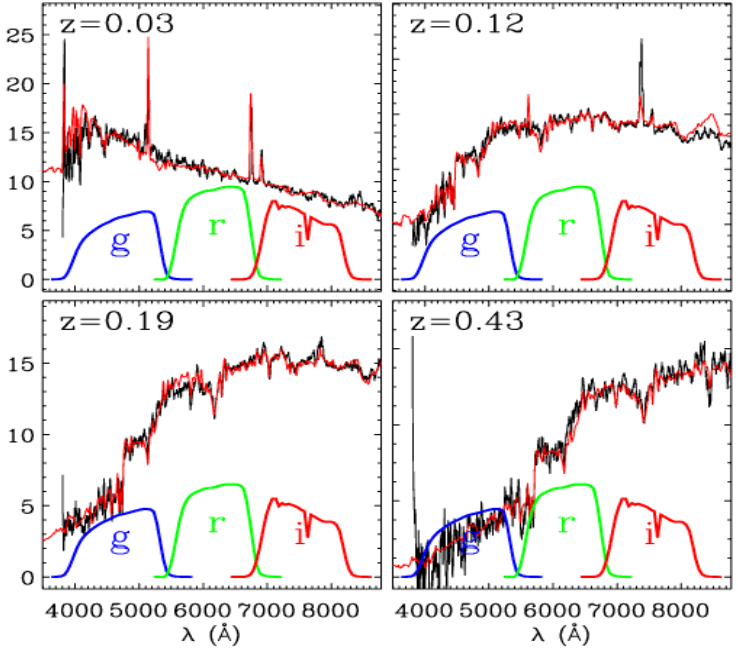
- **The Ultimate Optical Transient Survey Facility in the SKA Era**
  - 8-m class telescope equipped with 3.2 Gigapixel camera
  - Operational 2020, US-led international collaboration
- LSST & SKA share a number of Science Goals and Data Challenges and are timed and placed so as to be able to be exploited jointly

Multi-wavelength data encode redshift and physics

Spectral energy distributions of different types of emission

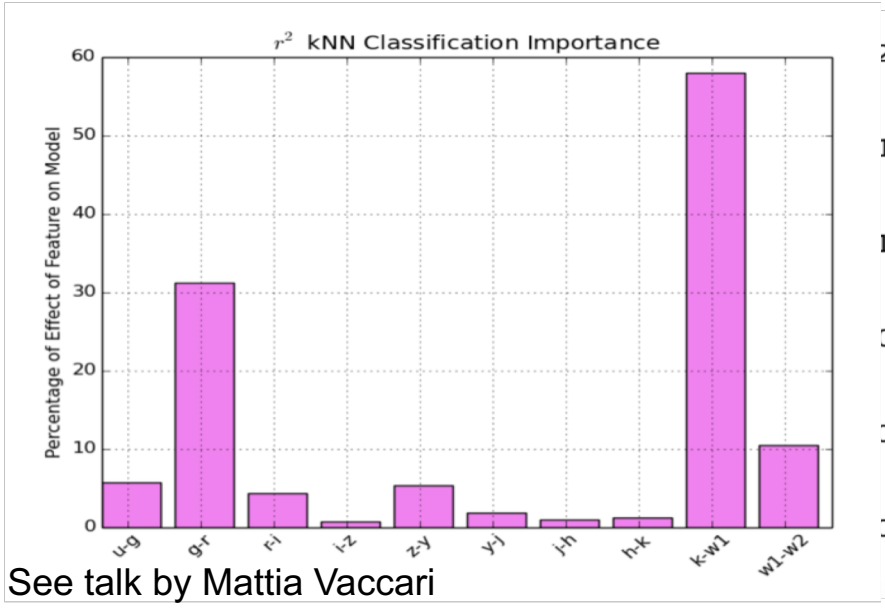


Relative Photometry changes with redshift

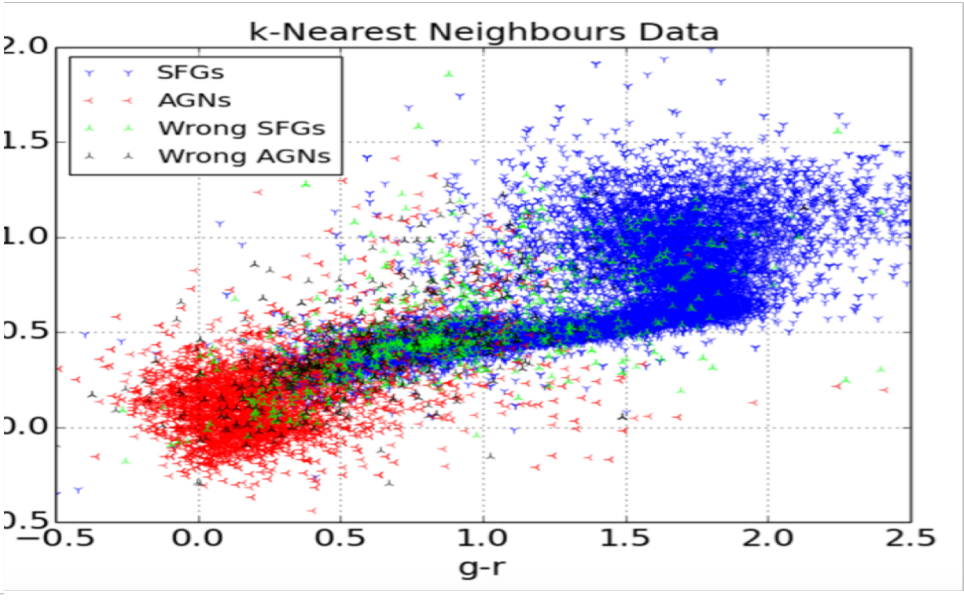


Cloud-enabled **Machine Learning**

- Case Study : combining SDSS & WISE multi-band photometry to classify sources as SFGs/AGNs
- Feature Ranking to Optimise Feature Selection



See talk by Mattia Vaccari

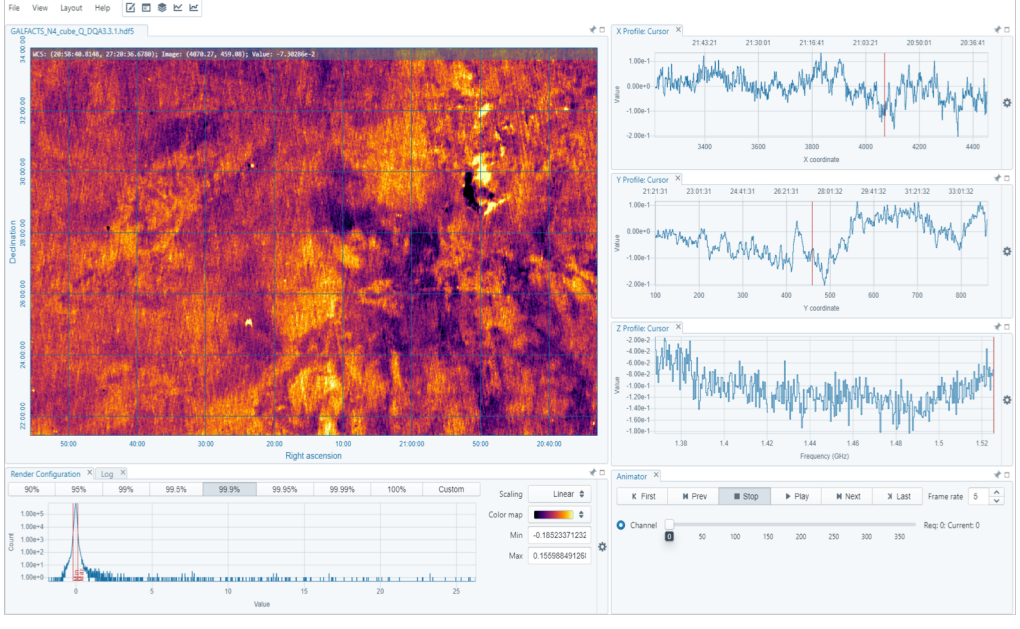


Visual analytics of big data in the cloud

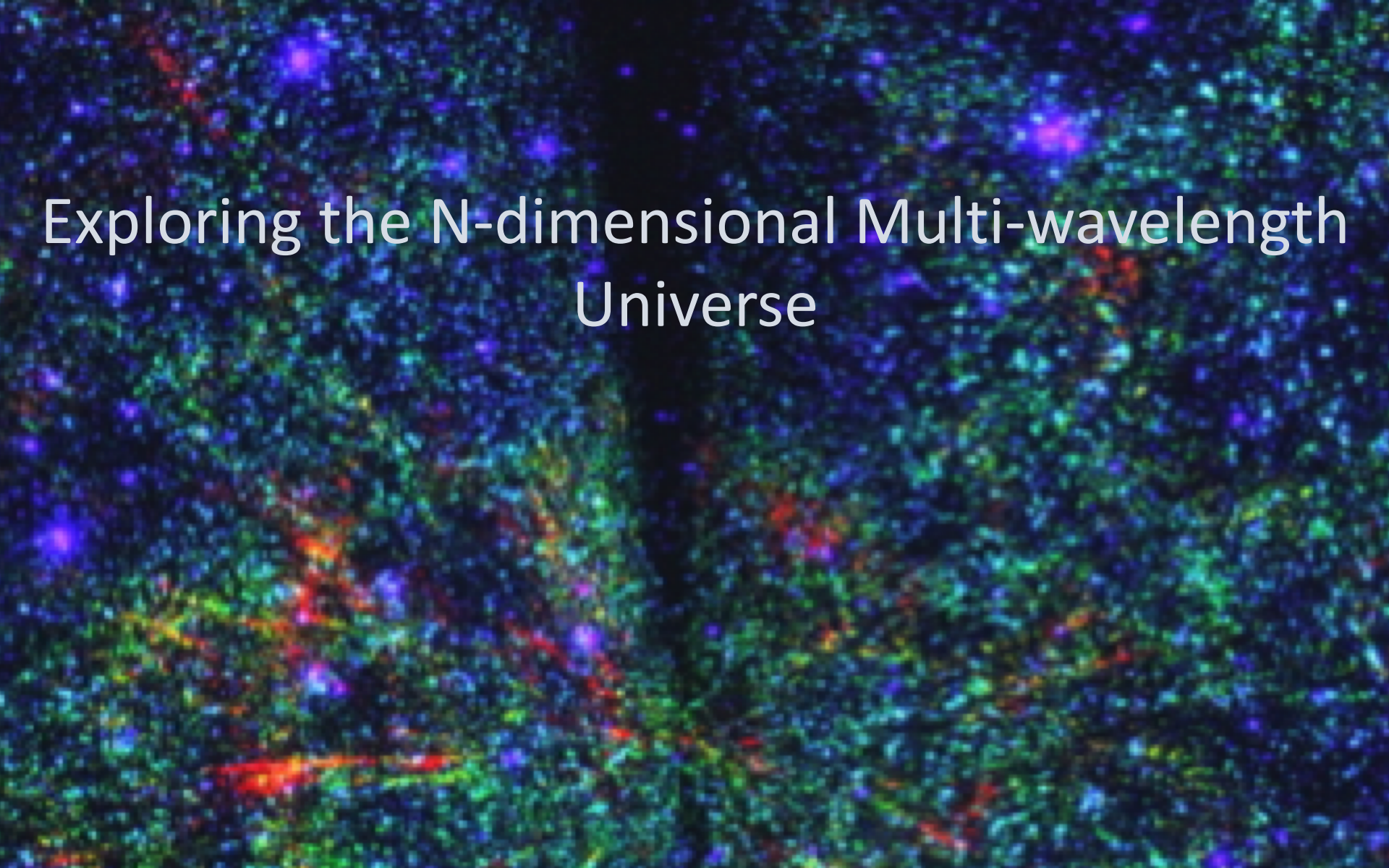
Cube Analysis and Rendering Tool for Astronomy

IDIA (South Africa) – NRAO (US) – ASIAA (Taiwan)

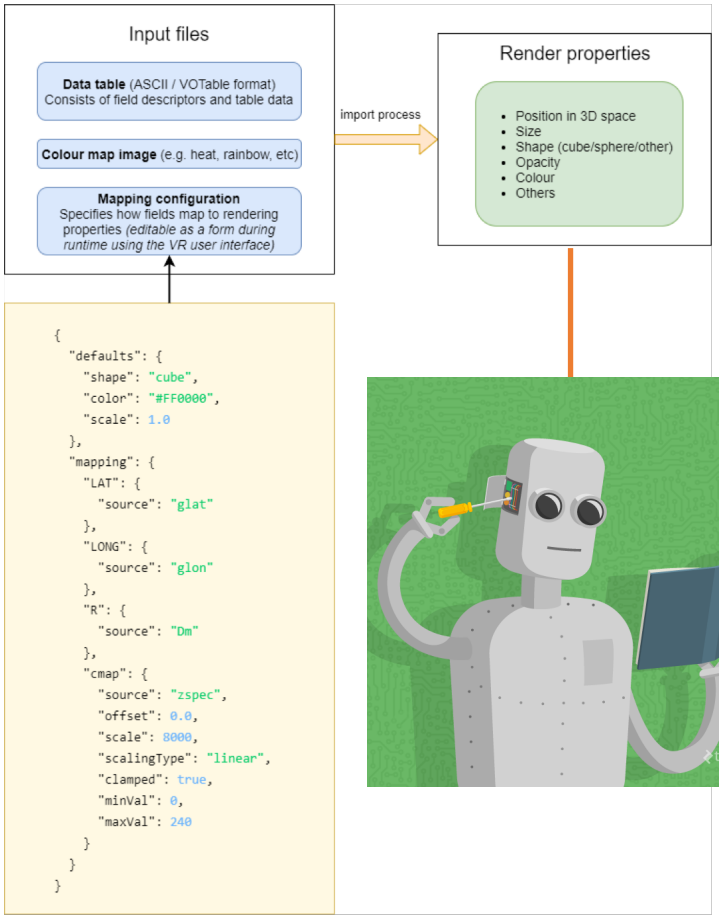
- To be deployed at ALMA Regional Science Centres
- Cloud-based Visual analytic of remote large MeerKAT cubes



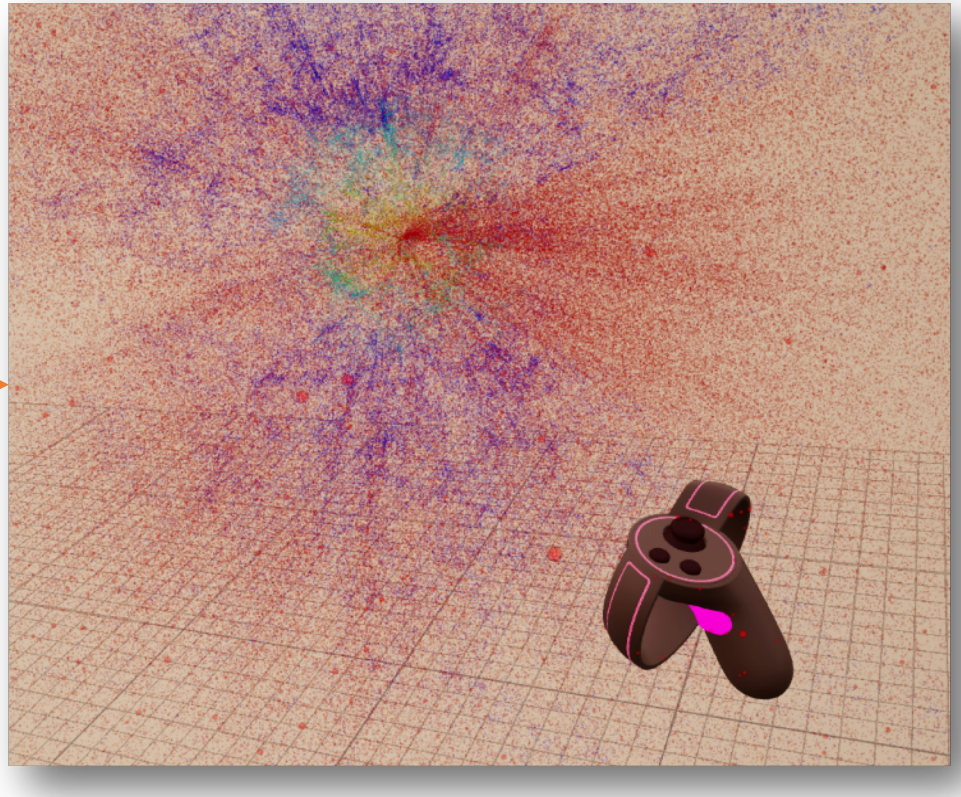
http://cartavis.github.io


Exploring the N-dimensional Multi-wavelength Universe

Big data in Virtual Reality



- supports over 1M data points
- User can move around data by walking around room
- Web-based user interface on user's (virtual) wrist



Cloud for HCD, Development and Outreach

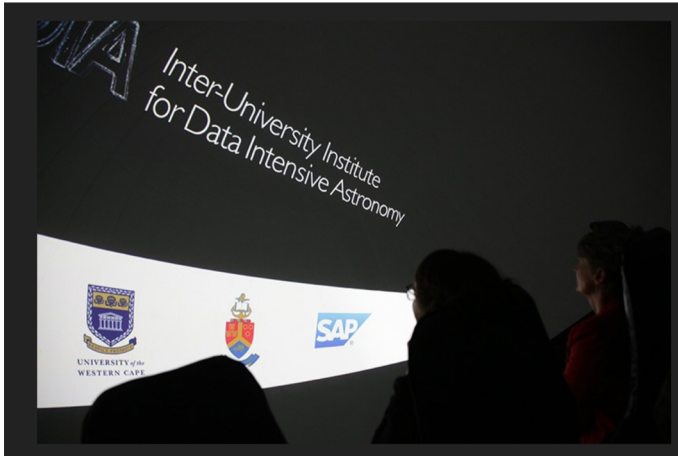
IDIA partners with the SARAO Big Data Africa school.

Published on September 11, 2018 by carolune

DARA BIG DATA

Development in Africa with Radio Astronomy

The week of 10 – 17 September 2018 sees the 2nd Big Data Africa school take place near Cape Town. Organised by SARAO as part of the Newton Fund's DARA Big Data project, the school brings 27 students from 8 African countries together with lecturers, local and international, to take part in an intense, hands-in academic school where they get first hand experience with big data. IDIA has partnered with the school and the students develop and run projects on the IDIA research cloud. The projects offered to the students of the school range from radioastronomy to cybersecurity.



On July 23 2018, we invited members of the media and of the astronomy community to an event at the Iziko Planetarium and Digital Dome to celebrate the IDIA Research Cloud. The event consisted of an introduction, a show and a panel discussion followed by refreshments. The show itself featured previously unseen and unpublished astronomical images from the MIGHTEE experiment on MeerKAT and from the MeerLICHT telescope, overlaid to illustrate the complementarity of astronomy at different wavelengths.

https://youtu.be/Il0luYyS0N4

BRICSKA:
Big **D**ata **R**esearch **I**nfrastructure **C**ollaboration toward the **SKA**

Short term implementation (1-3 years)

- Workshops and student/researcher exchanges focused on the development, testing and improvements of big data technologies in the areas of
    1. Open source cloud implementation and technologies
    2. Development of novel algorithms and HPC software for processing and analysis of big astronomy data
    3. Build open source based architectures and tools to enable deployment of cloud-based HPC systems, and
    4. Human capital development (in collaboration with OAD).

Medium-term implementation (3-5 years)

- continued workshops and student/researcher exchanges with active development in high performance computing data processing and the use of cloud-based tools for data visualization and analytics
- The establishment of a BRICS federated network of cloud-based infrastructure for data intensive astronomy and related research

Additional Resources

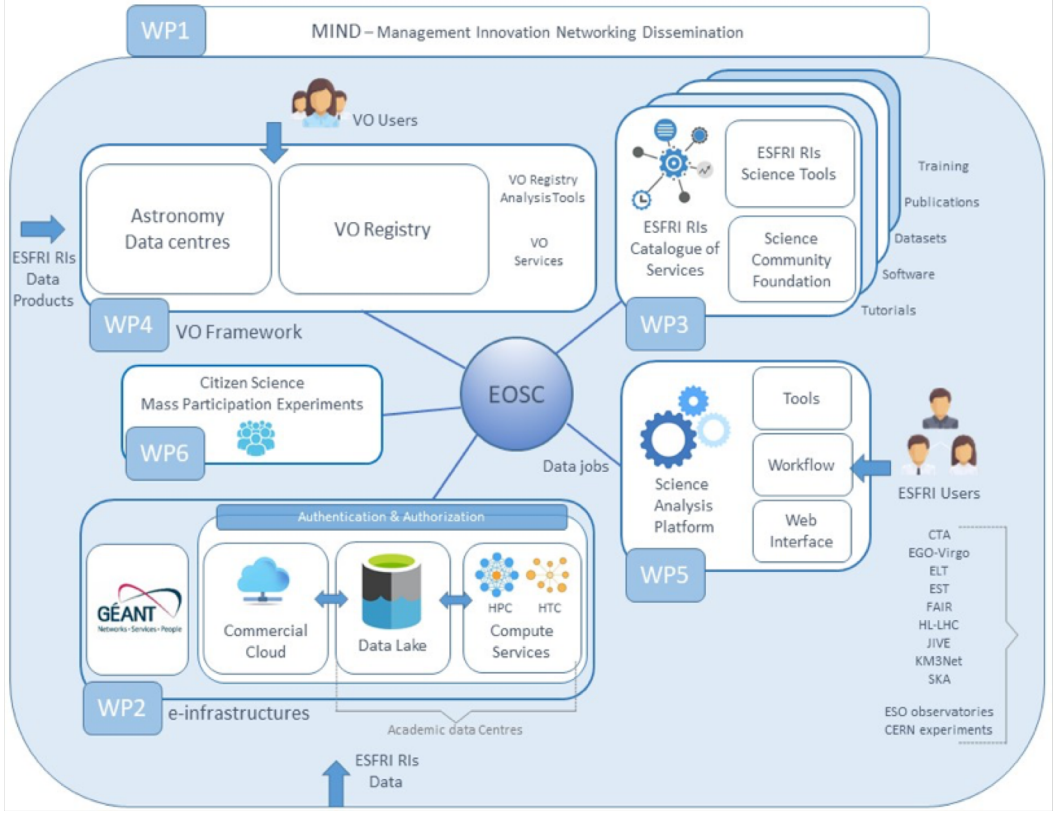
- Funding for equipment to build BRICS cloud infrastructure (hardware and software)
- Student and postdoctoral fellowship bursaries to develop joint BRICS big data projects

ESCAPE

European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

ESCAPE aims to address the Open Science challenges shared by ESFRI facilities as well as other pan-European research infrastructures in astronomy and particle physics

- EC H2020 (16 M€, 2019-2023)
- Partners include SKA, CTA, KM3Net, EST, ELT, HL-LHC, FAIR, CERN, ESO, JIVE
- Led by CNRS, 32 different EU institutions
- ASTRON leading Science Platform WP
- Work kicks off in January 2019



Australia Development 2018-2021

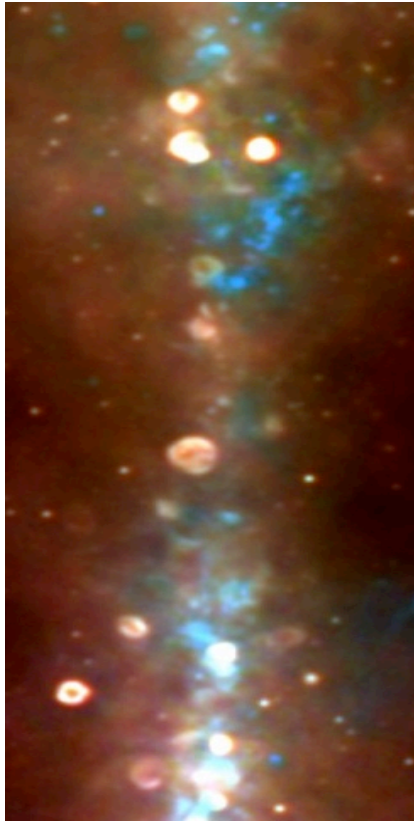
- $4M seed funding received
- AusSRC Management Committee has been formed (ICRAR, CSIRO, ASKAP, MWA, PAWSEY)
- Science Advisory Panel to the Project Teams has been formed
- AusSRC Project Team forming will be completed in Q1 2019
- AusSRC major funding request for SKA1 (Q2 2020)

Canadian Initiative for Radio Astronomy Data Analysis

- **CIRADA** (cirada.ca)
  - enhanced data products (EDPs) for VLASS, CHIME and ASKAP surveys
  - advanced re-processing of raw data
  - unified processing software stack
  - cross-matches, advanced analytics, visualisation
  - long-term archiving and data access
  - ***enables full science return from major Canadian science & instrumentation programs***

- Administrative structure
  - CFI Innovation Fund 2017: $10.6M
  - PI: Bryan Gaensler ; Deputy PI: Erik Rosolowsky
  - five year program, commenced April 2018
  - six Canadian universities: Toronto, Alberta, McGill, Queen's, UBC, Manitoba
  - plus NRC/CADC, Compute Canada, NRAO, ASTRON, IDIA, Cornell, Berkeley, Minnesota

Dunlap Institute for Astronomy & Astrophysics
UNIVERSITY OF TORONTO

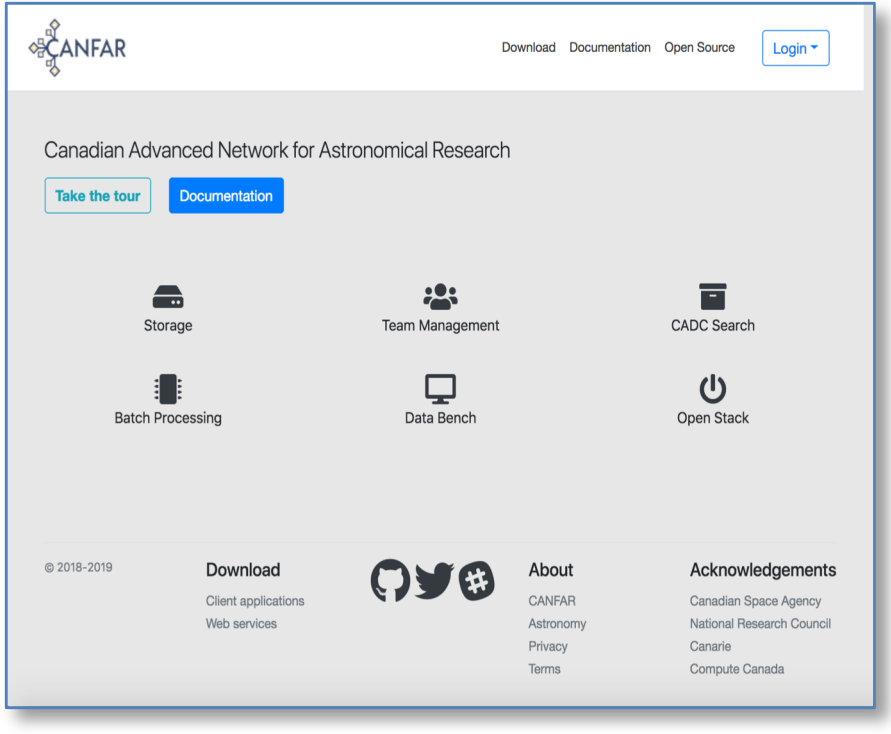


Hindson et al. & Johnston-Hollitt et al.

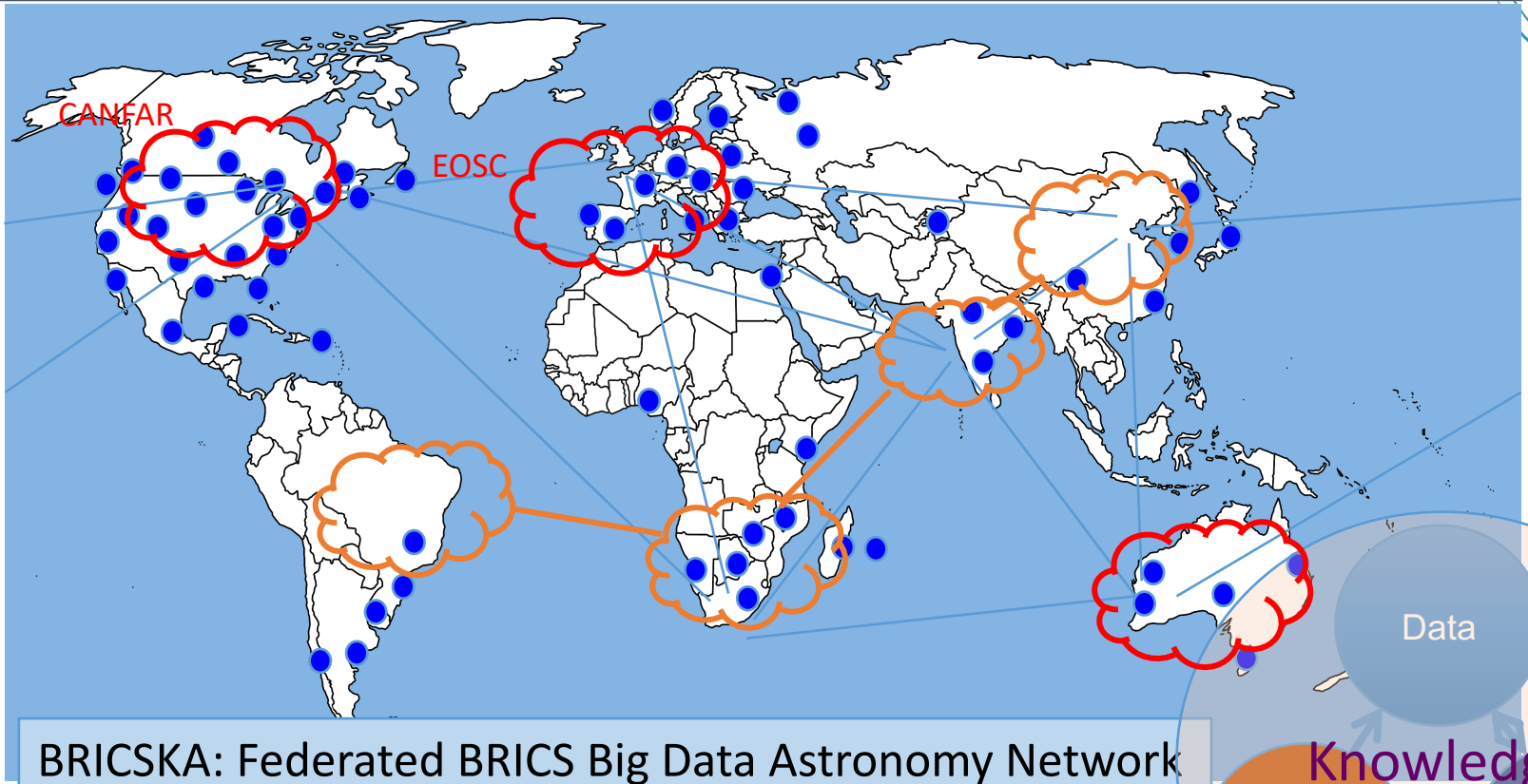
Canadian Advanced Network for Astronomical Research

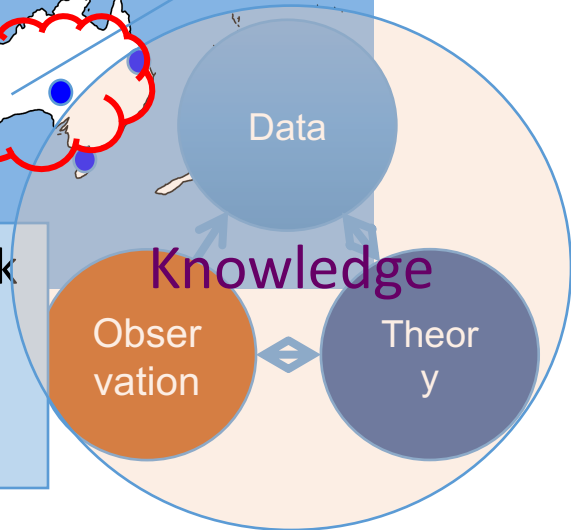
Science platform and cloud ecosystem for data intensive astronomy

- User services
  - User storage
  - Interactive and batch processing
  - User databases
- Federated research cloud resources provided by Compute Canada
- Integrated:
  - Authentication and authorization
  - Access to telescope data
  - Access to user storage
- Extending capabilities and capacity to support radio data projects (CIRADA)



Big Data Astronomy in the SKA Era





BRICSKA: Federated BRICS Big Data Astronomy Network

- Collaborative development of big data technologies for joint exploitation of big astronomy data
- Integrate into global network

BRICSKA:
Big **D**ata **R**esearch **I**nfrastructure **C**ollaboration toward the **SKA**

Tangible outputs:

- A network of federated cloud-based infrastructure and e-science tools
- Cloud-based provisioning of HPC for automated processing pipelines
- A wide-field VLBI post-processing data pipeline and source extractor as Jupyter notebook, new technologies and modalities for visualization such as machine learning enhanced visual analytics of cloud-based remote data
- A visual classification tool improving upon radio galaxy zoo to browse, classify and annotate radio sources to discover the unexpected and leverage the power of multi-wavelength data
- Cloud-enabled systems for fusion of and joint analytics of LSST and radio data
- Innovate data analytics tools for large data sets driven by machine learning techniques
- Cloud-based tools for collaborative exploration by distributed research teams
- A cloud-based astronomy and data science toolkit for outreach and development

We expect this collaboration to have far-reaching effects toward developing data intensive expertise and experience in South Africa.

CHINESE – SOUTH AFRICAN COLLABORATIONS IN ASTRONOMY: WORKSHOP PROGRAM

The Big Data Challenge in Astronomy

With Cui Chenzhou: Director, Centre of Information and Computing, NAOC
Project Manager, Chinese Virtual Observatory

Final workshop: First quarter 2019