ilifu

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
DIRISA

The ilifu Cloud Computing Facility

Enabling MeerKAT Science

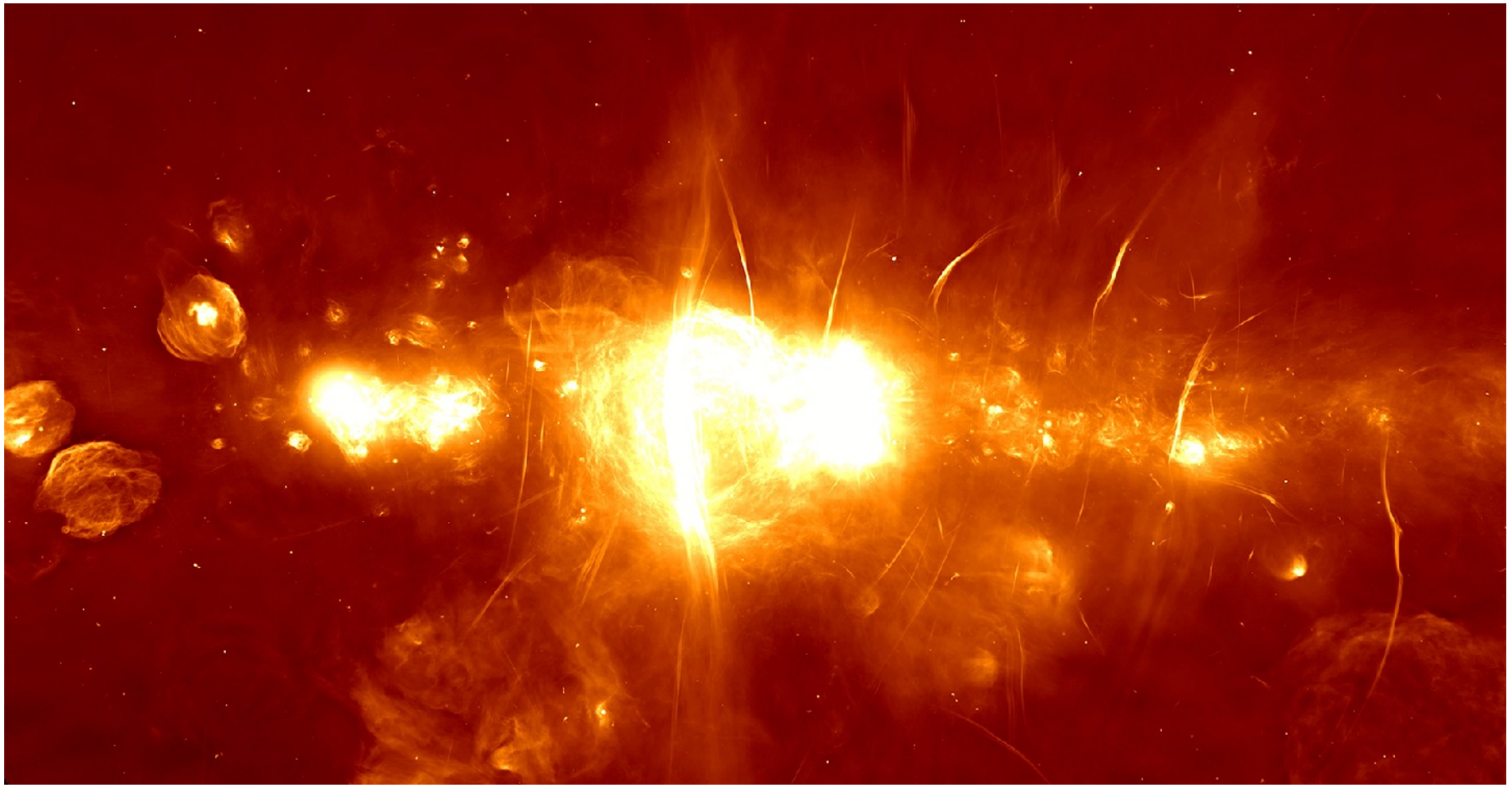
IDIA

Mattia Vaccari - IDIA/UWC
www.mattiavaccari.net

SKA Italy III, 8 October 2021

RESPICE PROSPICE
UNIVERSITY of the WESTERN CAPE

MeerKAT : South Africa's SKA "Precursor"



- Completed (on schedule and within budget) in Mid-2018
- Delivering Transformational Science from Day One
- Will be owned and operated by South Africa for 5 years

SKA Timeline & South African Facilities

2010 | 2015 | 2020 | 2025 | 2030

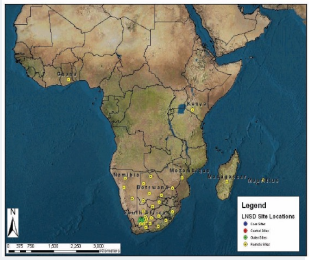
1% | 10% | 100%

KAT-7

MeerKAT

SKA1

SKA2



Science

Pathfinder Science
global facilities
Early studies (e.g. VLA, GMRT deep fields, CHILES,..)

→

Precursor Science
MeerKAT, ASKAP
Large Survey Projects, VLASS

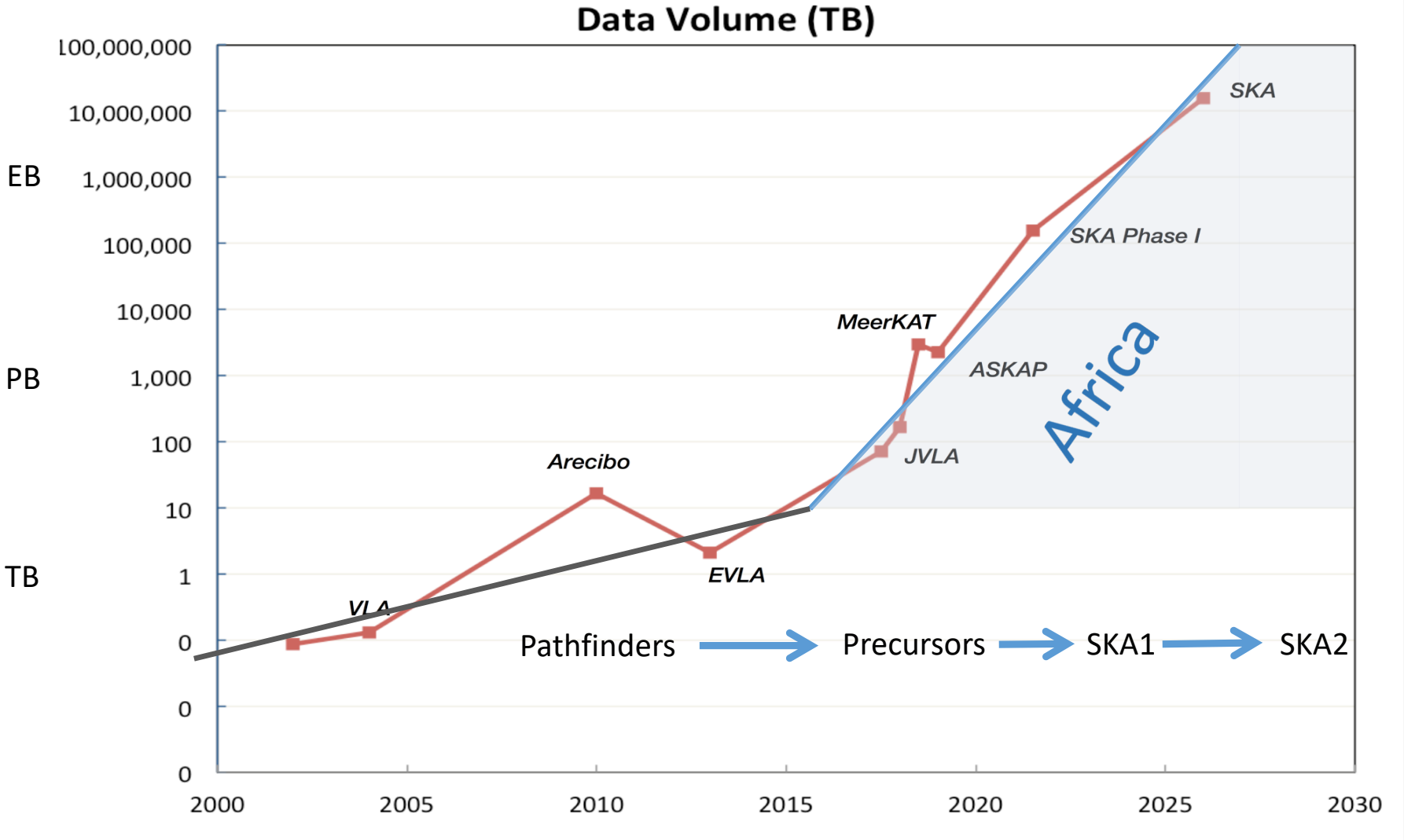
→

SKA1
Key Science Survey Programs

→

SKA2
Wide-field
Long baseline survey
radio astronomy

Growth of Data Volumes in Radio Astronomy



Changing Sociology of Radio Astronomy

- Much of the key science en route to the SKA will be achieved via large-scale observing programs executed by globally distributed teams of researchers working on the data in a collaborative manner

IDIA Inter-University Institute for Data Intensive Astronomy

from big data to big ideas

https://www.idia.ac.za

- IDIA (Inter-University Institute for Data Intensive Astronomy) was launched in Sep 2015
- Driven by MeerKAT/SKA Data Delivery, Processing & Mining challenges while leveraging South Africa's growing involvement in multi-wavelength projects (SALT, LSST, HESS/CTA)

MeerKAT Key Science Large Survey Projects

Imaging

Time Domain

- 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)
- MALS (extragalactic HI absorption)
- ThunderKAT (exotic phenomena, variables and transients)
- TRAPUM (pulsar search)
- MeerTime (pulsar timing)
- MESMER (High-z CO)
- MeerGAL (Galactic Plane Survey)

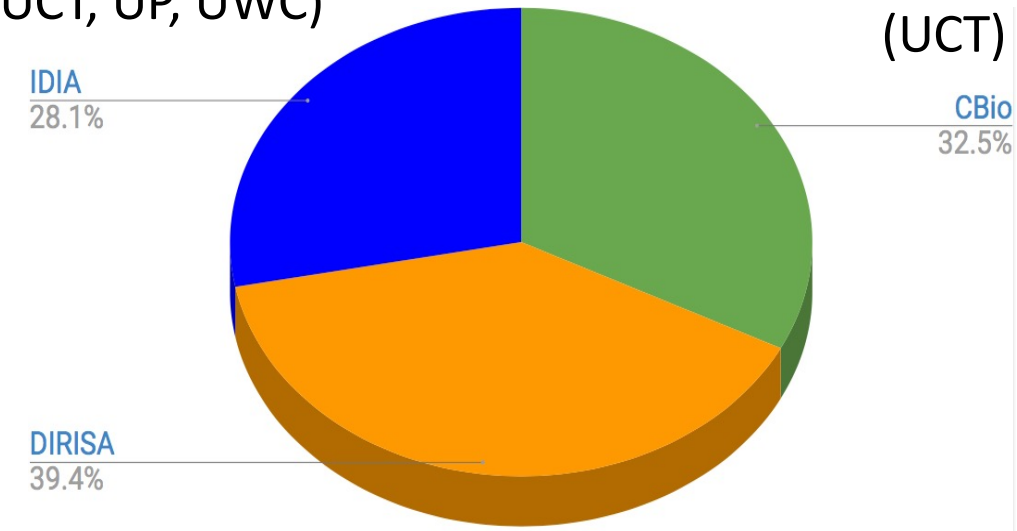
https://www.sarao.ac.za/large-survey-projects

Ilifu (Cloud) Facility Staged Roll out 2018-2021

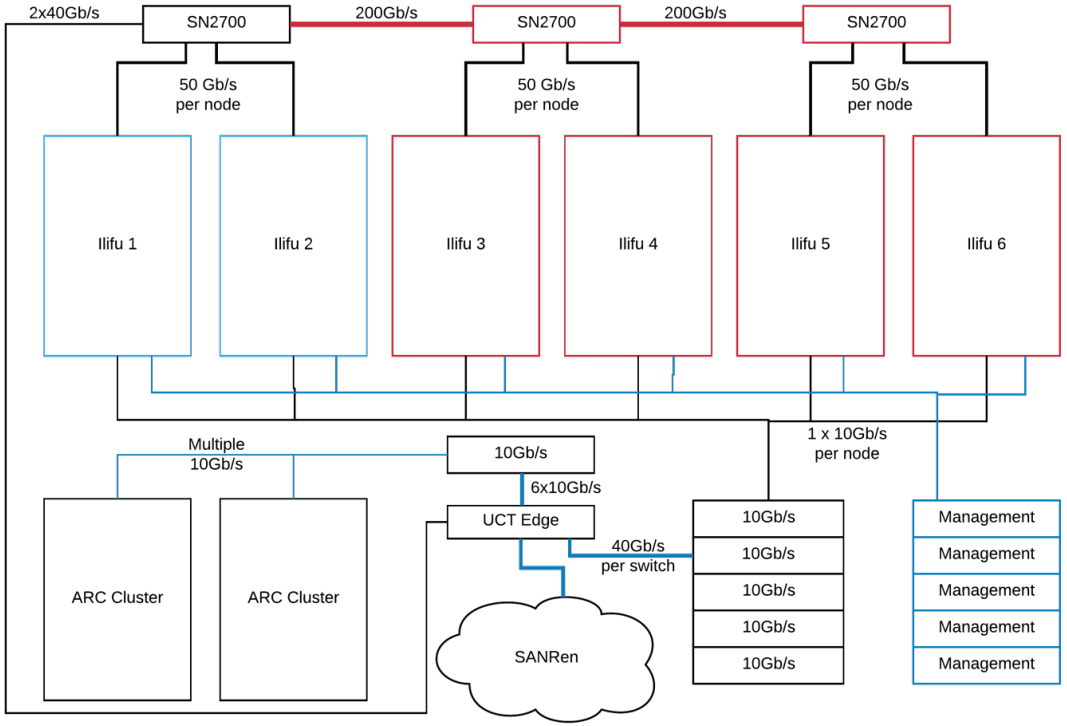
ilifu

http://www.ilifu.ac.za

(UCT, UP, UWC)



(UCT)

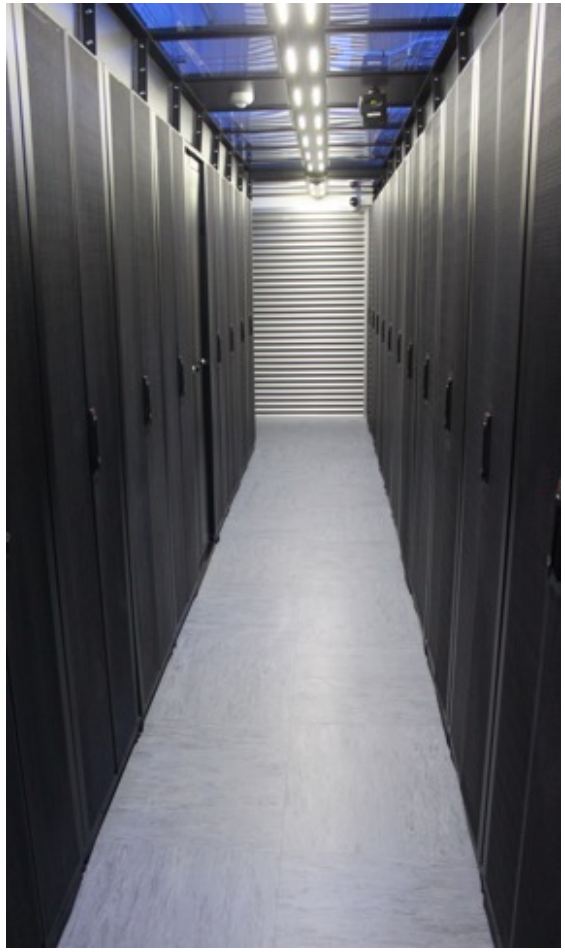


(UCT, UWC, CPUT, SU, SPU, SARAO)
Astronomy & Bioinformatics

- Entire system available to all partners via fair share and managed by university researchers
- Implemented as data intensive research cloud (v2.0) based on IDIA astronomy cloud (v1.0)
- IDIA and CBIO resources are allocated and managed by the relevant consortia
- DIRISA-funded resources allocated to ilifu partners via competitive process

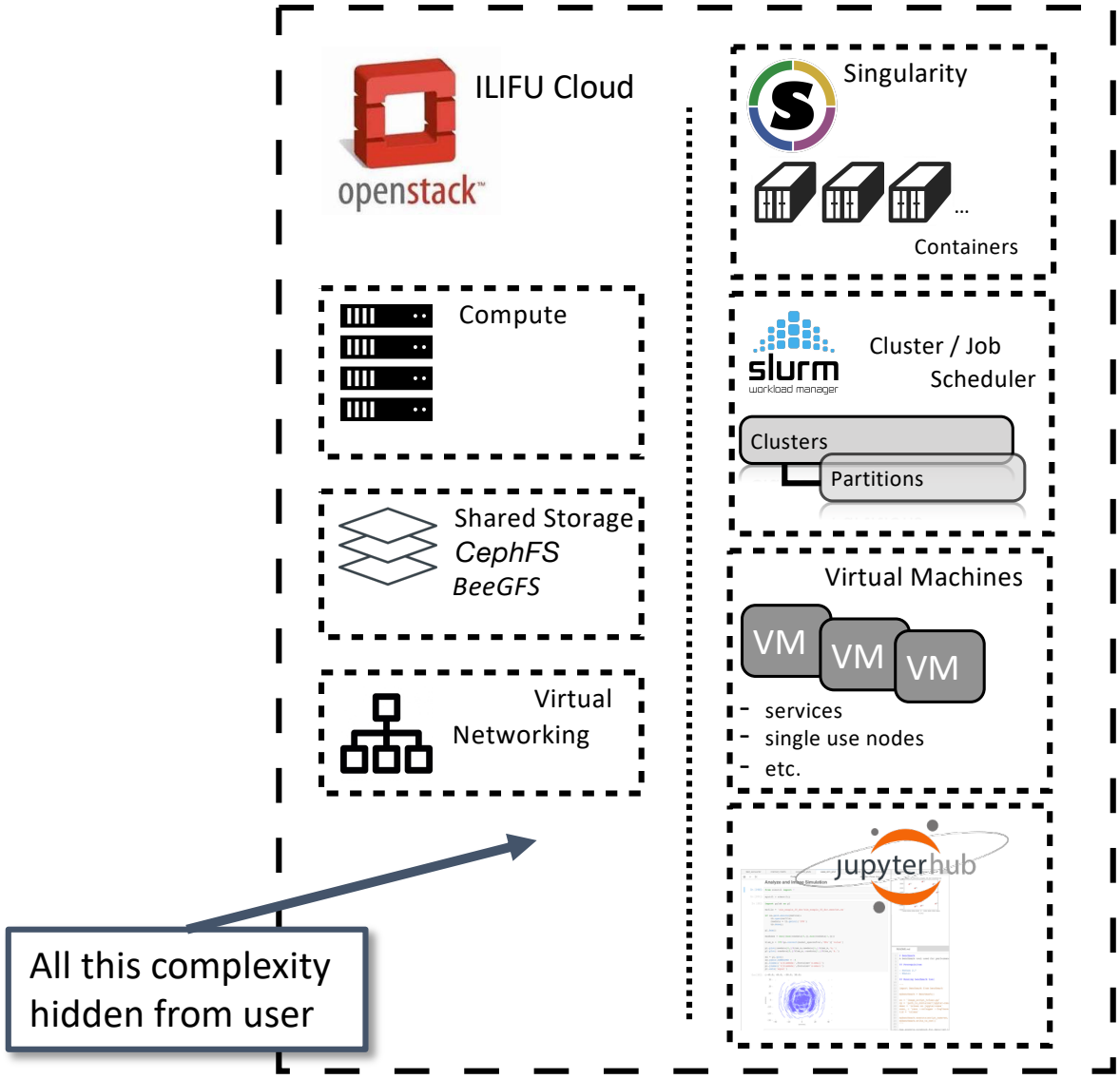
Ilifu Data Intensive Research Cloud: 2015-2021

- Big data astronomy and bioinformatics research in SA
- Cloud platform on data-centric computing architecture
  - Design by Rob Simmonds (IDIA/UCT)
  - Builds on core services provided by its predecessors
    - ARC Prototype (UCT/NWU, 2015)
    - IDIA Cloud (UCT/UP/UWC, 2017)
  - Data-centric architecture with large local POSIX storage
  - 120 Compute nodes
    - 2.6 GHz Intel Processors, 32 cores, 256 GB RAM / node
    - 2 nodes have 512 GB RAM each
    - 4 nodes also have 2 x Nvidia Tesla P100 GPUs
  - 4.2 PB CephFS + 0.4 PB BeeGFS (Usable)
  - Hosted by UCT ICTS with 10 Gb/s network to SANReN
    - Ready for upgrade to 100Gb/s after SANReN upgrade



Ilifu Data Intensive Research Cloud (v2.0)

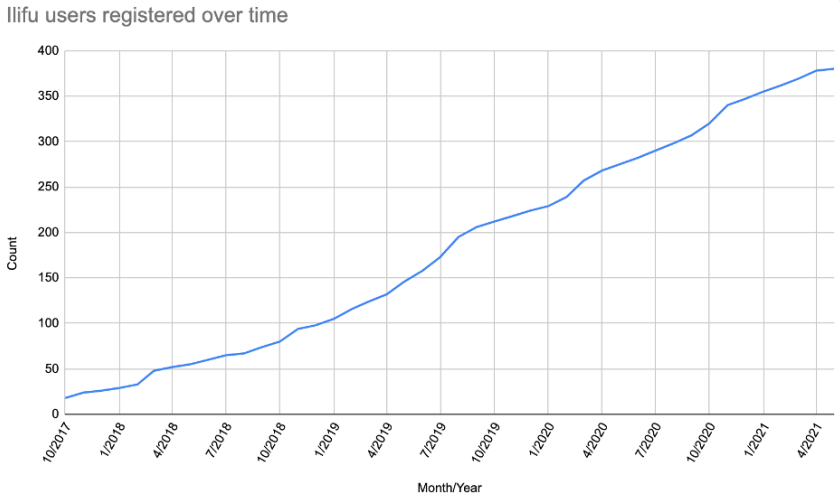
ilifu

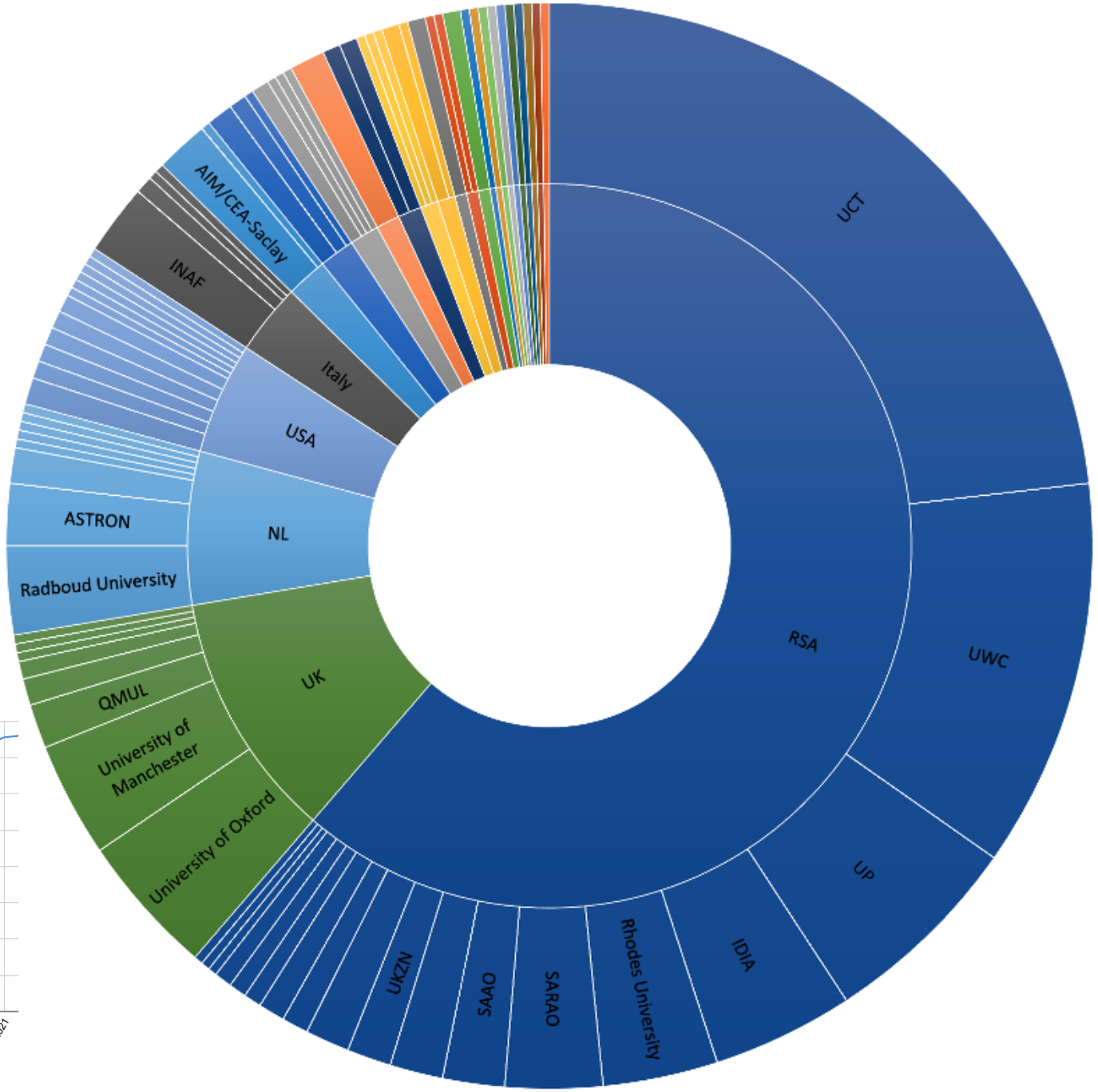


User Interface

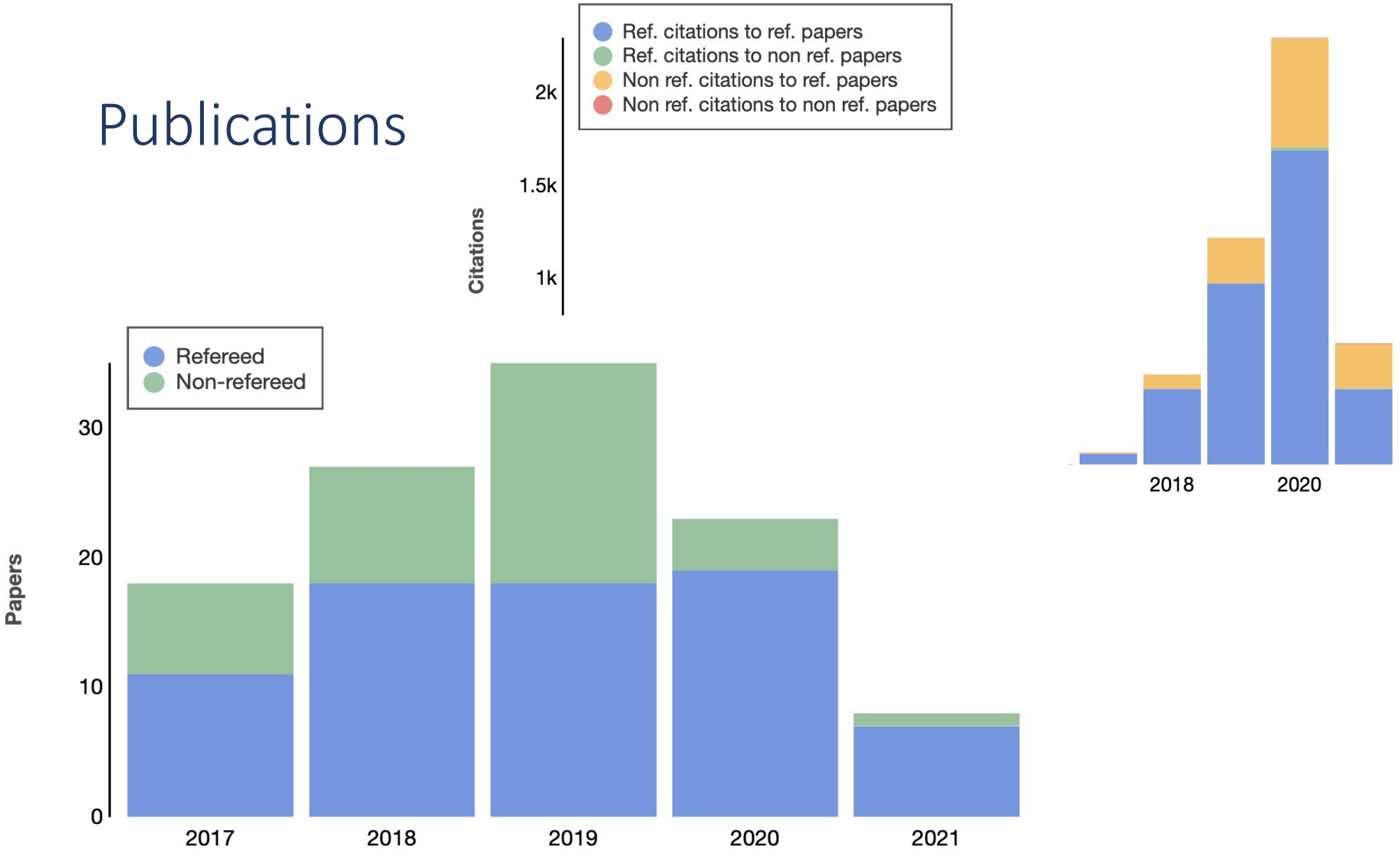
Usage

- Almost 400 users (60% SA)
- 80% Compute Target Usage Routinely Achieved





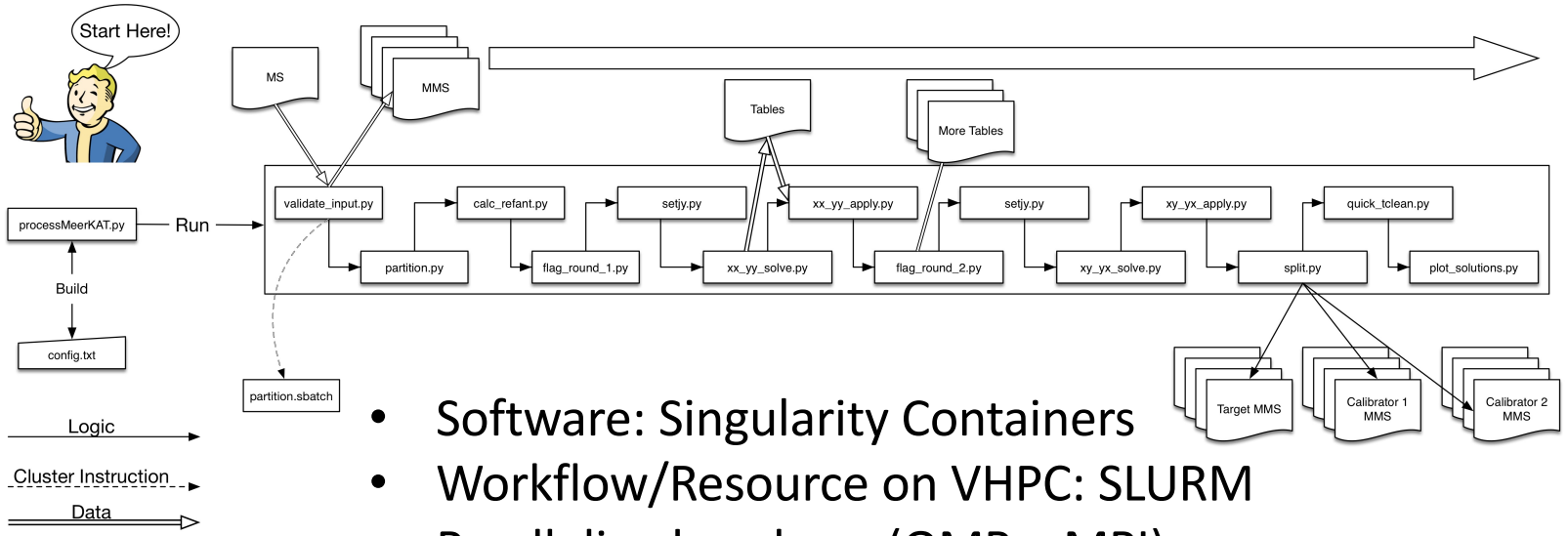
Publications



Human Capacity Development Programme

- Big Data Hackathons (2-3 days) & Schools (10-14 days)
  - Projects in Agriculture, Astronomy, Health and more
  - Guest Speakers & Training from both Industry & Academia
  - Physical, virtual and hybrid events throughout COVID
- H3ABioNet Pan-African Bioinformatics Training

The IDIA MeerKAT Data Reduction Pipeline

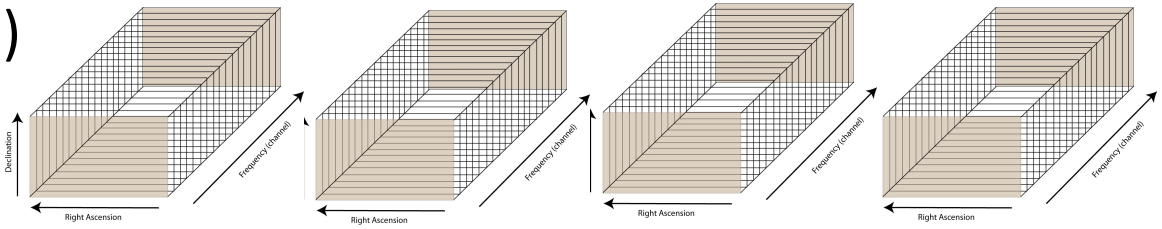


- Software: Singularity Containers
- Workflow/Resource on VHPC: SLURM
- Parallelised package (OMP + MPI)
- User configurable and executable

Data products

- Broad band multi-frequency synthesis images
- 4D spectro-polarimetric data cubes (1k channels)
- 3D HI spectral cubes (32 k channels)

https://github.com/idia-astro/pipelines



The CARTA Big Data Visualization Tool

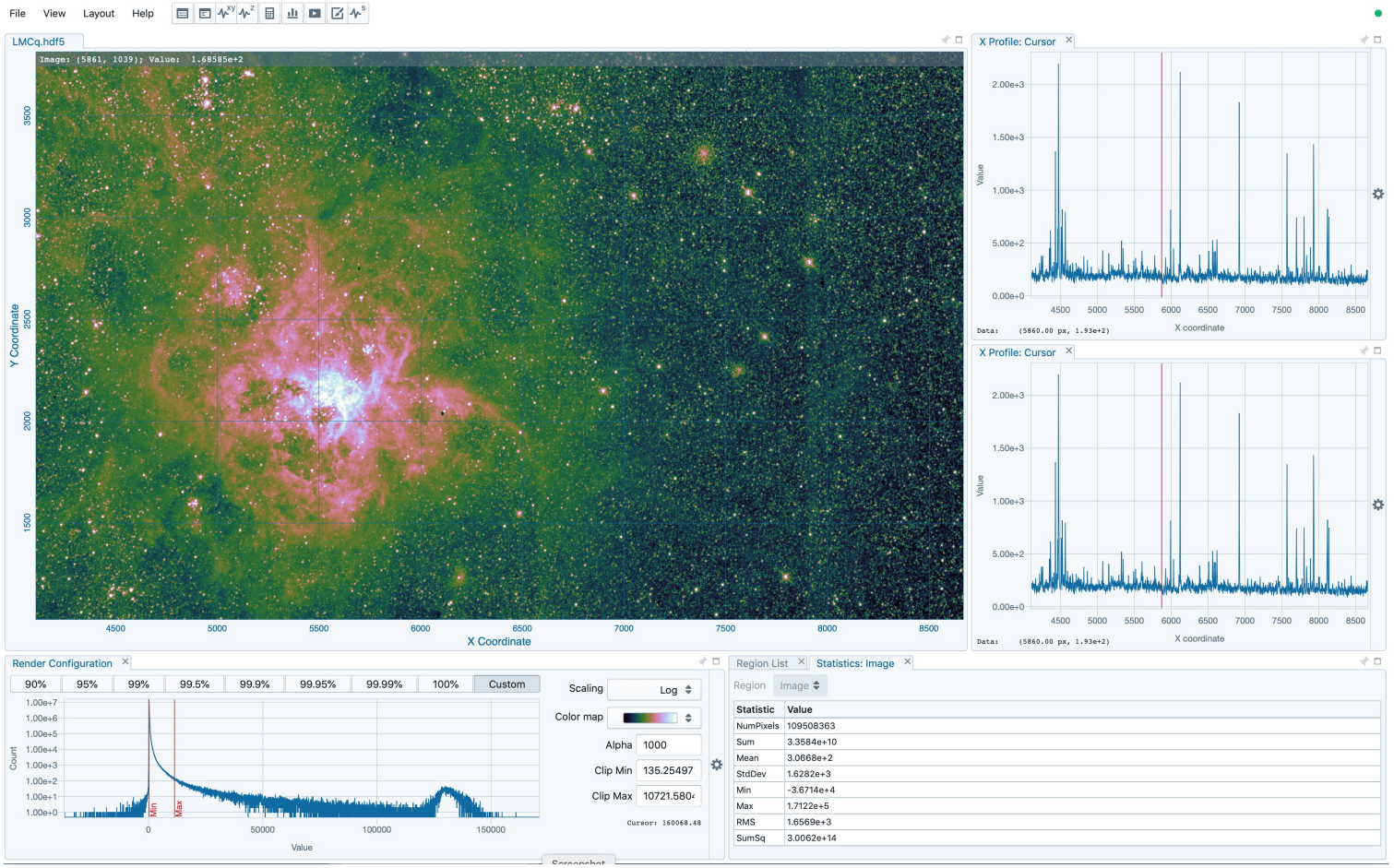
CARTA = Cube Analysis and Rendering Tool for Astronomy - https://cartavis.org

- Enables cloud-based visual analytics of remote large image cubes
- HDF5 Support
- Images & Data Cubes
- Catalogues & Regions

- 1 TB image loading in 5 s

- First deployed on ilifu

- Adopted by an increasing number of Data Centres



Cloud Federation & The EGI-ACE Project

- EGI-ACE's main goal is to implement the compute platform of the European Open Science Cloud and contribute to the EOSC Data Commons by delivering integrated computing platforms, data spaces and tools as an integrated solution that is aligned with major European cloud federation projects and HPC initiatives

- IDIA/ilifu's main task is to enable EGI / EOSC platforms to be accessible by South African researchers and to have (Radio) Astronomy software containers more widely deployed

EXPECTED OUTPUT

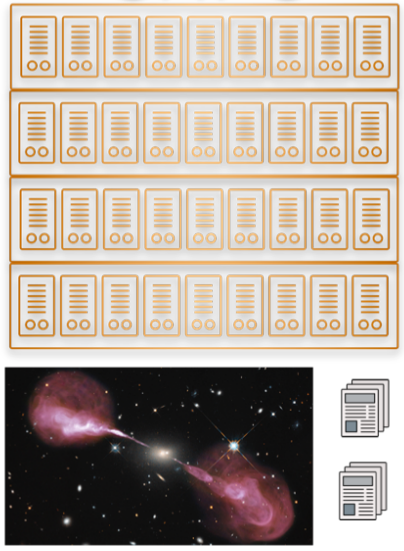
https://www.egi.eu/projects/egi-ace

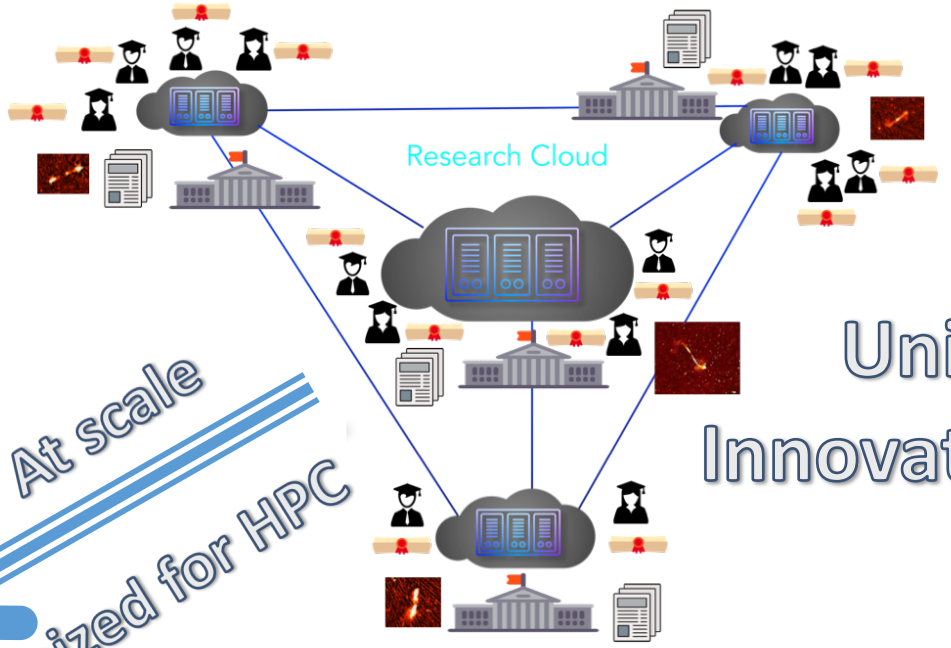
3,000 RESEARCHERS TRAINED

SUPPORT FOR **50,000** USERS

FROM MORE THAN **20** DIFFERENT SCIENTIFIC DISCIPLINES

A South African Data Intensive Research Cloud?





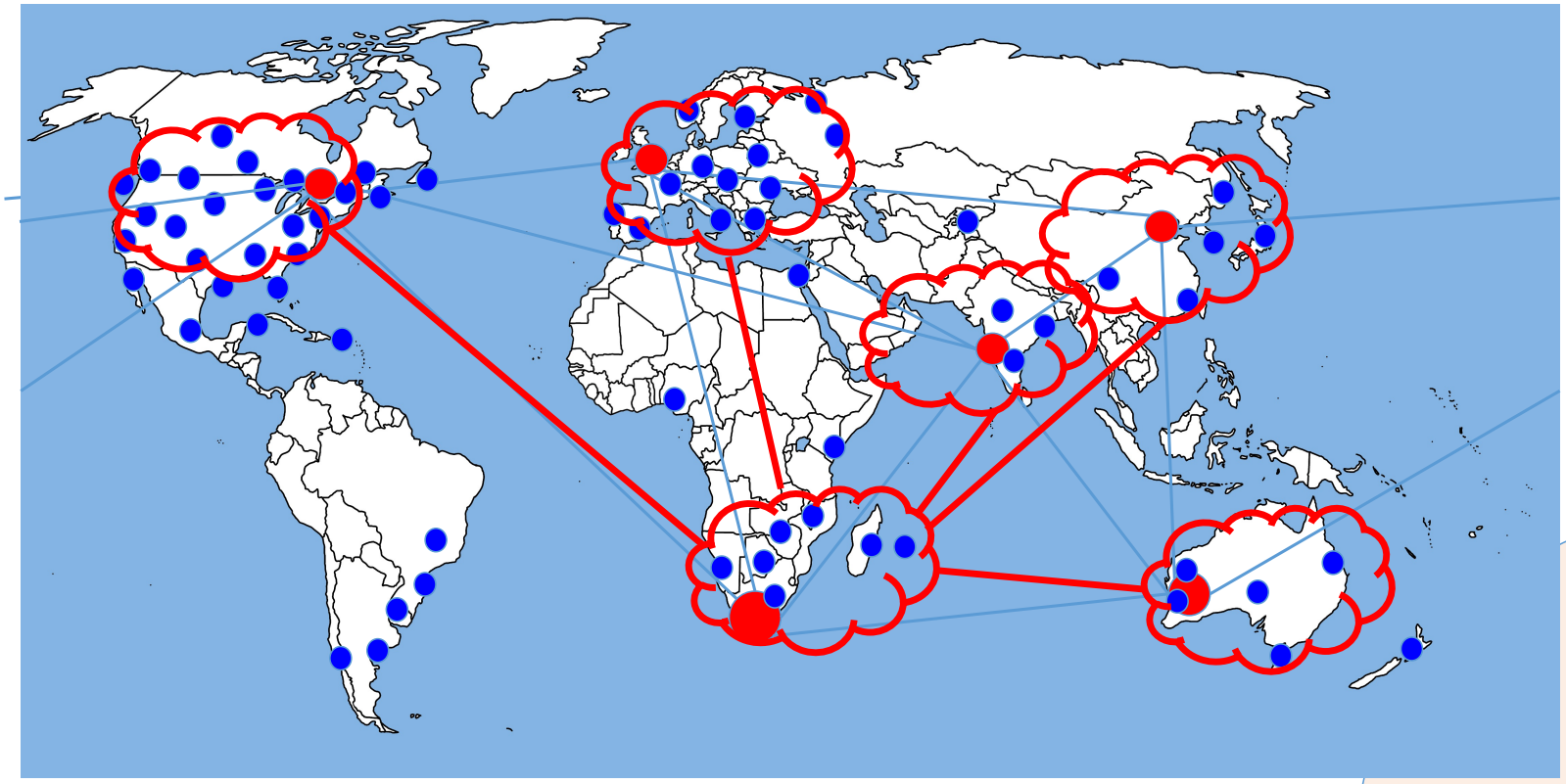
At scale

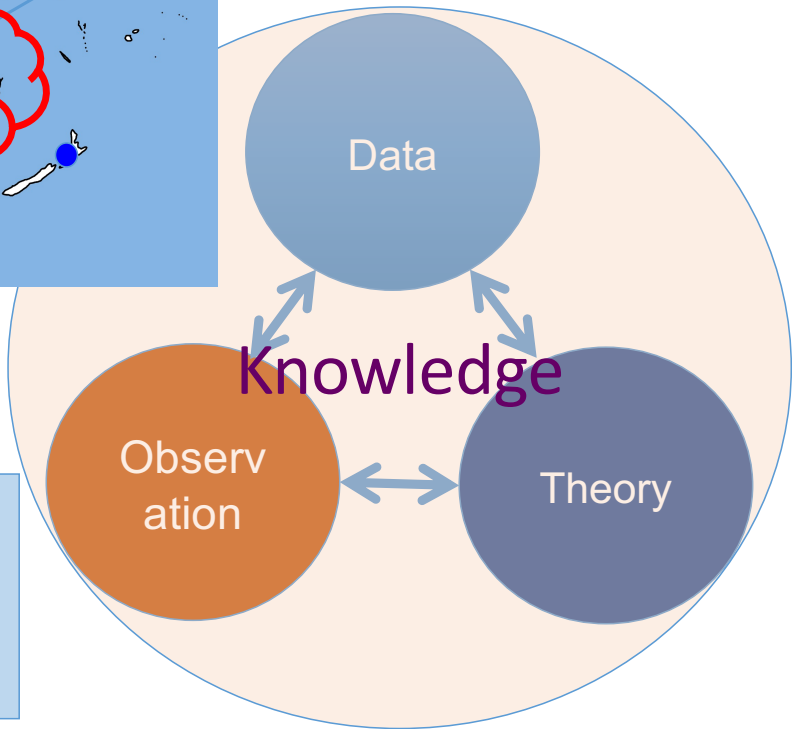
Optimized for HPC

University Innovation cluster

Phase 1	Phase 2	Phase 3
Further develop core technologies for cloud base layer and middleware deployment. Implement federated identity management. Deployments on initial sites. Build technical communities. Initial focus on limited application domains.	Add more sites, including wider African. Develop domain specific application stacks. Add common storage and compute pools. Demonstrate scalable compute within cloud and to national HPC centre	Further development of tools and techniques. Link to international clouds.

SKA Regional Science and Data Centres



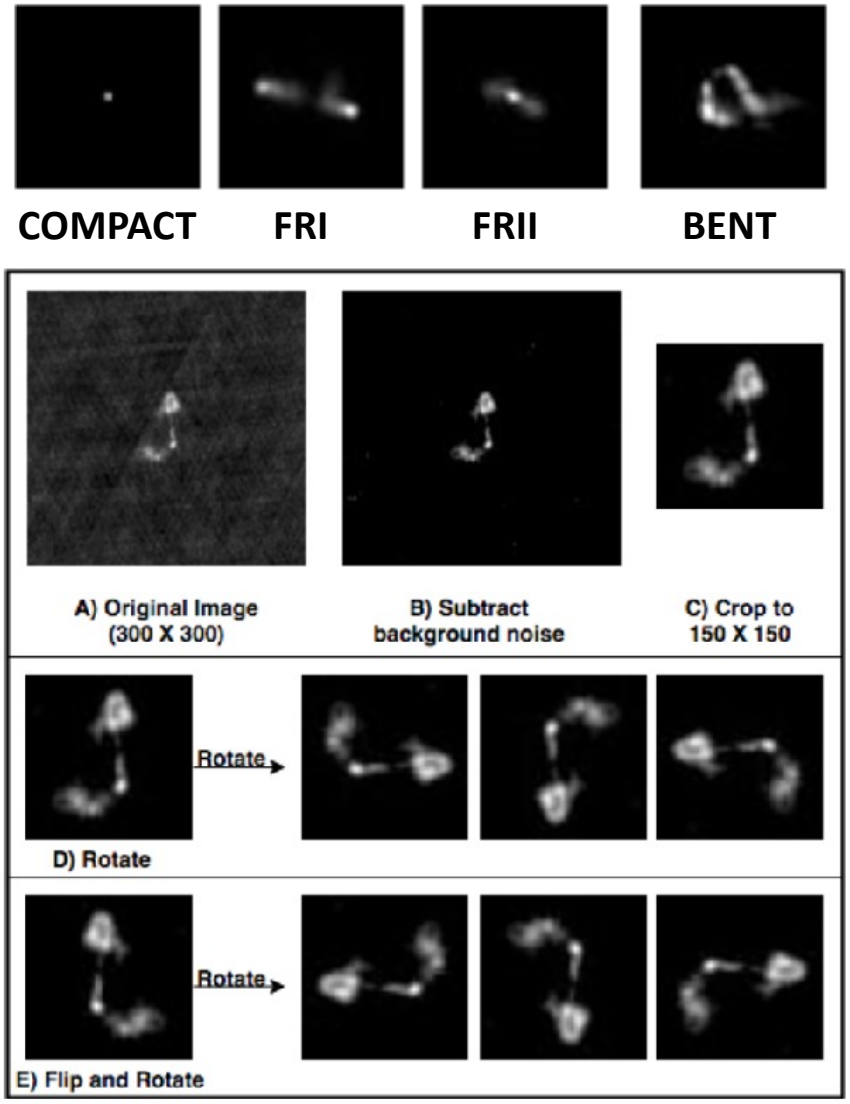


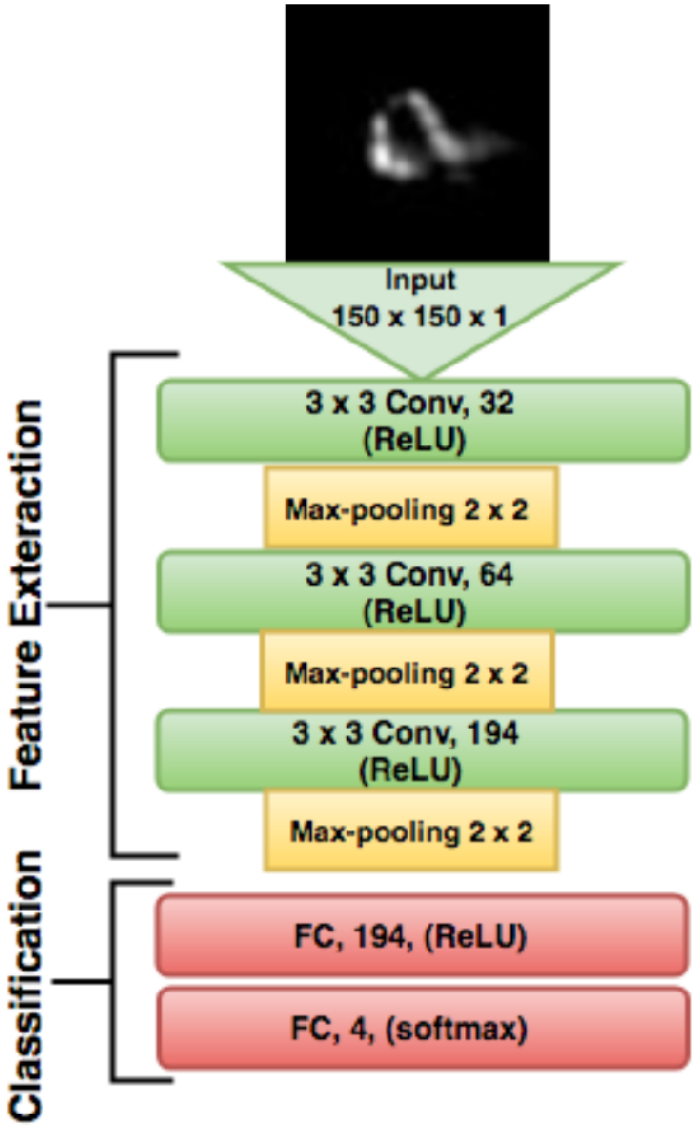
The SKA Data and Science Engine

Key technologies:
Network = connectivity
Cloud = access to data, infrastructure, software

Radio Source Morphological Classification

Alhassan, Taylor & Vaccari 2018





Collaborative Data Annotation

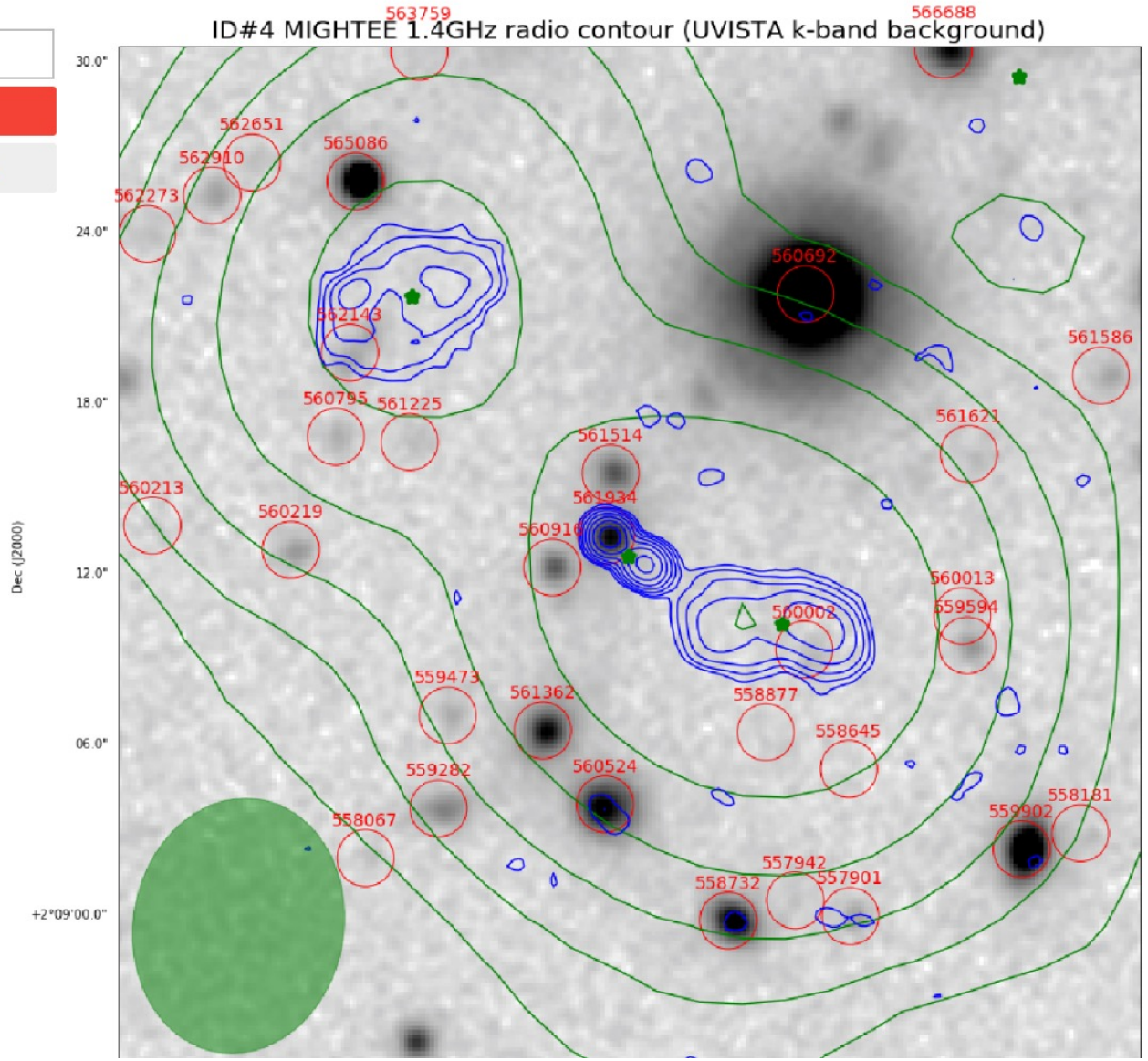
MATCH ID | Write IDs here!

GO | BACK

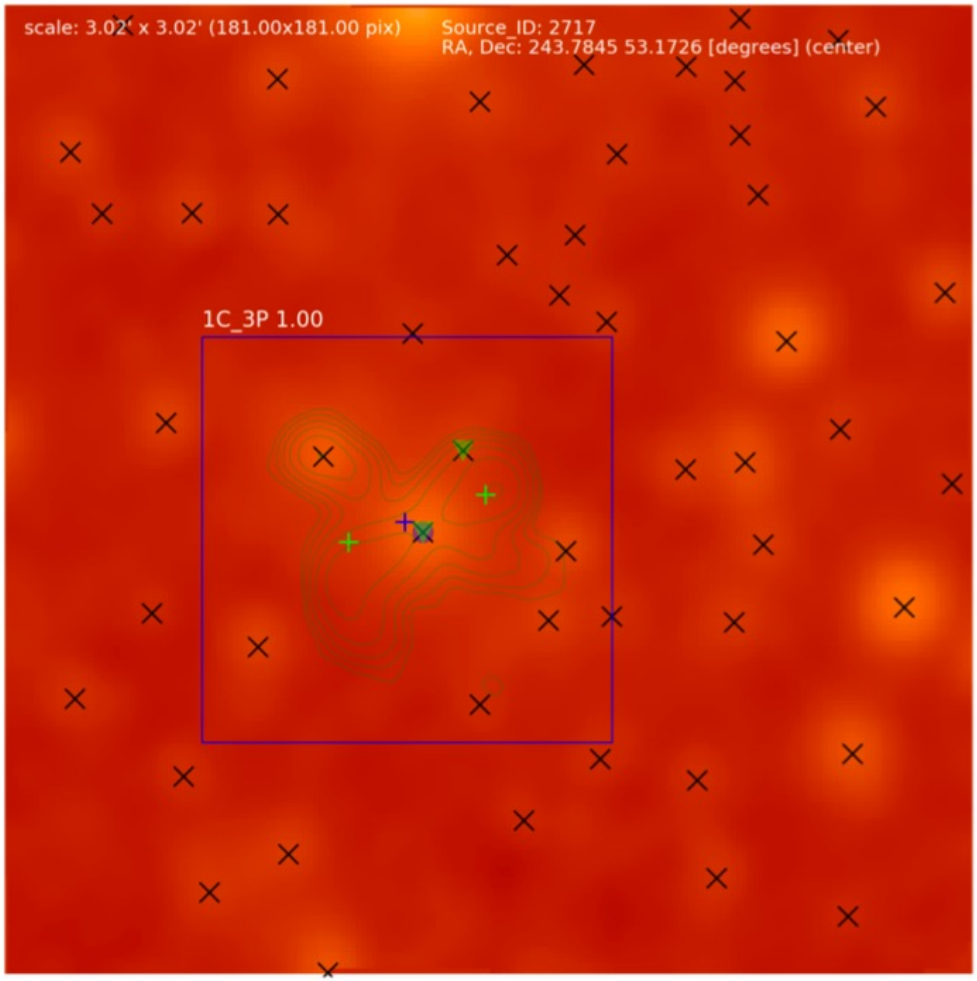
ZOOM OUT | ZOOM BACK

RANDOMISER!

Matt Prescott



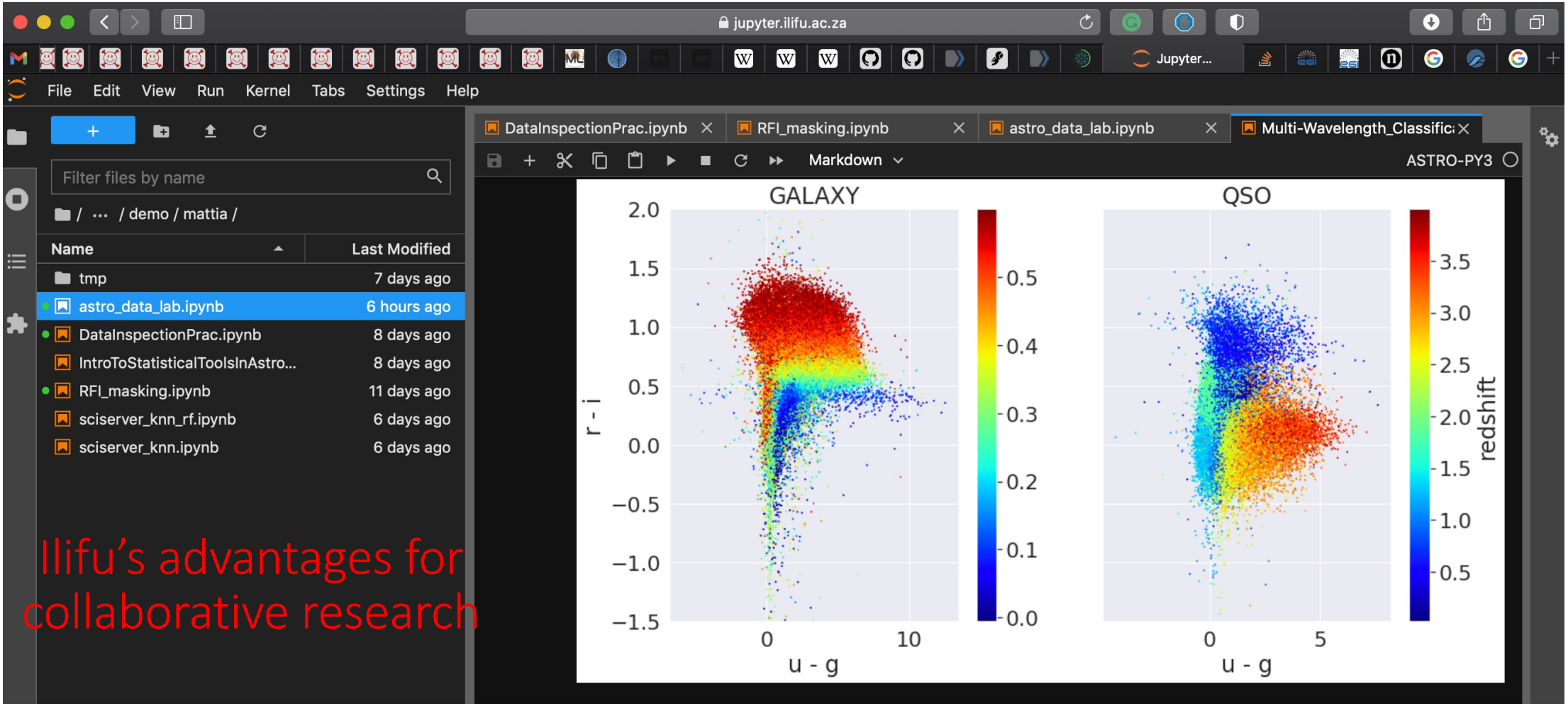
A Full Source Characterization Pipeline



Automate the characterisation (detection, identification and classification) of radio sources

MeerKAT's coverage and depth will allow us to create precious "training sets" to improve the source characterisation pipeline for EMU/VLASS/SKA

Chaka Mofokeng (MSc Thesis) - Applied to GMRT Data

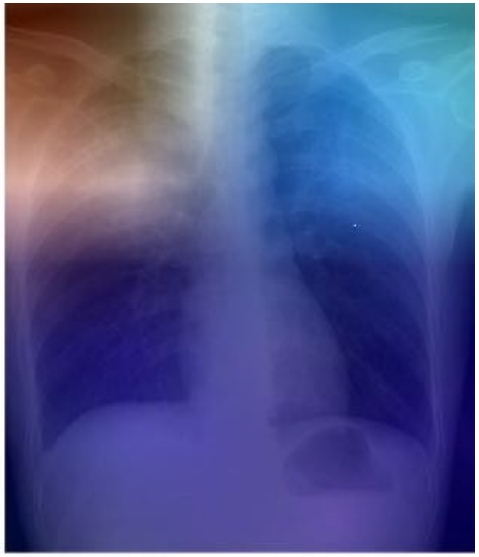


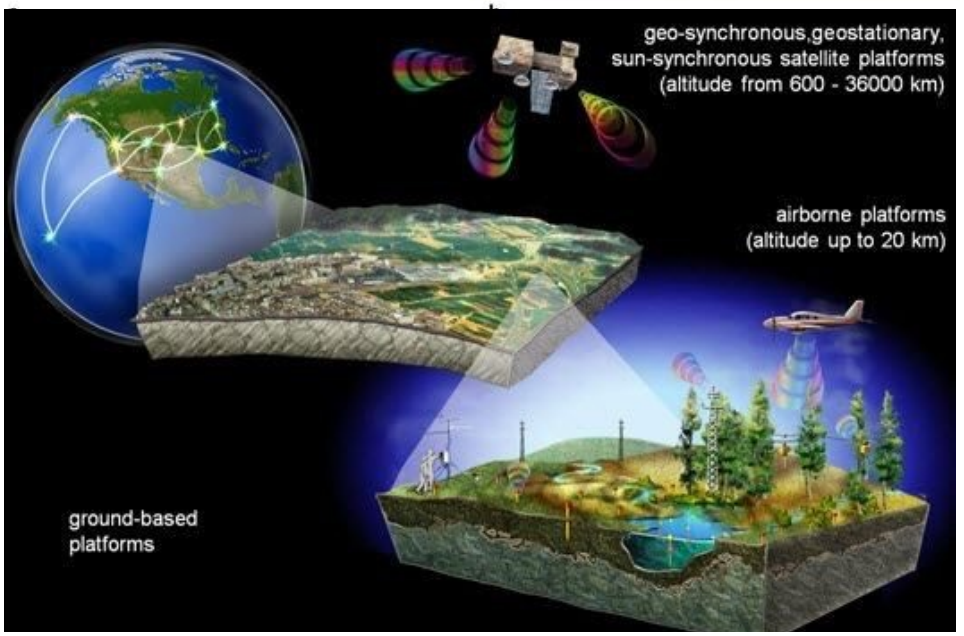
Ilifu's advantages for collaborative research

- (Big) Data Transfer, Storage, Processing and Visualization happens remotely
- Staff & students can collaborate on shared data, scripts and notebooks
- Software install is (mostly) managed by the ilifu support team, but users can also create their own software containers and/or virtual environments
- Allows a distributed research community to keep working in COVID times!

Ilifu : Enabling SA Science via Cloud Computing

ilifu





Summary

- ilifu is a custom cloud infrastructure developed in South Africa by a multi-disciplinary distributed university team
- Democratizes big data research by providing a flexible platform for interactive access to process, analyse, and visualize big data
- Serves a distributed community of researchers in MeerKAT and other SKA pathfinder key projects and South African bioinformatics
- Is providing capacity for training in data science and data intensive science
- Is working on technology for federation of clouds with global partners
- Can be the kernel to grow a South African and Pan-African federated research cloud with potential to transform data intensive research in Africa
- Provides Ample Opportunities for Collaboration in Cloud Infrastructure, Data Processing, Machine Learning, Visualization, Human Capacity Development