

South Africa: SKA Regional Centre Activity

Dr. Rob Simmonds
Associate Director
IDIA



Overview

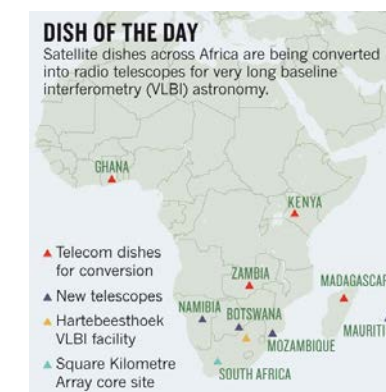
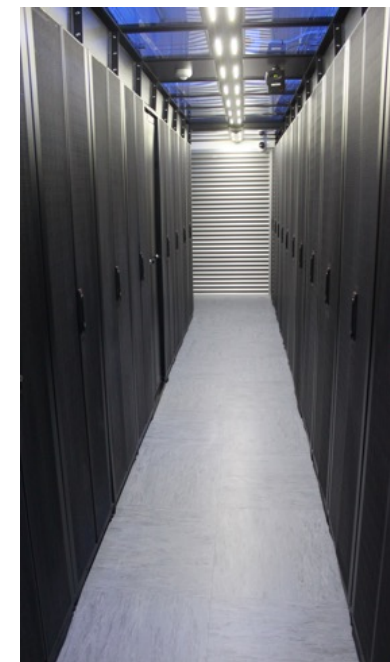


- MeerKAT RC / Tier 2 facilities
- SDP DELIV
 - Data delivery architecture for the SKA
- Data distribution and data processing for MeerKAT data
 - Collaboration between IDIA, ASTRON and SKA-SA
- SKA-SA activities
 - MeerKAT Commissioning
 - Archiving MeerKAT data
 - Pipelines / Machine learning
- CyberSKA portal and visualization activities

ARC / IDIA / Tier 2 systems



- Hardware deployed at UCT and NWU to act as MeerKAT Tier 2 system
- Currently 48 compute nodes (1500 cores) / 1.5PB storage
 - Includes 8 P100 GPUs for experimentation
 - 50Gb/s Open Ethernet core network
 - Managed using OpenStack IaaS framework
- RFP for additional compute and storage about to be issued
- Additional systems being deployed at UWC, UP and Wits will be used for distributed prototyping
- Exploring expansion of ARC to AVN / SKA partner countries



IDIa storage



- Using CEPH for cloud storage
 - Ephemeral storage to VMs
 - Block storage for creating small / mid-sized file system volumes
 - Object storage for scale out applications
- CEPH 3x replication but could reduced duplication using software erasure coding on current systems (on SAS backed volumes)
- Have BeeGFS for large data volumes
 - Higher performance (+)
 - Less data replication (using RAID 6) (+)
 - System level authorization (-)
 - Provide home, data and scratch BeeGFS volumes to trusted OpenStack hosted VMs

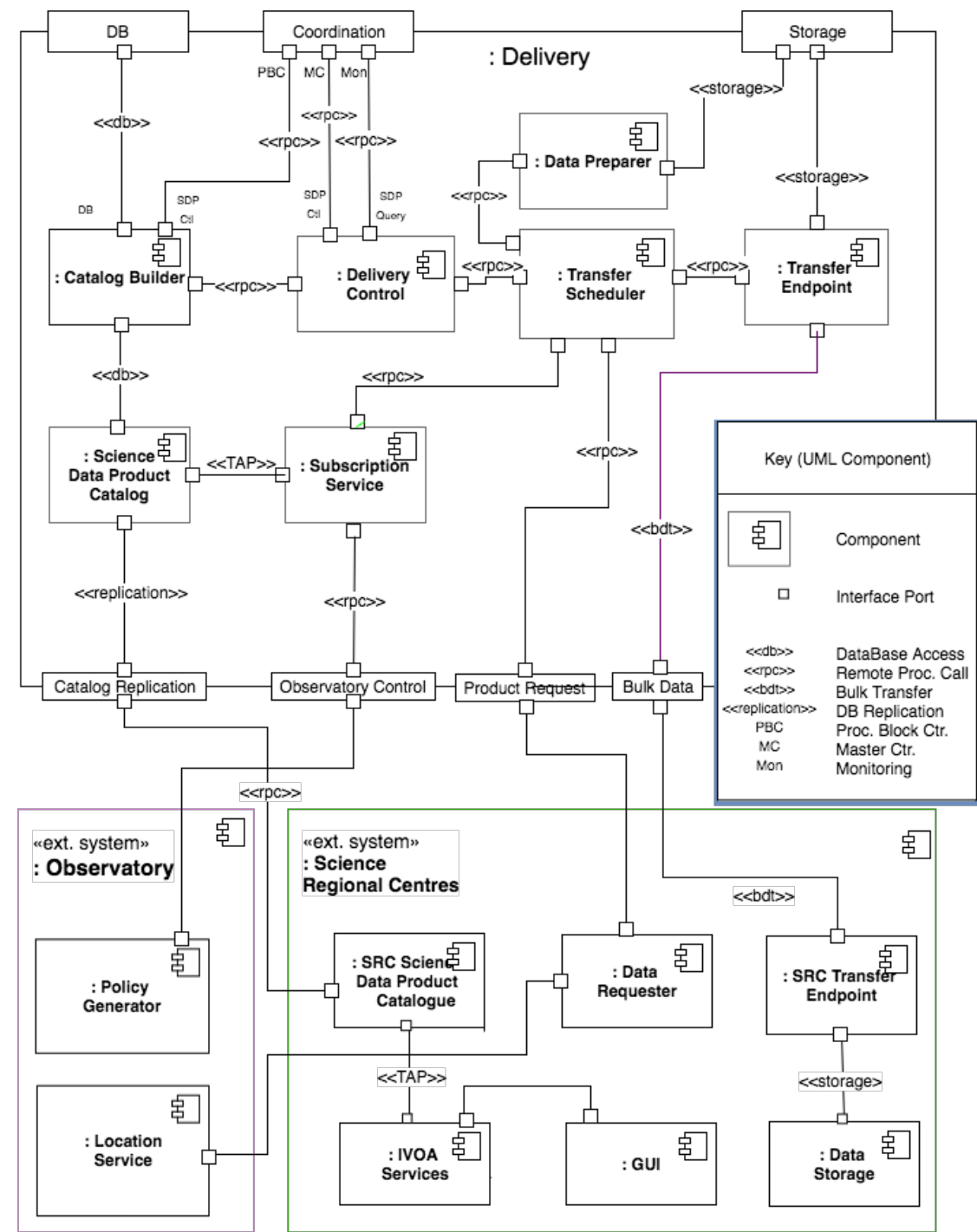
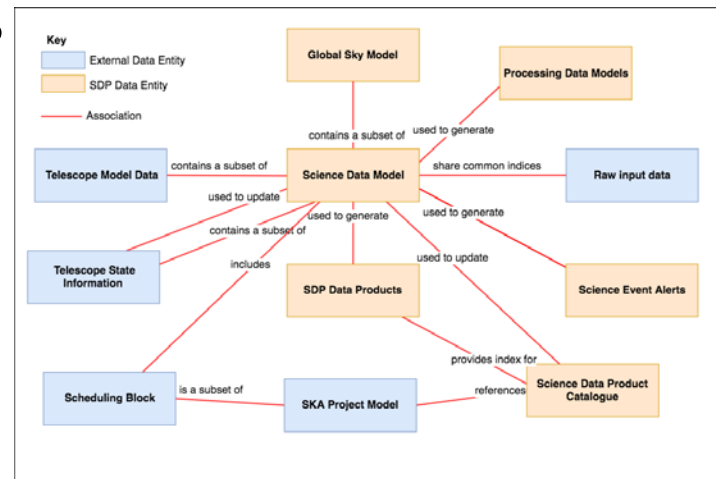
IDIA POSIX storage access



- Access to shared POSIX volumes from trusted machines
 - I.e., users don't have sudo access on these machines
- Users execute applications Singularity containers
 - Repository of application containers being created
- Users can create their own containers on non-trusted VMs
 - These can access shared POSIX volumes using SSH-FS
- Note: moved from Docker to Singularity containers due to concerns about privilege escalation

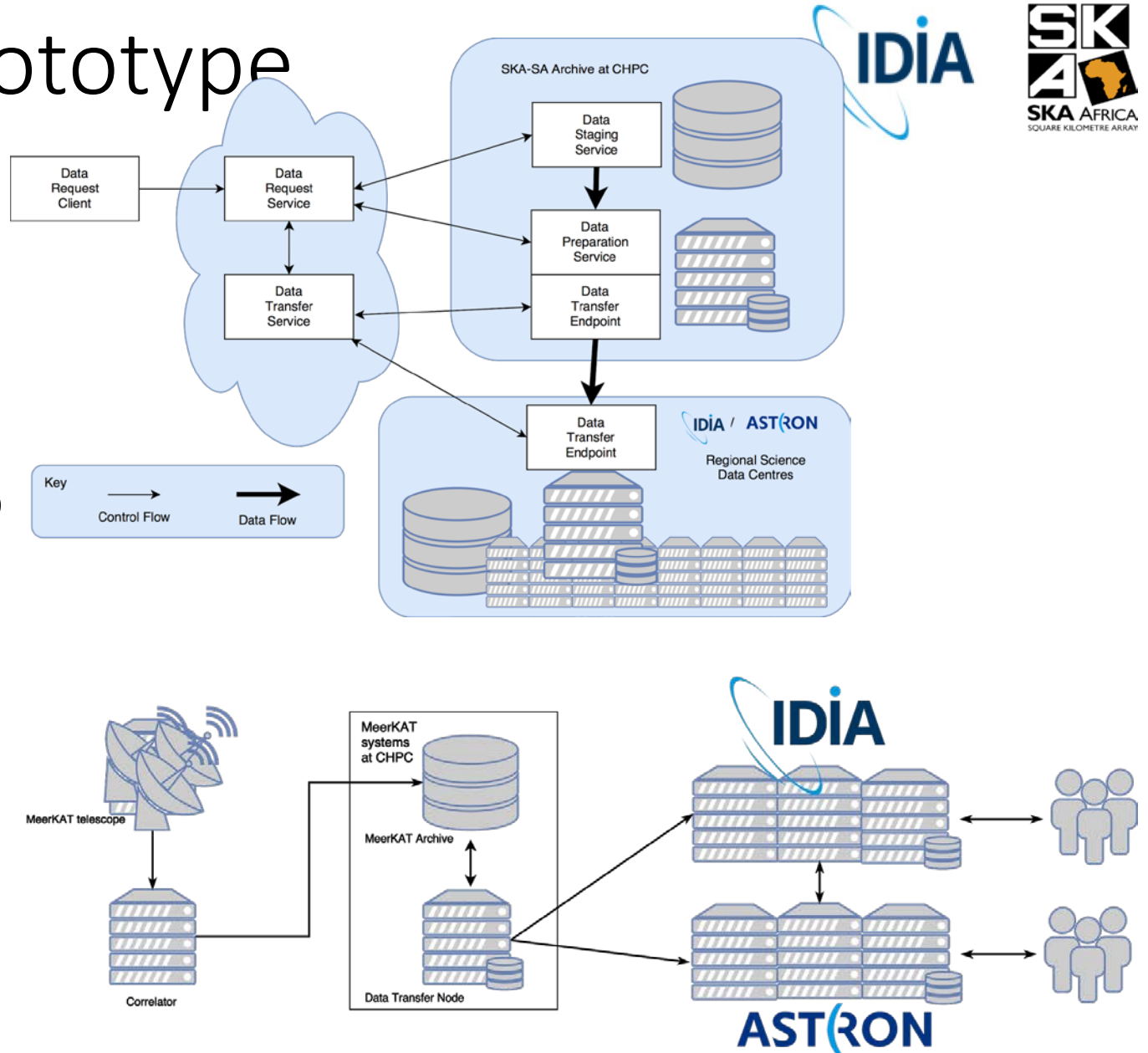
SDP Delivery Architecture

- Refining the delivery architecture to fit with revised SDP architecture
- Includes the SDP SRC interface
- Partners from UCT, CADC, SKA-SA/SAC, ASTRON, IAA and Oxford
- Still needs some thought about the what is needed in Science Data Product Catalog at SRCs



Data transport prototype

- Deployed as part of DOME project
- Utilizes CERN's FTS to control network transfers
- Queuing system used to stage data from Archive and to keep network loaded with high priority transfers
- Implements Request, Prepare and Deliver components of SDP architecture
- Deployed at SKA-SA/CHPC, IDIA and ASTRON



IVOA service prototyping



- Have access to sandbox services running at CADC
- SKA-SA providing access to MeerKAT metadata
- Mapping this metadata to COAM2 data model for use with CADC services
- Will provide prototype for SDP Query service once data modelling is complete

Pipelines



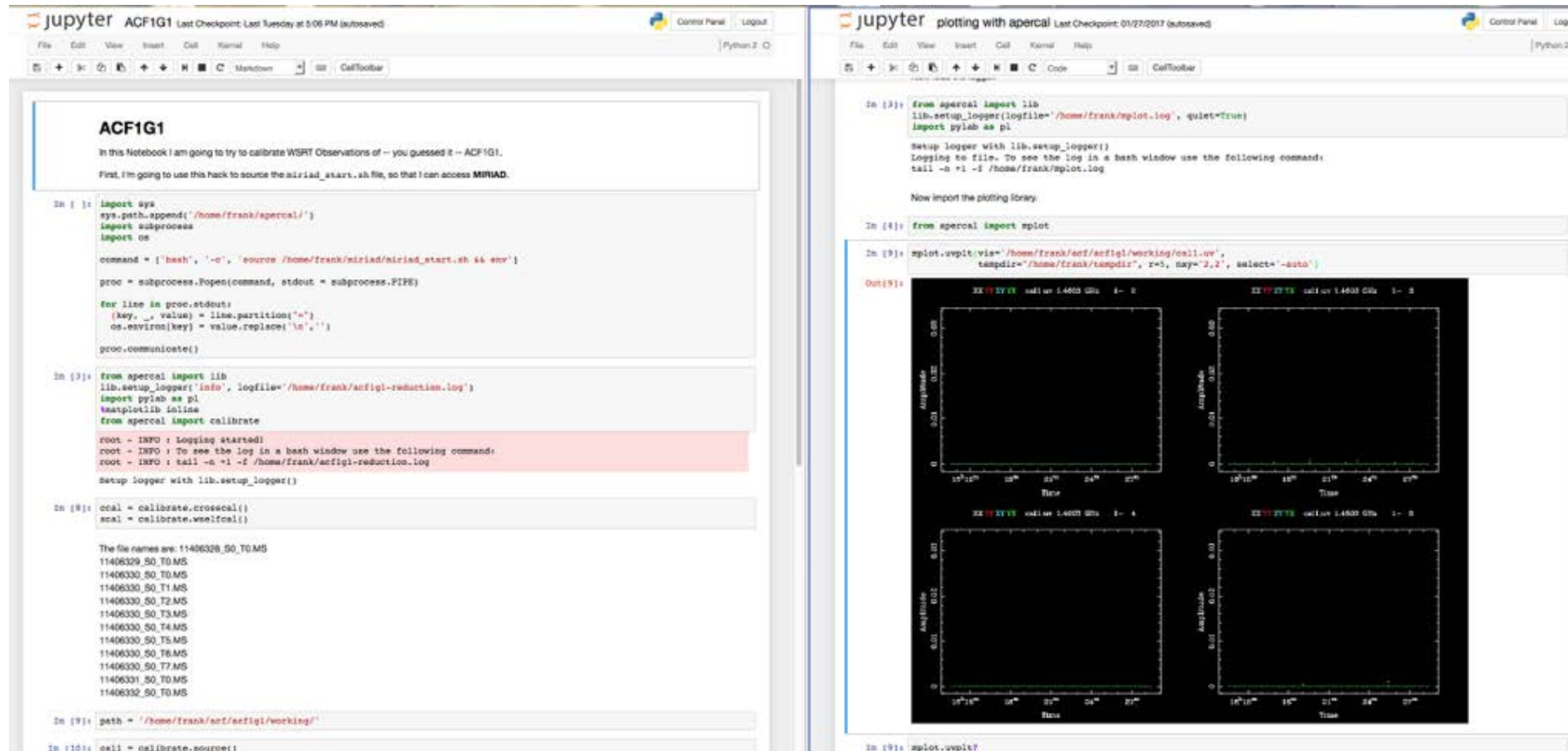
- Singularity container with:
 - CASA, drive-casa, python libraries, Jupyter Notebook & Hub.
- Continuing to develop software framework for running pipelines in Notebook
- Adopted code from JIVE project that was more robust than our initial prototype
- Being combined with data transport system to provide automated archive to product execution
- Adding DAGman/HTCondor and automated interaction with Git to produce reproducible workflows

Current Notebook Dashboard on ARC

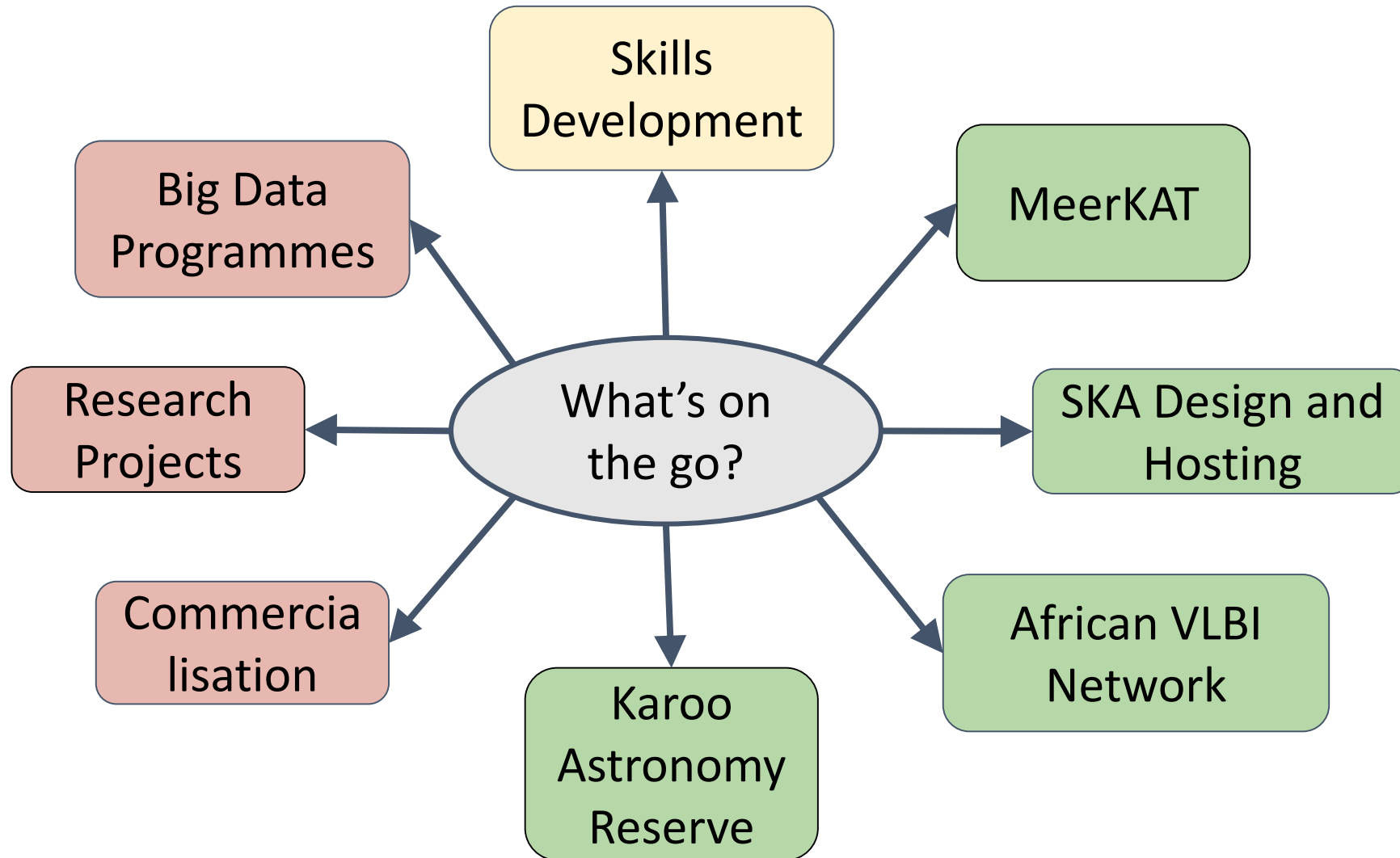


Pipeline Control &
Data Exploration

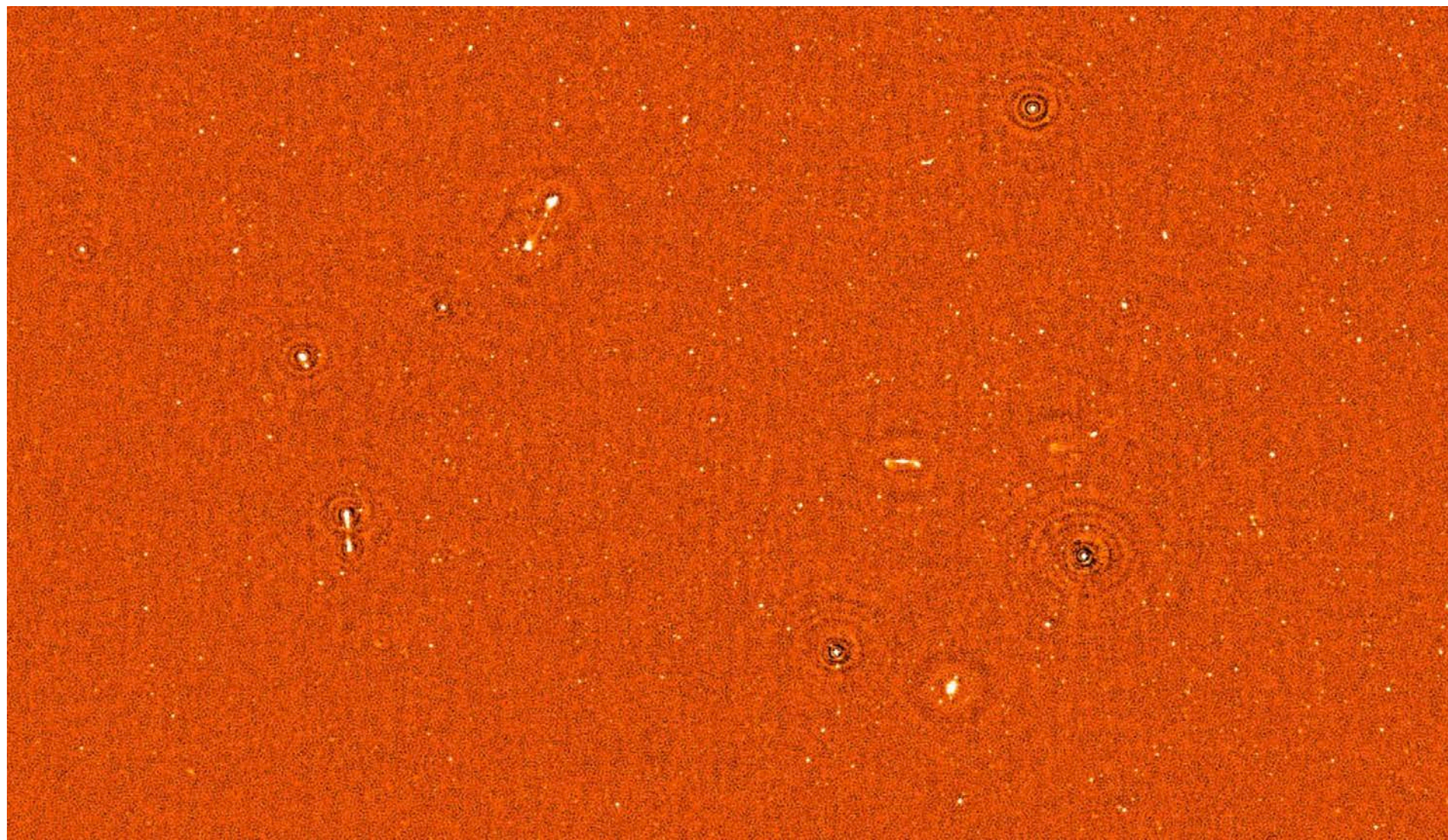
Visualization



SKA SA - Programmes

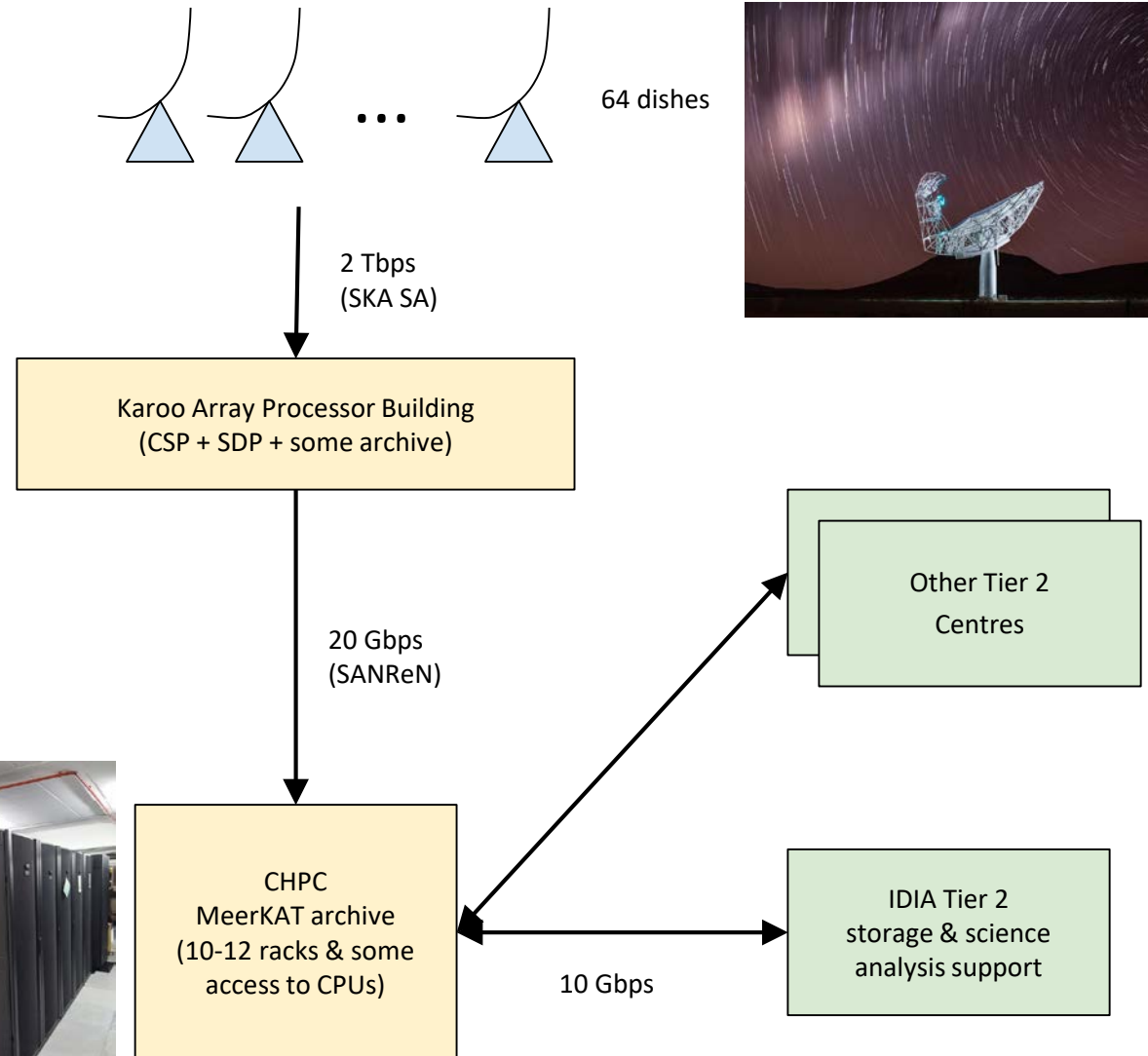


MeerKAT data processed at IDIA

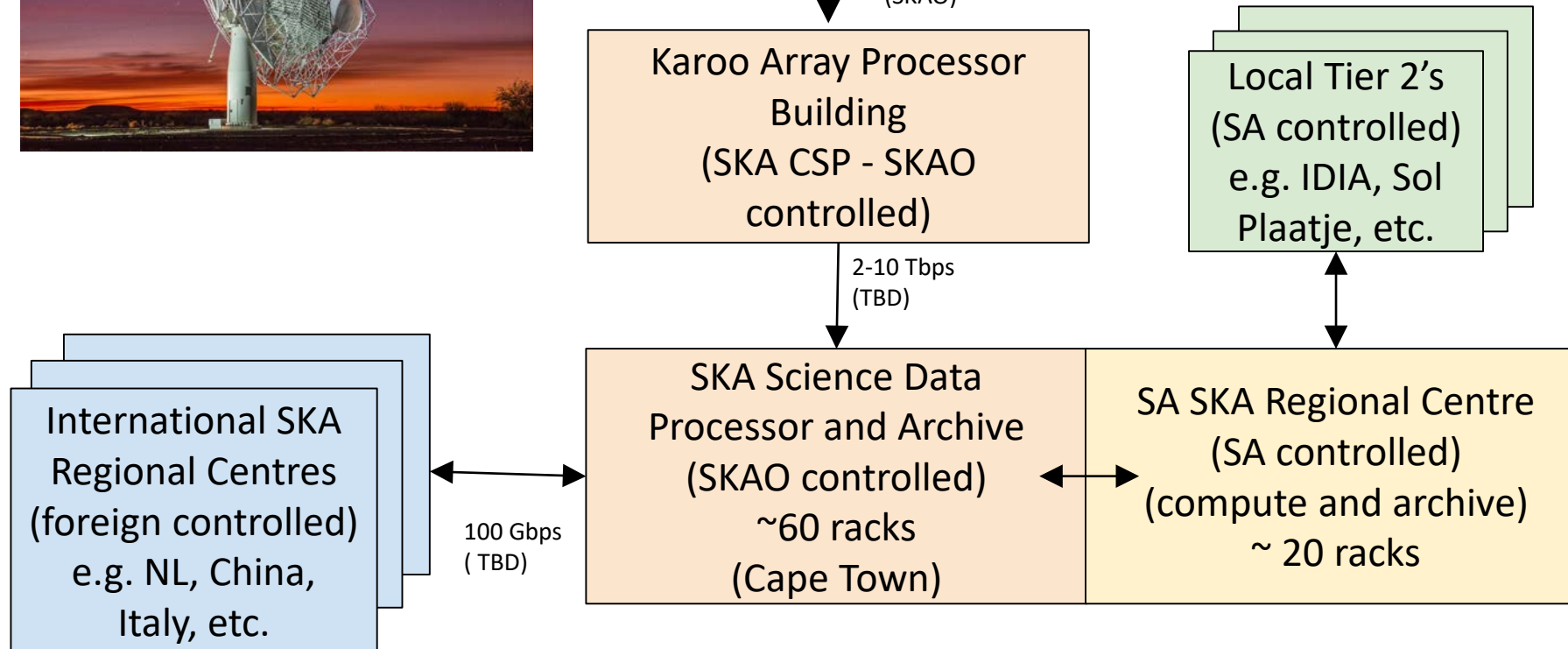
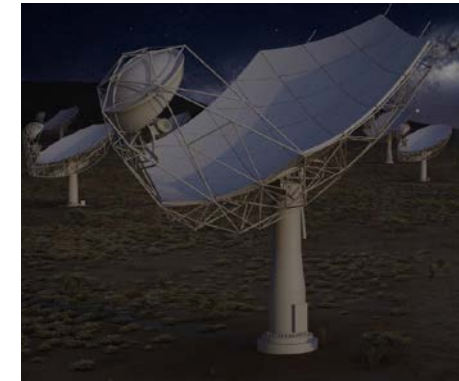
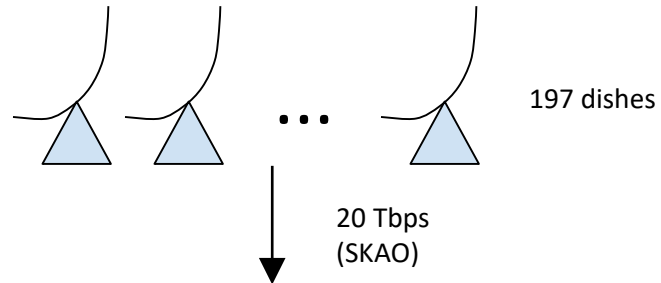


MeerKAT Processing and Archive

- Currently have storage in place in Carnarvon and CHPC
- Arriving today:
 - 200 TFLOPS compute
 - 10PB tape library
- Will have 20PB raw disk storage by Feb 2018 (~10TPB now)

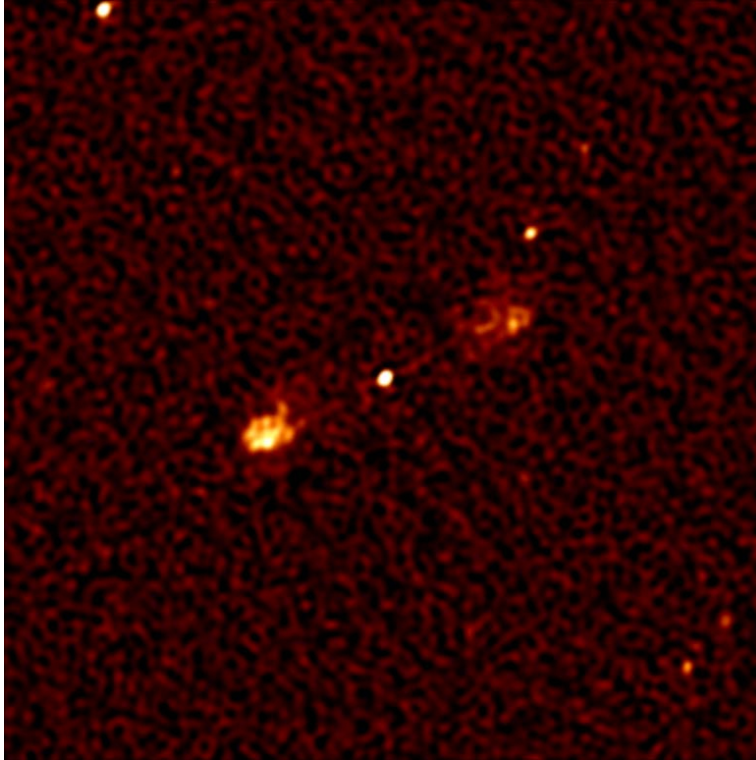


SKA1 Era Regional Centres



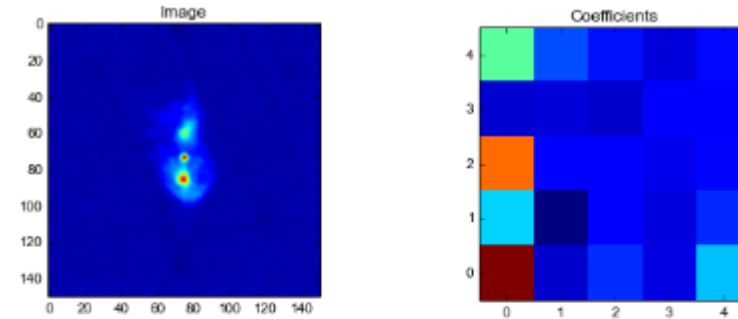
Galaxy Morphology

Arun Aniyen

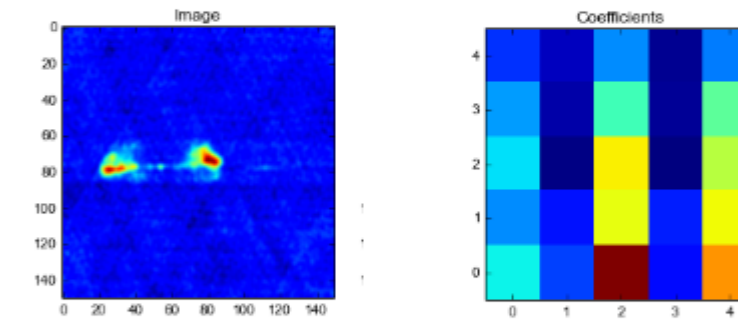


Shapelet and Deep Learning approaches
under development

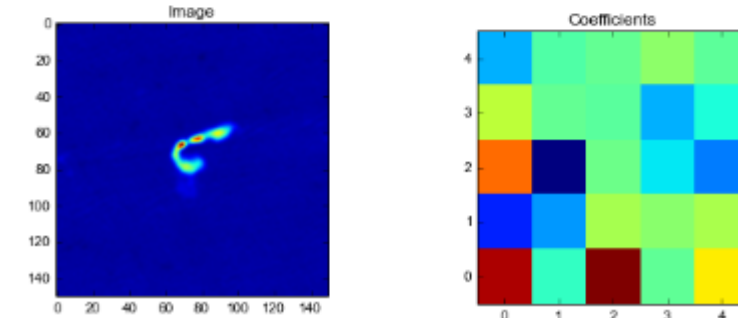
FRI Type Galaxy



FRII Type Galaxy



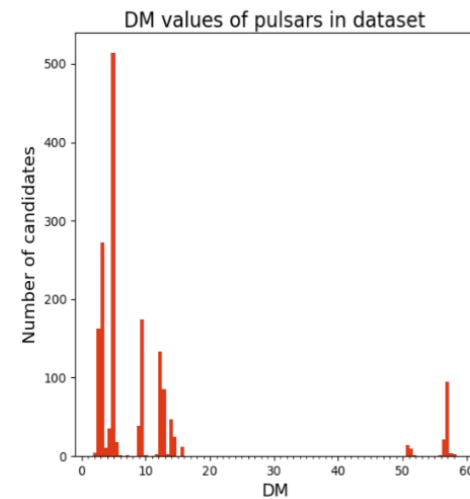
Bent Tail Galaxy



Pulsar Candidate Selection



Image: NASA



Please enter the DM limits for the training set

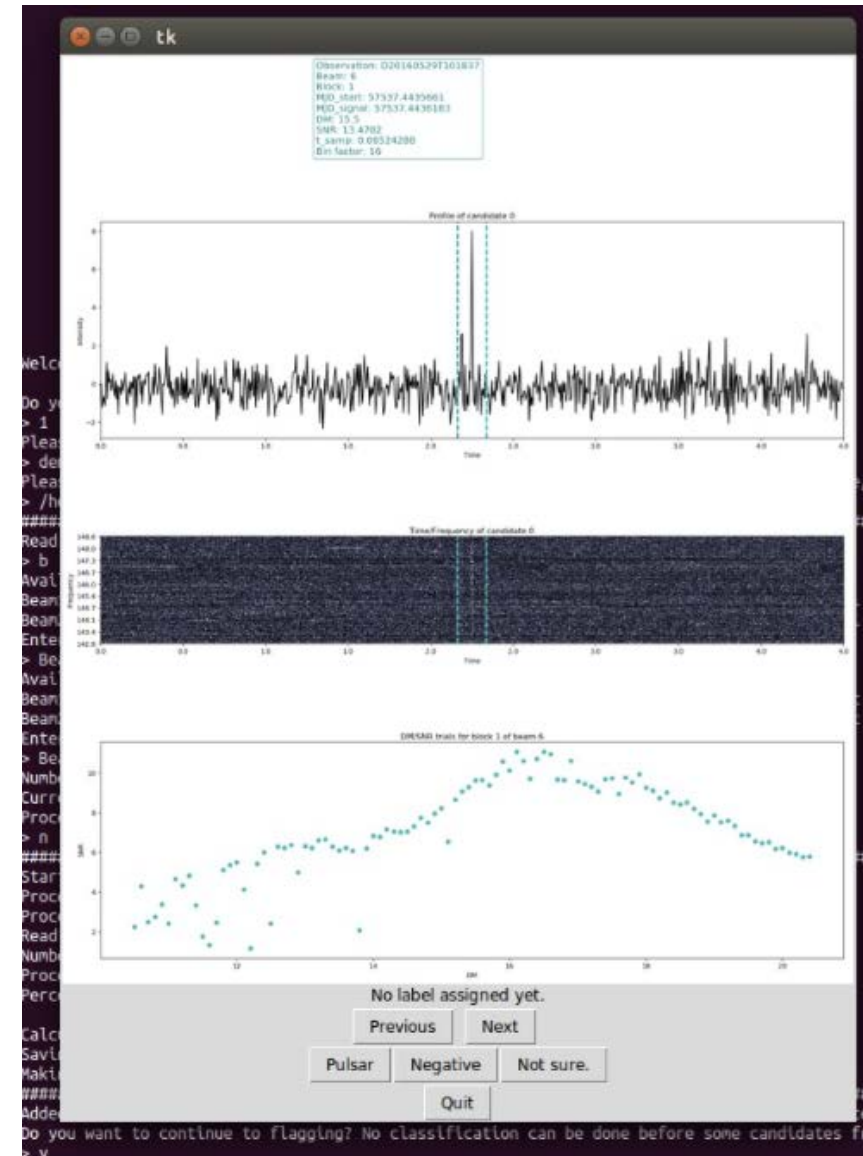
Start DM: 9

End DM: 53

Submit

Quit

Illustration 5: Training / testing split interface;



CyberSKA Portal / Gateway



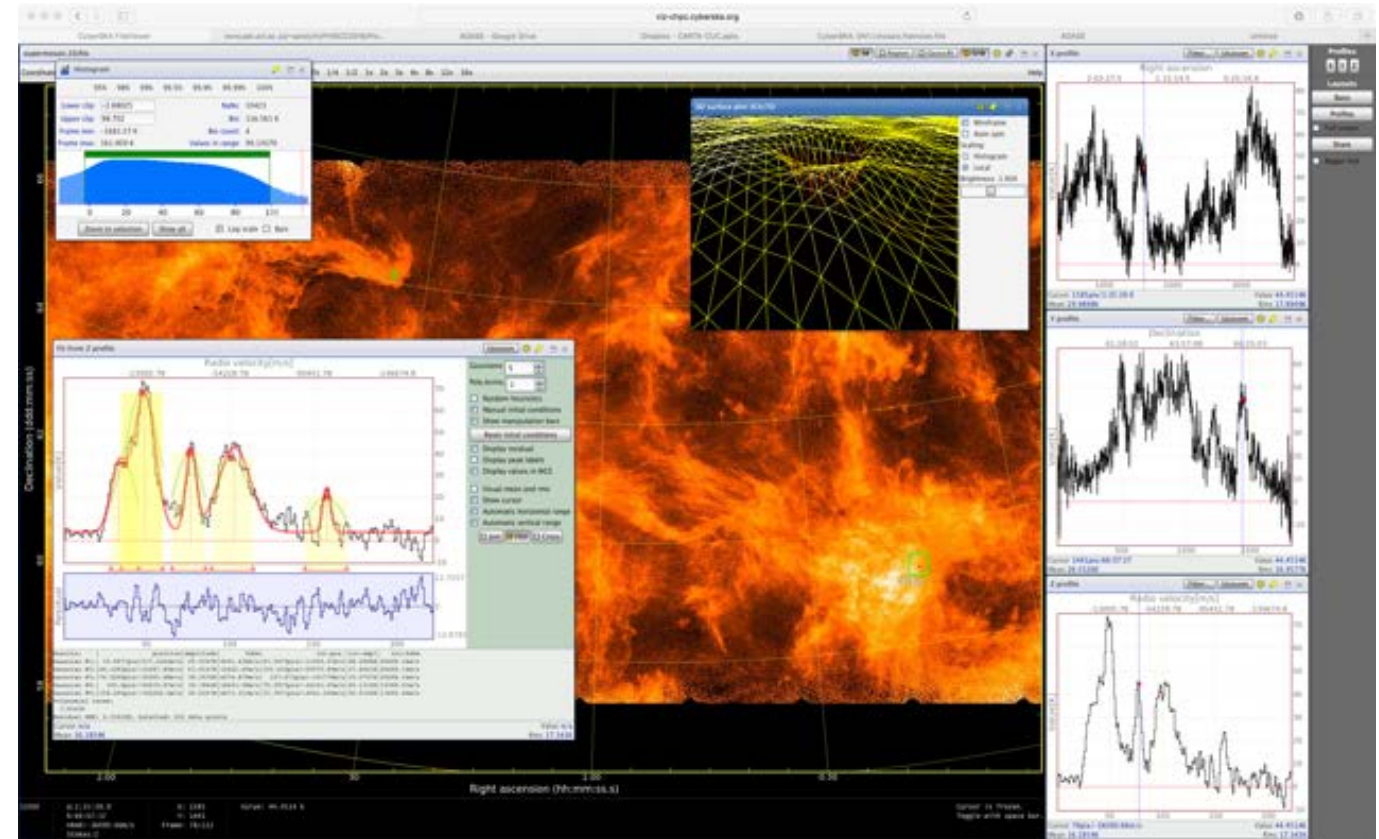
The screenshot shows the CyberSKA Portal / Gateway interface. At the top is a dark banner with the text 'CYBERSKA' in large white letters, followed by 'A Cyberinfrastructure platform to meet the needs of data intensive radio astronomy on route to the SKA' in smaller white text. Below the banner is a navigation bar with links: Home, Profile, Settings, myDashboard, myGroups, Tools, About, Help. There is also a search bar and a 'Log out' button. The main content area is divided into several sections. On the left is a user profile for 'Cameron Kiddle' with a 'Subscribe to feed' button and a 'Bookmark this' button. Below this is a 'My applications' section with links to 'My blog', 'My bookmarked items', 'My event calendar', 'My files', 'My pages', 'My publications', and 'My tasked items'. The 'Contacts' section shows a grid of 24 user avatars. The 'Event calendar' section lists three events: 'Imaging Science Technical Meeting' (10:00 - 11:00, 28 Jun 2011), 'CANARIE Site Visit (Tentative)' (Semi-annual site visit by CANARIE staff, 1:00 - 3:00, 29 Aug 2011), and 'ADASS XXI' (Astronomical Data Analysis Software and Systems Conference, 6 Nov 2011 - 10 Nov 2011). The 'Recent Astro-ph Eprints' section lists four eprints with PDF links. The 'Active Users' section shows two active users: 'Cameron Kiddle' (Busy preparing for SKA 2011 - http://ska2011.org/ update (12 days ago) Location: calgary, alberta, canada) and 'Samuel George' (my 'Job' is waiting for processing, please check back later for results"... (5 days ago) Location: astrophysics group, university of cambridge, cavendish laboratory cambridge cb3 0he, uk). The 'Pages' section shows one page: 'CyberSKA Updated Collaboration Requirements - Phase III' (Last updated 21 days ago by Cameron Kiddle). The 'Group membership' section shows three groups: 'DMS support of Astronomical Data' (This subgroup is for planning and developing specialized support Astronomical Data in the CyberSKA Data Management System.), 'CyberSKA Sys Admins', and 'Application Developers' (Group for developers working/creating portal applications.). The 'Activity' section shows three recent activities: 'Samuel George bookmarked Detection Thresholds and Bias Correction in Polarized Intensity (4 hours ago)', 'Russ Taylor updated a page titled SubGroups (9 hours ago)', and 'Mircea Andreut has posted a new comment on this discussion topic | SKA 2011 Travel Plans for UoFC participants' (I can take one person. (16 hours ago)).

- Over 700 users
- Latest version enables federation of multiple portals
- iRods used for data management
- Provides access to data sharing, collaboration, visualisation and data search tools

Visualisation



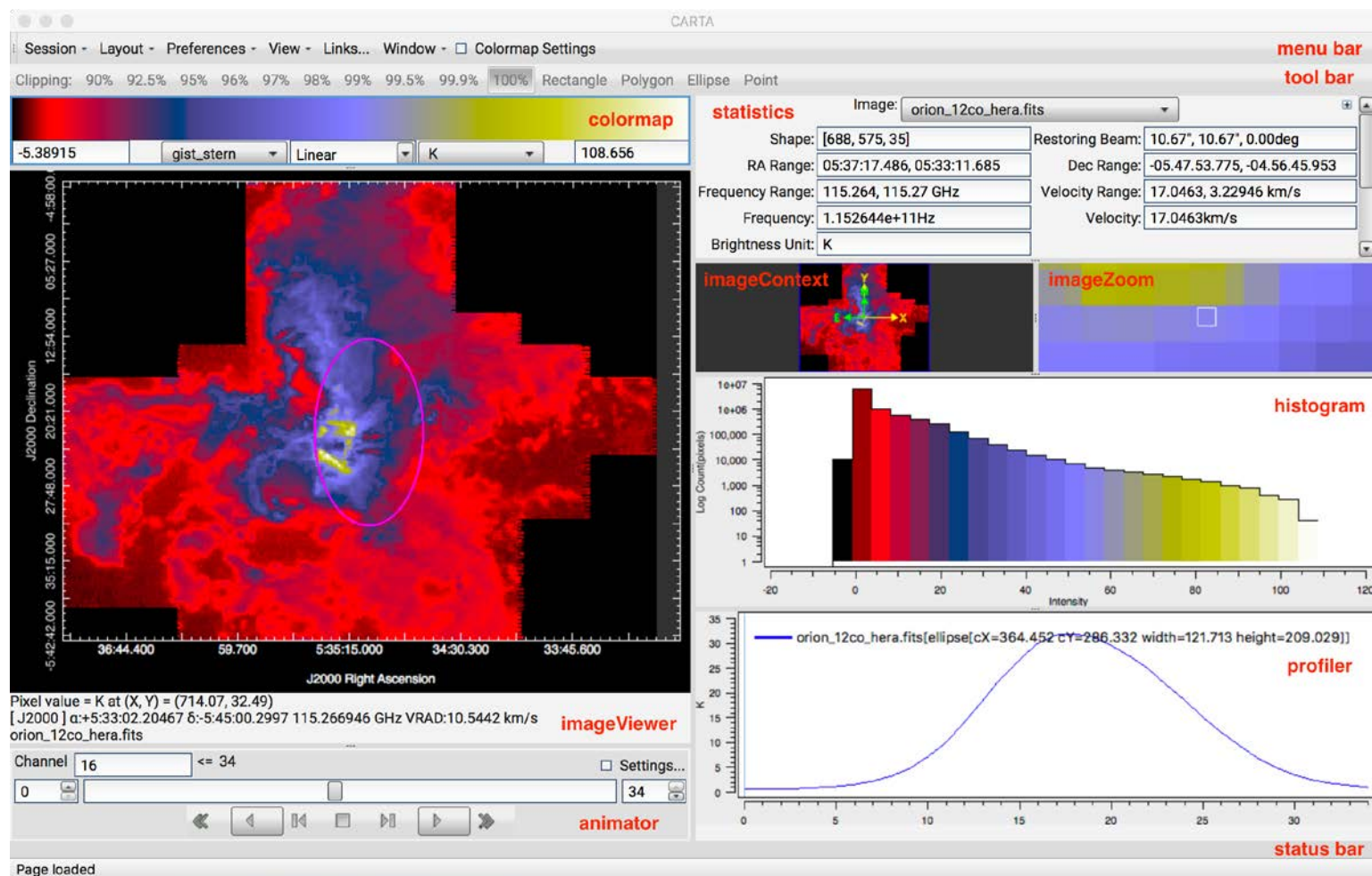
- CyberSKA remote radio astronomy viewer is primary visualization tool in CyberSKA portal
- Currently used with up to 360 GB data cubes
- Provides range of visual analytics algorithms
- Enables sharing of visualization sessions between distributed participants



CARTA viewer



- Developing new viewer
 - NRAO collaboration
 - Will replace CASA viewer
- Aim to scale visualisation and analytics to multi-terabyte cubes
- Exploring HDF5 formats supporting parallel I/O
- Currently reworking architecture to provide common GUI across workstation and remote access platforms



Summary



- SKA SA aim to support activities for SRC development
- SDP has developed baseline architecture for delivering data to SRCs. SRC interface definition document to be reviewed soon
- MeerKAT RC framework being developed in multi-partner collaboration and initial data starting to be distributed
- CyberSKA portal and CARTA viewer development is ongoing with CARTA architecture update to unify GUI across platforms
- South Africa planning to have SRC in addition to SKA1 Mid Processing Centre



MeerKAT First Light (FR II galaxy)

