ilifu

The ILIFU Cloud & the HIPPO Project

Multi-Wavelength Astronomy in the Cloud

Mattia Vaccari
Inter-University Institute of Data Intensive Astronomy
University of the Western Cape
Cape Town, South Africa
ADASS, 7 October 2019

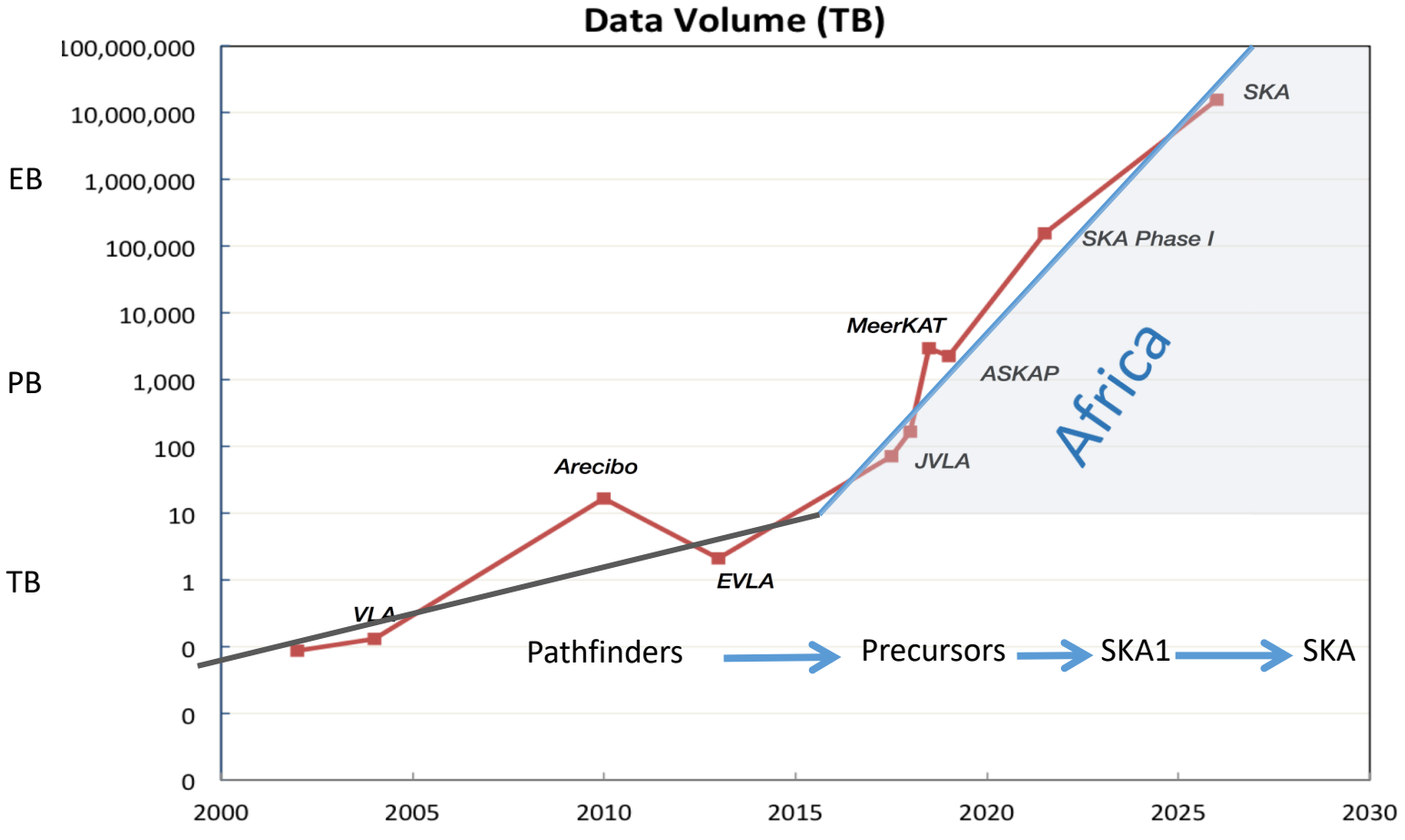
IDIA

RESPICE PROSPICE
UNIVERSITY of the WESTERN CAPE

MeerKAT : South Africa's SKA Precursor

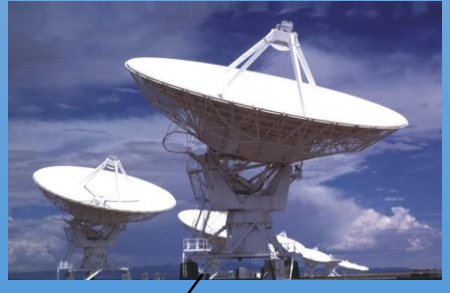
ilifu

Growth of Data Volumes to Radio Astronomers

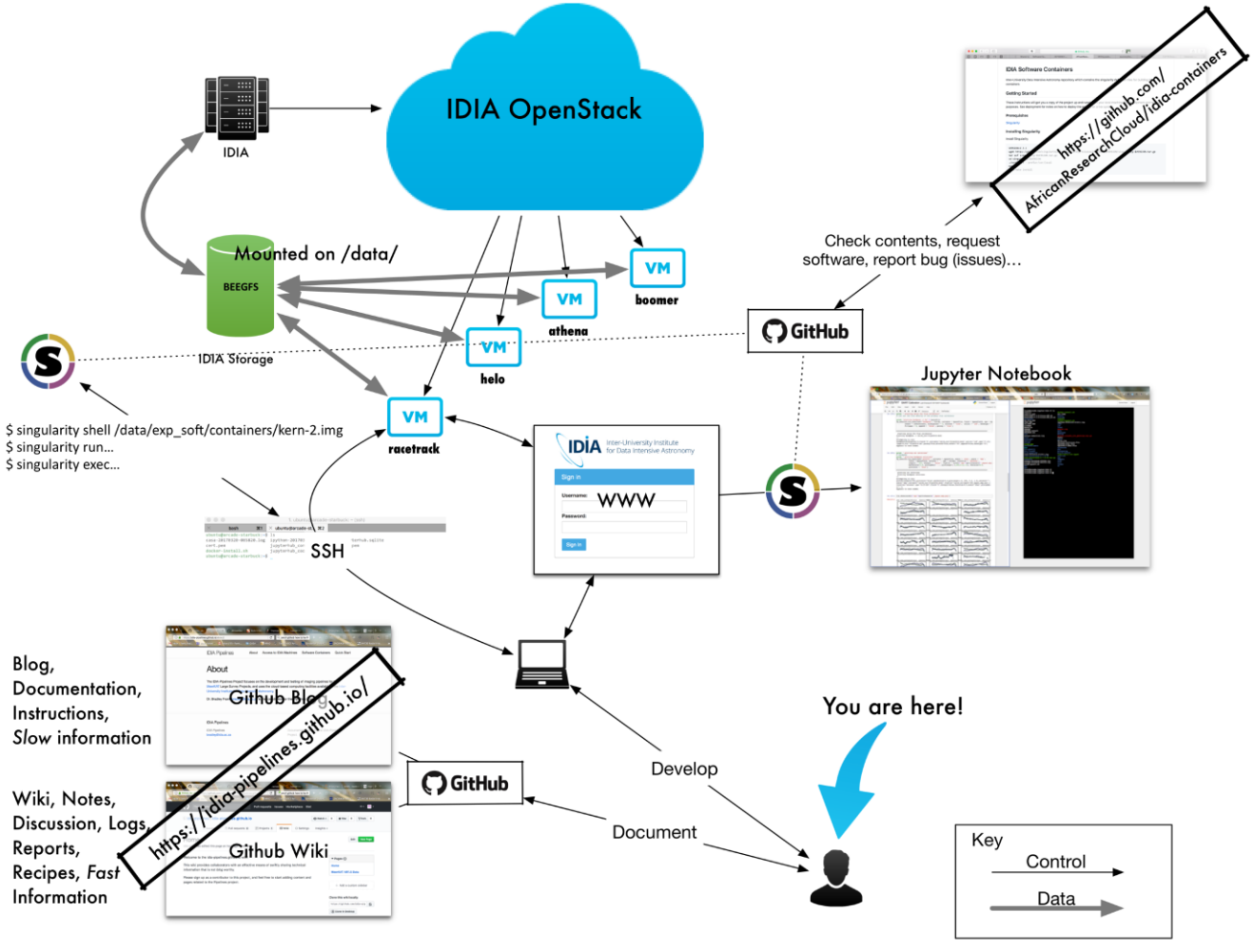


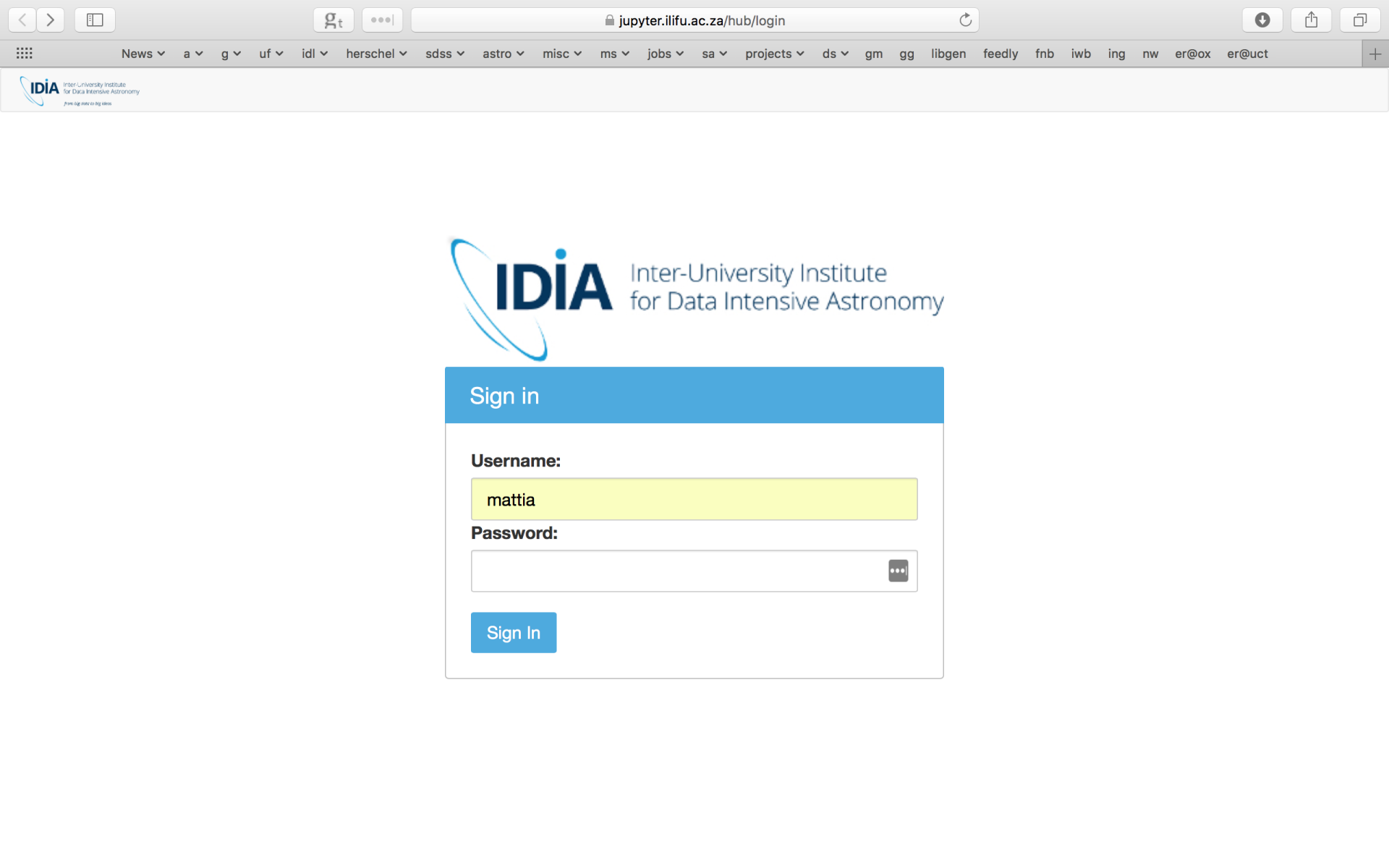
Changing Sociology of Radio Astronomy

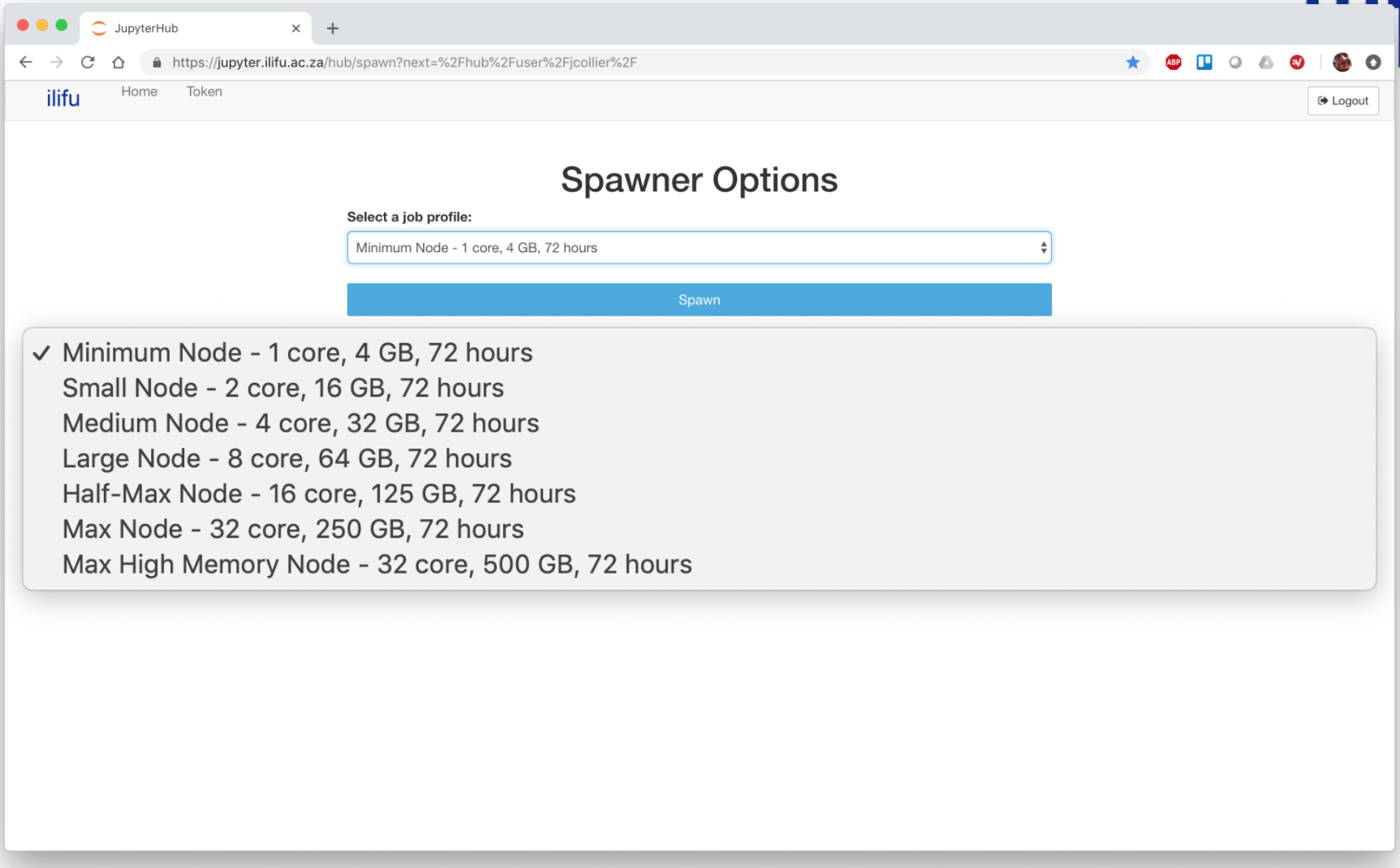
- Much of the key science en route to the SKA will be achieved via large-scale observing programs executed by globally distributed teams of researchers working on the data in a collaborative manner

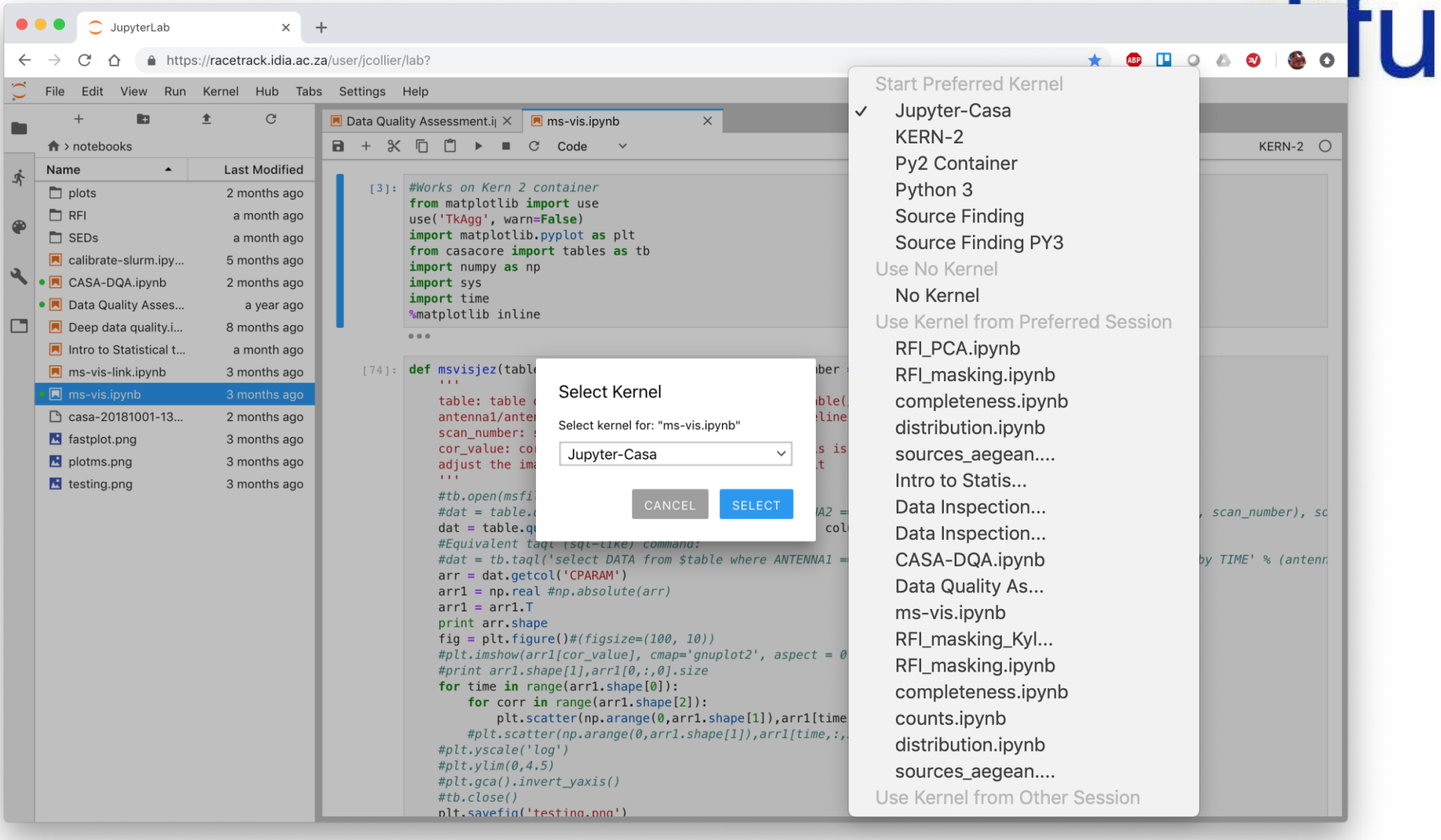


IDIA Data Intensive Astronomy Cloud (v1.0)

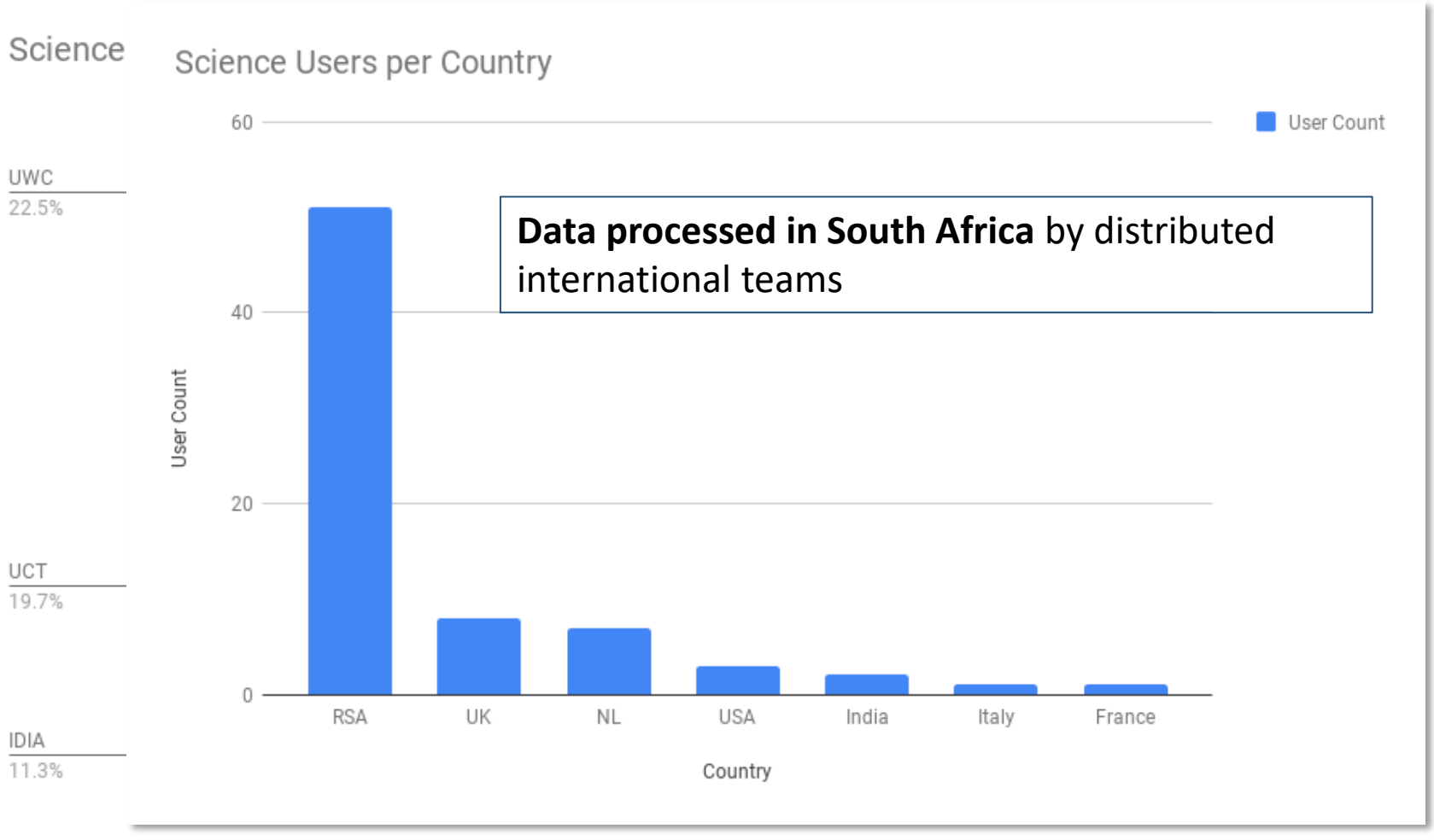



jupyter.ilifu.ac.za/hub/login
News a g uf idl herschel sdss astro misc ms jobs sa projects ds gm gg libgen feedly fnb iwb ing nw er@ox er@uct
IDIA Inter-University Institute for Data Intensive Astronomy
IDIA Inter-University Institute for Data Intensive Astronomy
Sign in
Username:
mattia
Password:
Sign In


JupyterHub
https://jupyter.ilifu.ac.za/hub/spawn?next=%2Fhub%2Fuser%2Fjcollier%2F
ilifu
Home
Token
Logout
Spawner Options
Select a job profile:
Minimum Node - 1 core, 4 GB, 72 hours
Spawn
Minimum Node - 1 core, 4 GB, 72 hours
Small Node - 2 core, 16 GB, 72 hours
Medium Node - 4 core, 32 GB, 72 hours
Large Node - 8 core, 64 GB, 72 hours
Half-Max Node - 16 core, 125 GB, 72 hours
Max Node - 32 core, 250 GB, 72 hours
Max High Memory Node - 32 core, 500 GB, 72 hours

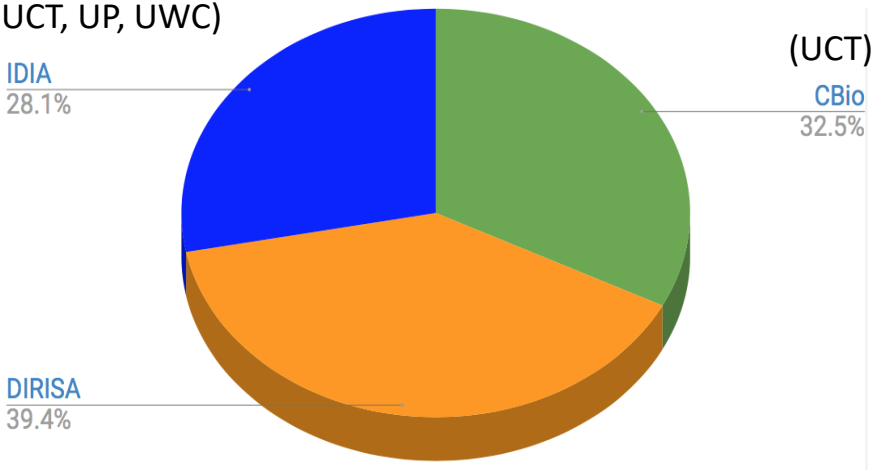

JupyterLab
https://racetrack.idia.ac.za/user/jcollier/lab?
File Edit View Run Kernel Hub Tabs Settings Help
notebooks
Name
Last Modified
plots 2 months ago
RFI a month ago
SEDs a month ago
calibrate-slurm.ipy... 5 months ago
CASA-DQA.ipynb 2 months ago
Data Quality Asses... a year ago
Deep data quality.i... 8 months ago
Intro to Statistical t... a month ago
ms-vis-link.ipynb 3 months ago
ms-vis.ipynb 3 months ago
casa-20181001-13... 2 months ago
fastplot.png 3 months ago
plotms.png 3 months ago
testing.png 3 months ago
Data Quality Assessment.i
ms-vis.ipynb
Code
KERN-2
[3]:
#Works on Kern 2 container
from matplotlib import use
use('TkAgg', warn=False)
import matplotlib.pyplot as plt
from casacore import tables as tb
import numpy as np
import sys
import time
%matplotlib inline
[74]:
def msvisjez(table
'''
table: table
antenna1/anter
scan_number:
cor_value: cor
adjust the ima
'''
#tb.open(msfi
#dat = table.
dat = table.qu
#Equivalent taql (sql-like) command:
#dat = tb.taql('select DATA from $table where ANTENNA1 =
arr = dat.getcol('CPARAM')
arr1 = np.real #np.absolute(arr)
arr1 = arr1.T
print arr.shape
fig = plt.figure()#(figsize=(100, 10))
#plt.imshow(arr1[cor_value], cmap='gnuplot2', aspect = 0
#print arr1.shape[1],arr1[0,:,0].size
for time in range(arr1.shape[0]):
for corr in range(arr1.shape[2]):
plt.scatter(np.arange(0,arr1.shape[1]),arr1[time
#plt.scatter(np.arange(0,arr1.shape[1]),arr1[time,:,
#plt.yscale('log')
#plt.ylim(0,4.5)
#plt.gca().invert_yaxis()
#tb.close()
plt.savefig('testing.png')
Select Kernel
Select kernel for: "ms-vis.ipynb"
Jupyter-Casa
CANCEL
SELECT
Start Preferred Kernel
Jupyter-Casa
KERN-2
Py2 Container
Python 3
Source Finding
Source Finding PY3
Use No Kernel
No Kernel
Use Kernel from Preferred Session
RFI_PCA.ipynb
RFI_masking.ipynb
completeness.ipynb
distribution.ipynb
sources_aegean....
Intro to Statis...
Data Inspection...
Data Inspection...
CASA-DQA.ipynb
Data Quality As...
ms-vis.ipynb
RFI_masking_Kyl...
RFI_masking.ipynb
completeness.ipynb
counts.ipynb
distribution.ipynb
sources_aegean....
Use Kernel from Other Session

IDIA Cloud MeerKAT Science Users



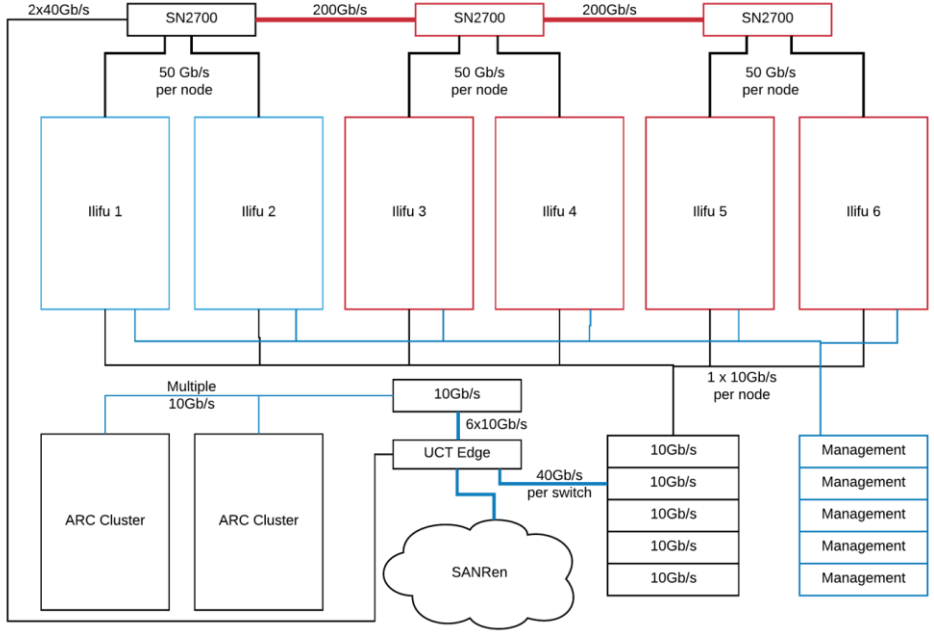
ILIFU Cloud Staged Roll out 2018-2020

(UCT, UP, UWC)



(UCT, UWC, CPUT, SU, SPU, SARAO)
AstroInformatics & BioInformatics

(UCT)



- 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 resources to be allocated to ilifu partners via an ongoing competitive process

MeerKAT Key Science Large Survey Projects

ilifu

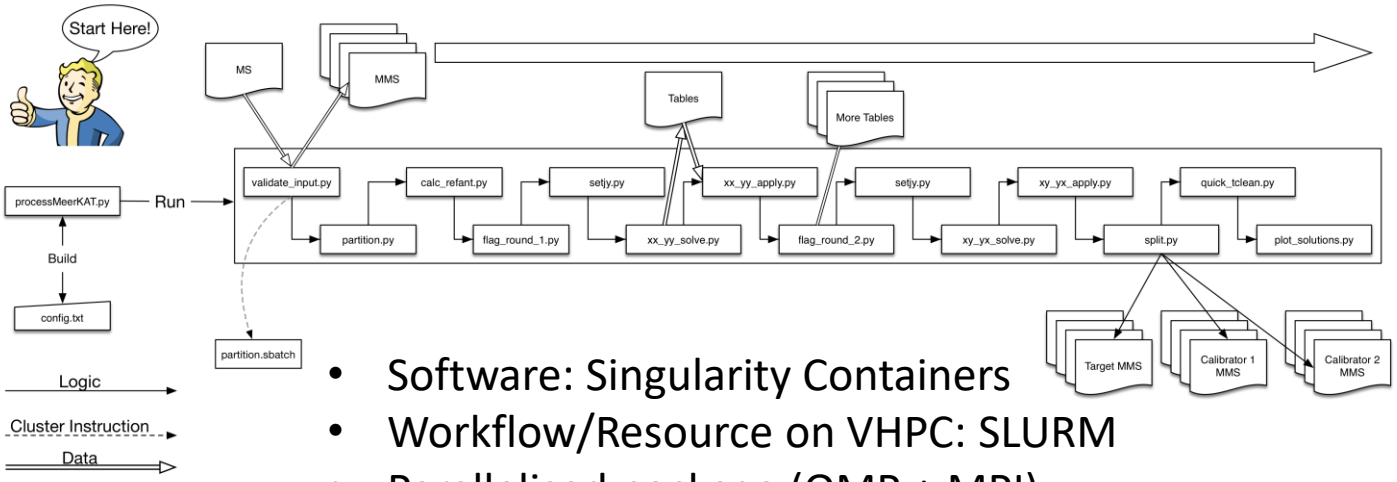
Imaging

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

http://public.ska.ac.za/meerkat/meerkat-large-survey-projects

The IDIA MeerKAT Data Reduction Pipeline

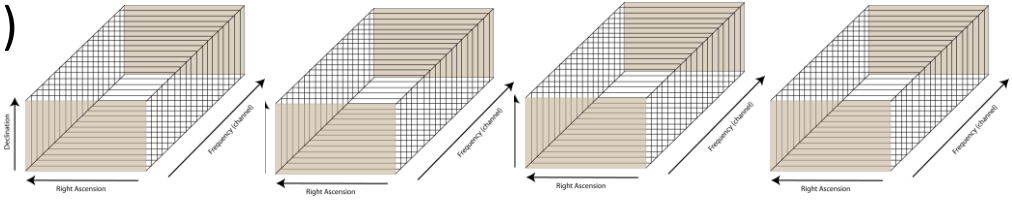


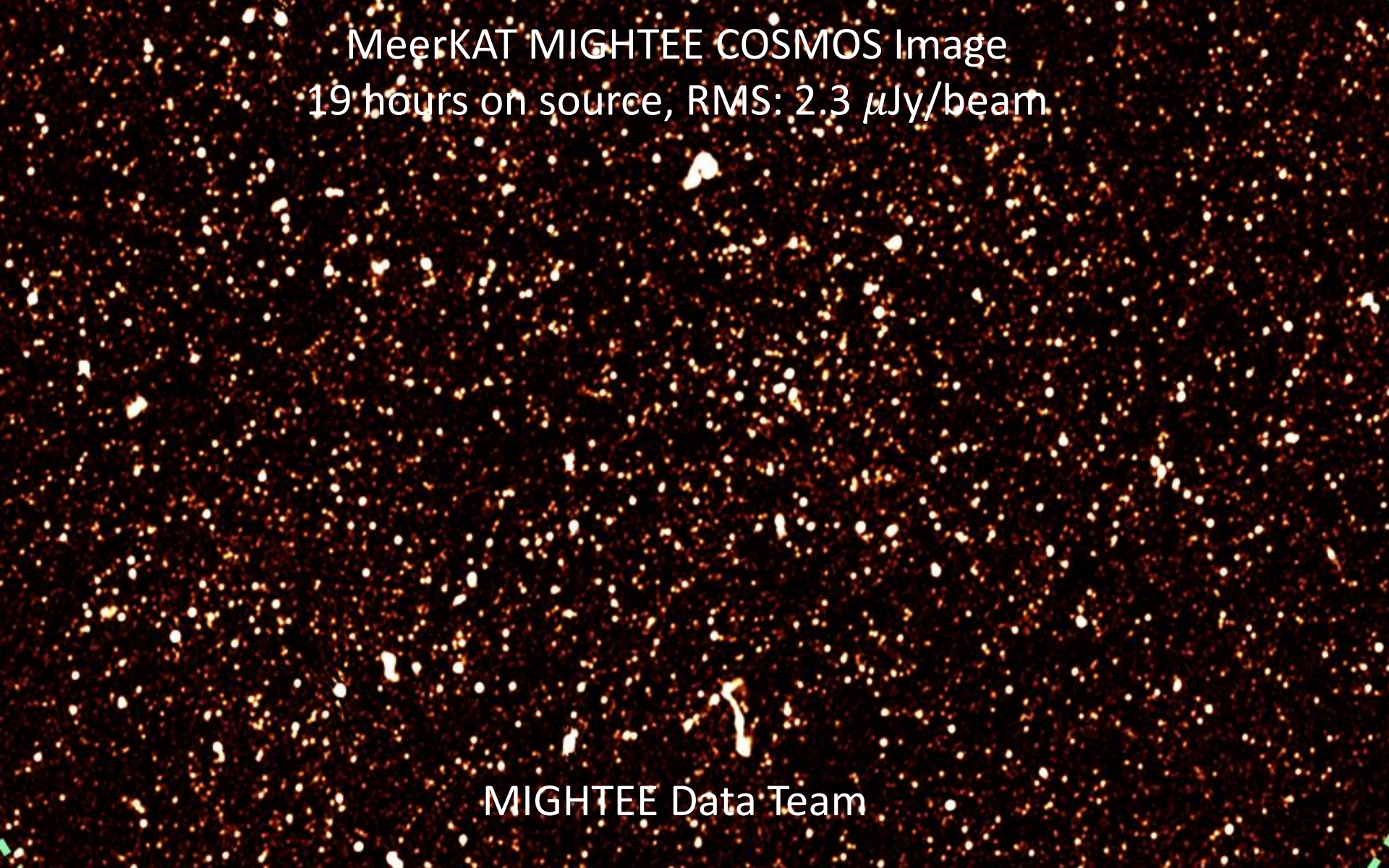
- Software: Singularity Containers
- Workflow/Resource on VHPC: SLURM
- Parallelised package (OMP + MPI)
- User configurable and executable

Data products

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

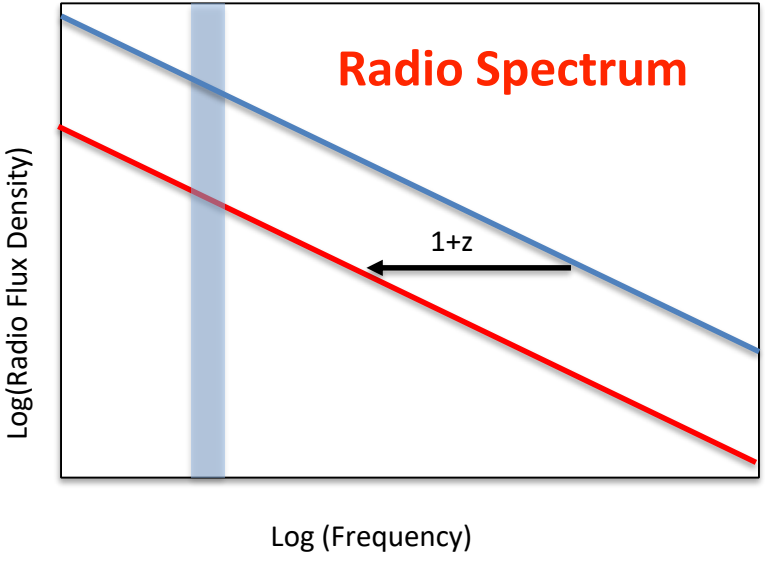
https://github.com/idia-astro/pipelines

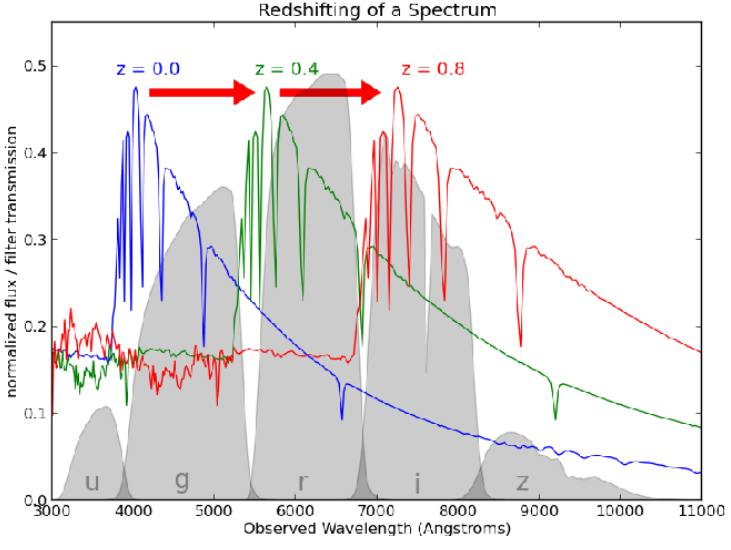



MeerKAT MIGHTEE COSMOS Image
19 hours on source, RMS: 2.3 μJy/beam
MIGHTEE Data Team

But of course…

There's nothing as useless as a radio source (Jim Condon)

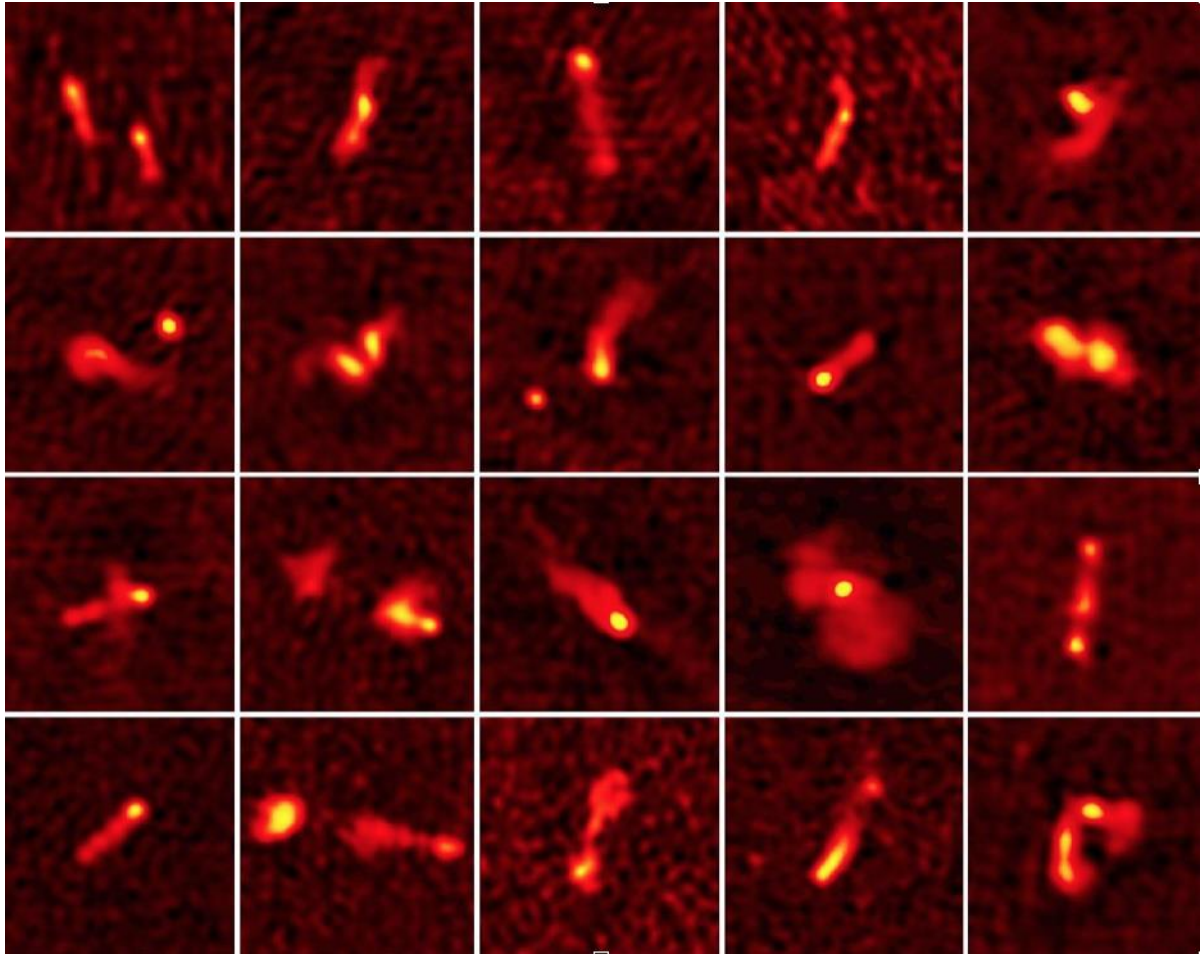




Optical Spectrum

- Radio provides no (or very little?) distance information
- Optical photometry provides much stronger constraints

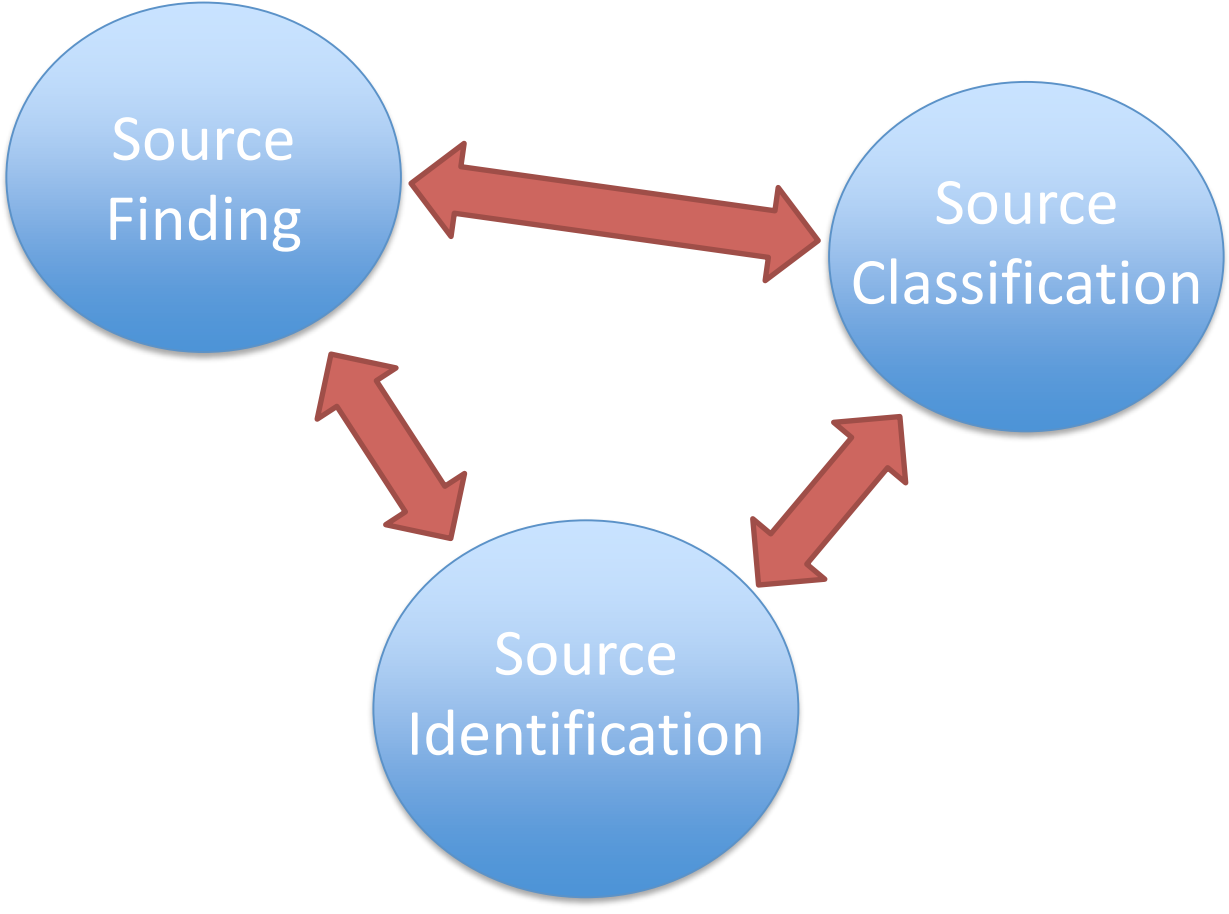
Radio Galaxies



VLSS - NRAO

Radio Source Characterisation

ilifu



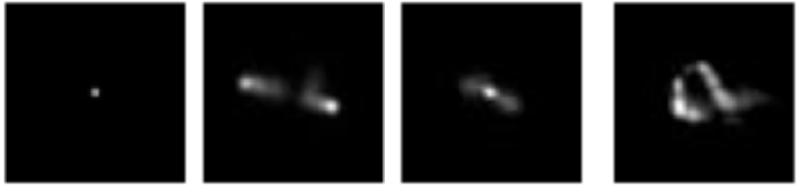
Multi-Wavelength Astronomy in the Cloud

ilifu

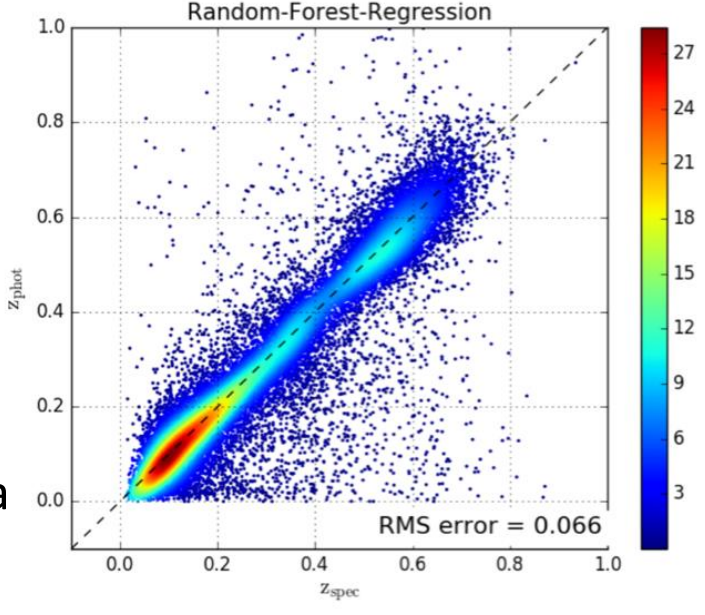
HIPPO : The HELP - IDIA Panchromatic PrOject

http://www.mattiavaccari.net/hippo

A Cloud-Based Environment for the Science Exploitation of Radio Surveys

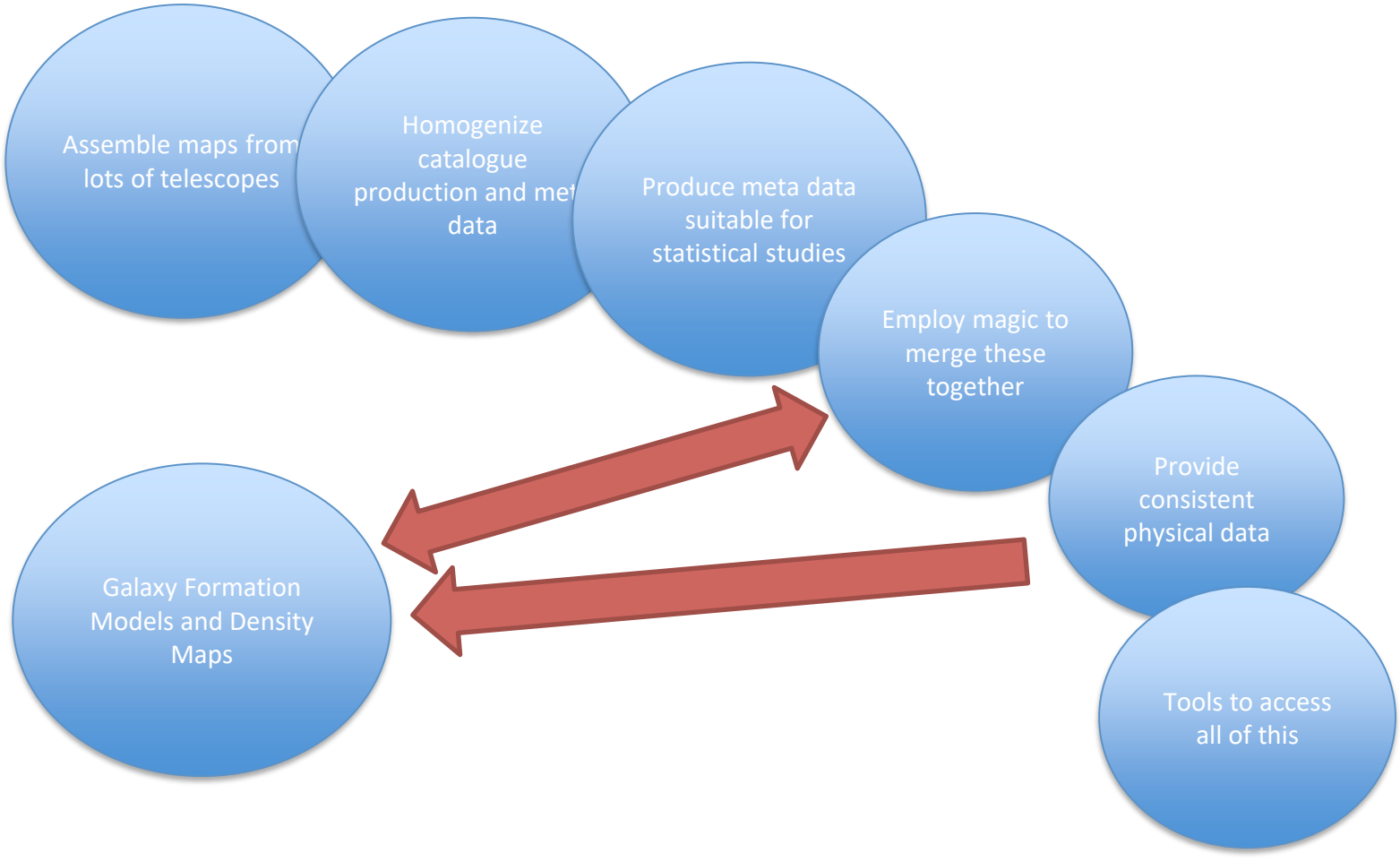


Working with IDIA programmers to create a cloud-based environment where scientists can exploit MeerKAT in the context of multi-wavelength data



HELP (EC-REA-FP7-SPACE) Concept

Vaccari 2016



HIPPO's First Steps

ilifu

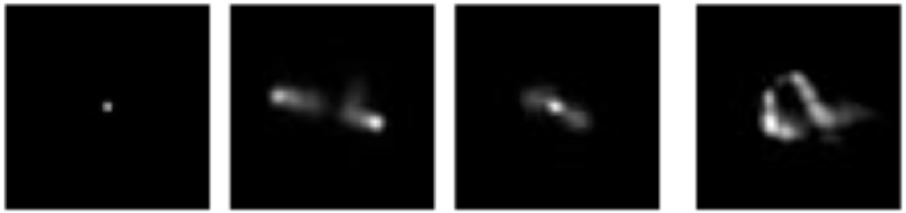
- Create a python-centric 'Software Container' for 'Source Characterization' on the ILIFU Cloud
- Assemble & Homogenize software tools to create cutouts and contours/overlays from most surveys
- Simple Visualization and Annotation Software
- Source Morphological Classification Software
- Source Spectro-Photometric Classification Software
- Extend Multi-Wavelength Ancillary Data (post-HELP)

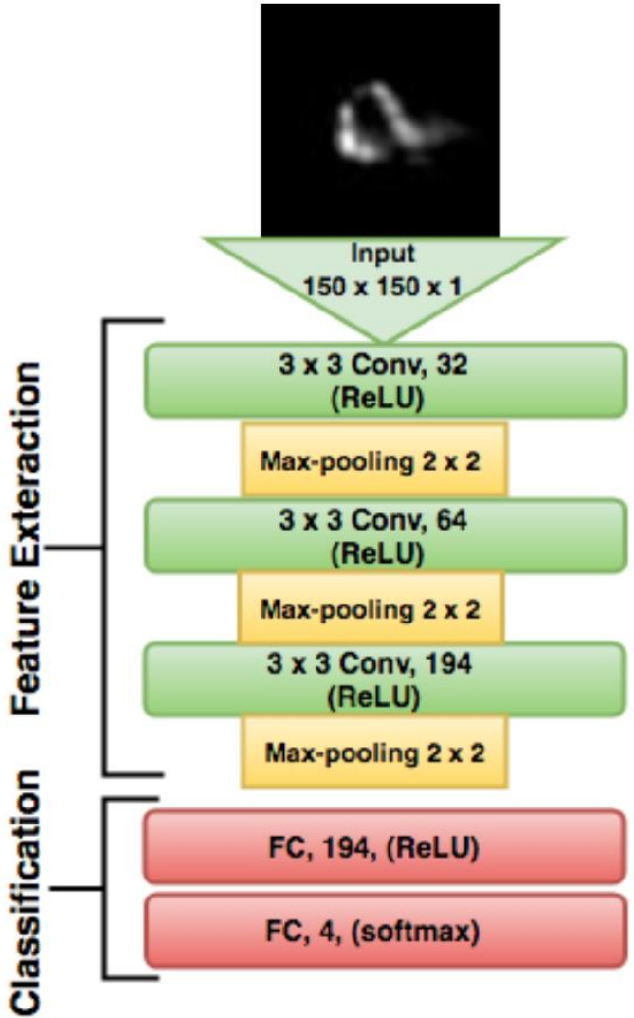
Deep Learning for Radio Source Classification

ilifu

Alhassan, Taylor & Vaccari 2018
https://github.com/wathela/FIRST-CLASSIFIER

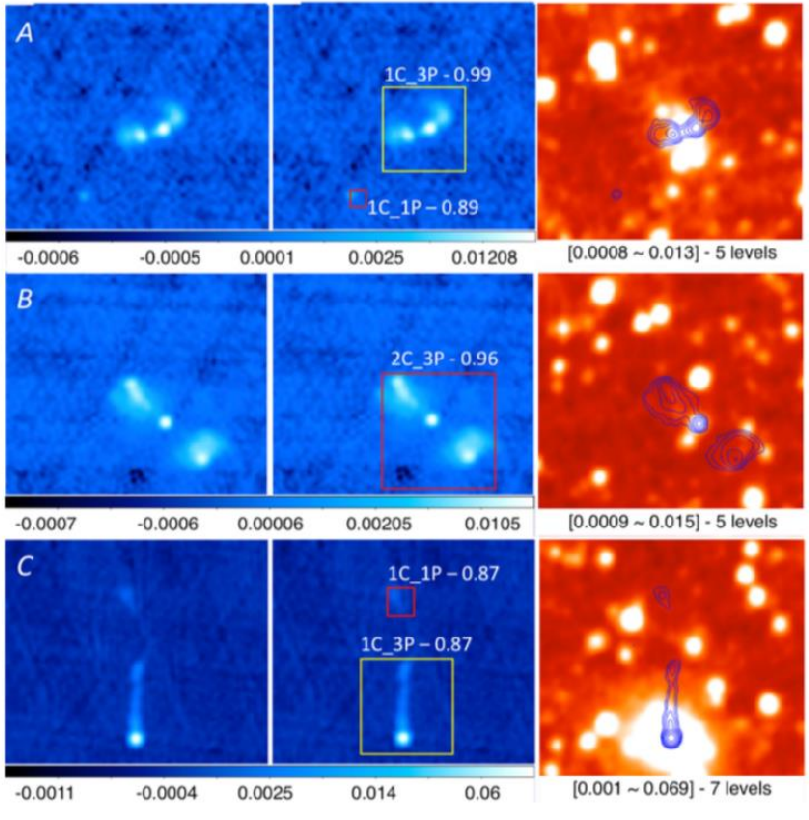
Type	Original Sample
COMP	121
FRI	201
FRII	338
BENT	177
Total	837





Faster Region-Based CNNs with ClaRAN for the Classification of Radio Sources

ilifu



Wu et al. 2019 - https://github.com/chenwuperth/rgz_rcnn/

Chaka Mofokeng (MSc) - Applied to GMRT(/MeerKAT) Data

MIGHTEE Early Work

ilifu

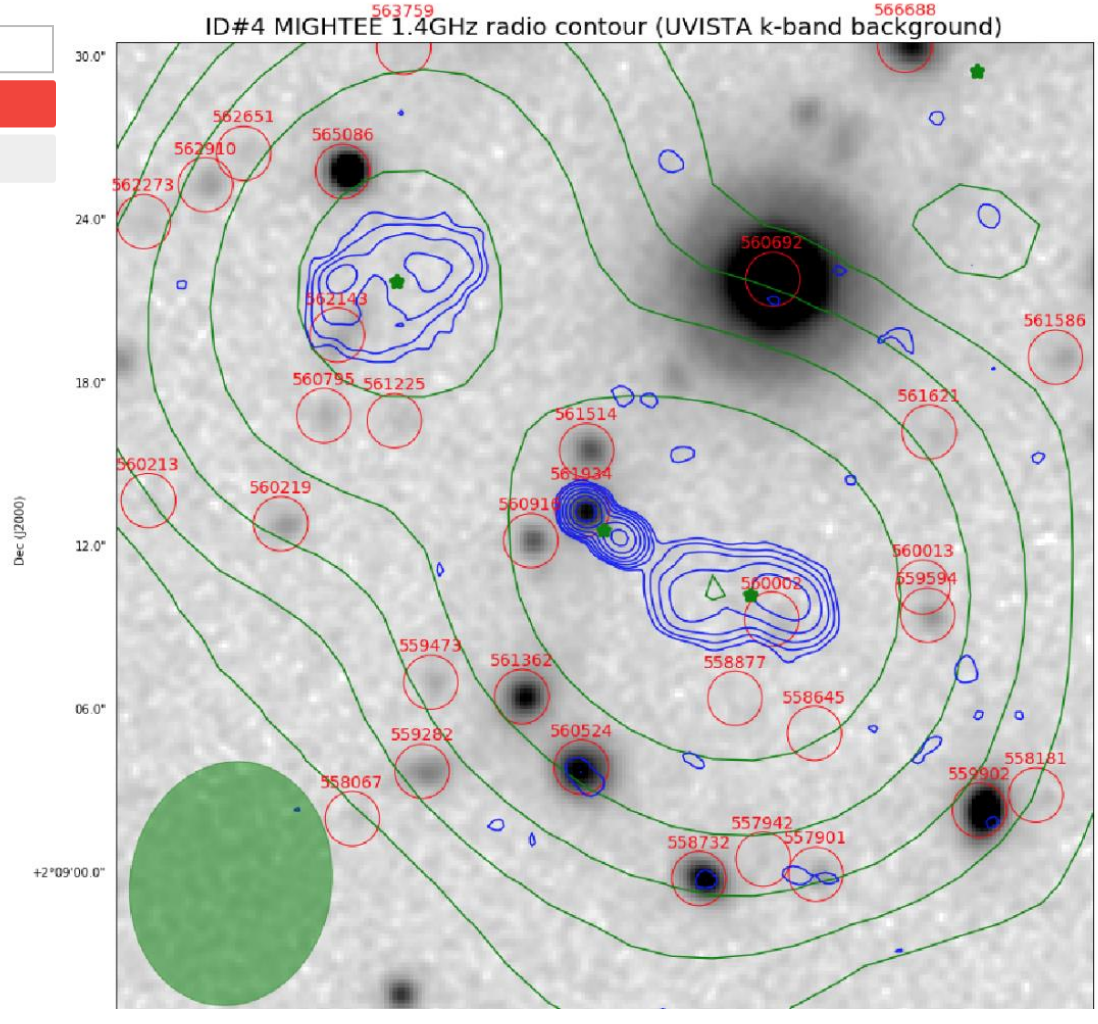
MATCH ID | Write IDs here!

GO | BACK

ZOOM OUT | ZOOM BACK

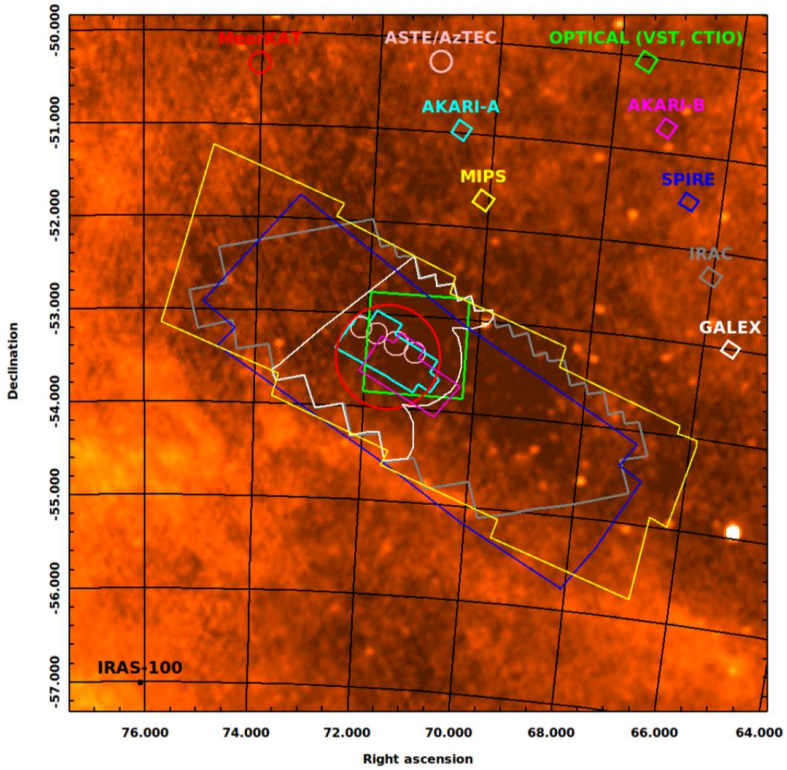
RANDOMISER!

Matt Prescott



ADFS MeerKAT/MeerLICHT Program

ilifu



http://www.mattiavaccari.net/adfs/

Multi-Epoch Simultaneous Optical & Radio Observations of a deep extragalactic field

LSST/SKA science pathfinder

Part of push to turn SAAO into "An Intelligent Observatory" in the multi-messenger era

Summary

- ILIFU is a modern, 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 by university researchers and students to process, analyse, and visualize big data
- Serves a distributed community of researchers in SKA pathfinder key projects and South African bioinformatics
- Can be the kernel to grow a South African and Pan-African federated research cloud with potential to transform data intensive research in Africa
- HIPPO is developing tools to use ILIFU for multi-wavelength astronomy, focusing on the challenges of source characterization and transient search
- Opportunity to build upon HIPPO@ILIFU to better support optical/infrared data processing pipelines and multi-messenger transient astronomy