IDIA Inter-University Institute for Data Intensive Astronomy

Data transport for MeerKAT and SKA regional centres

David Aikema
University of Cape Town

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

University of Cape Town • iYunivesithi yaseKapa • Universiteit van Kaapstad

UNIVERSITY of the WESTERN CAPE

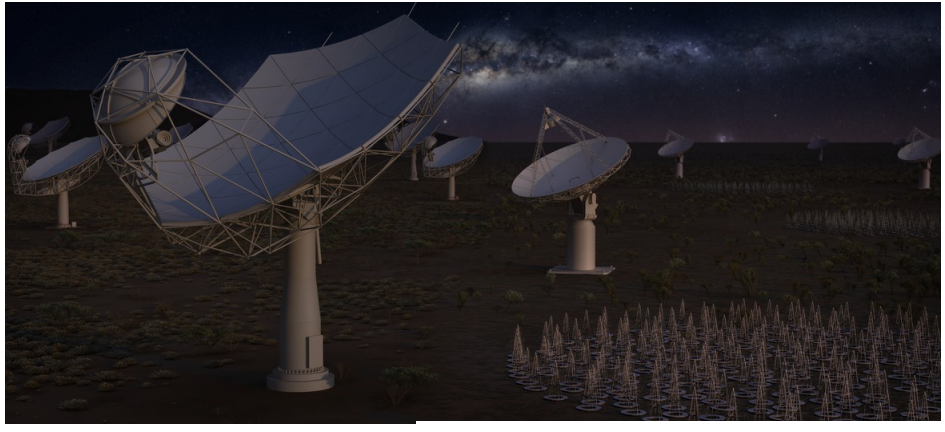
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

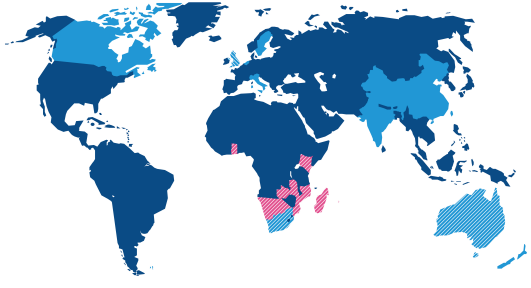
Outline

- Background:
  - SKA / MeerKAT
  - SKA SDP
- Regional Centres
- Scheduling Data Transfers
- Components
- Architecture Overview
- Status

SKA

- Square Kilometre Array
  - To be the world's largest radio-telescope
  - Global collaborative effort
  - Hosted in South Africa & Australia
- BIG data
- Currently in design phase
  - Aim start construction in 2018/2019





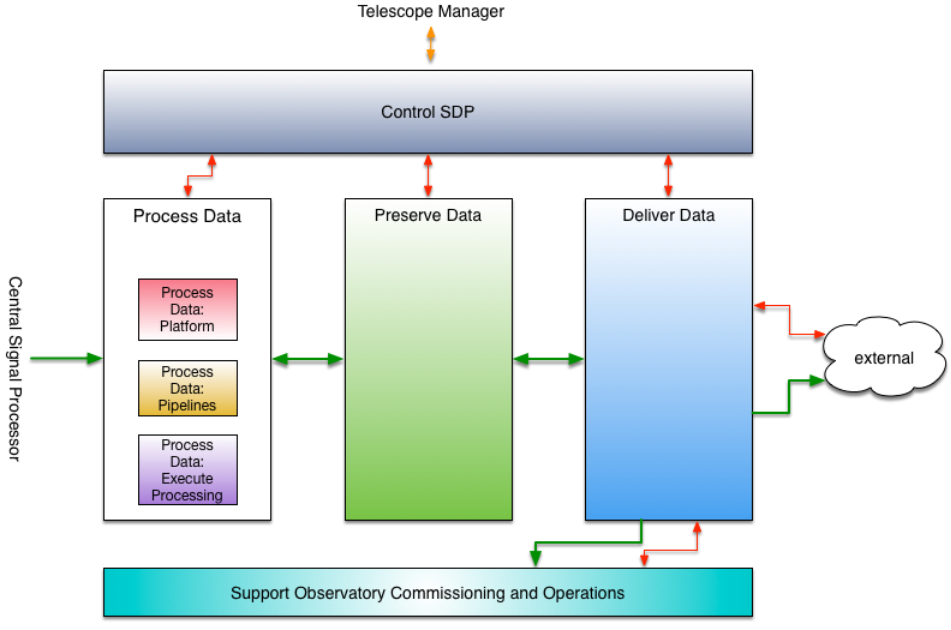
SKA
SQUARE KILOMETRE ARRAY

MeerKAT

SKA AFRICA
SQUARE KILOMETRE ARRAY

SKA Science Data Processor (SDP)

- One of several SKA work packages
- Science Data Processor
  - Receive data from correlator
  - Produce / store Science Data Products
  - Deliver Science Data Products
- Delivery only to observatory staff and SKA Regional Centres



SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

Regional Centres

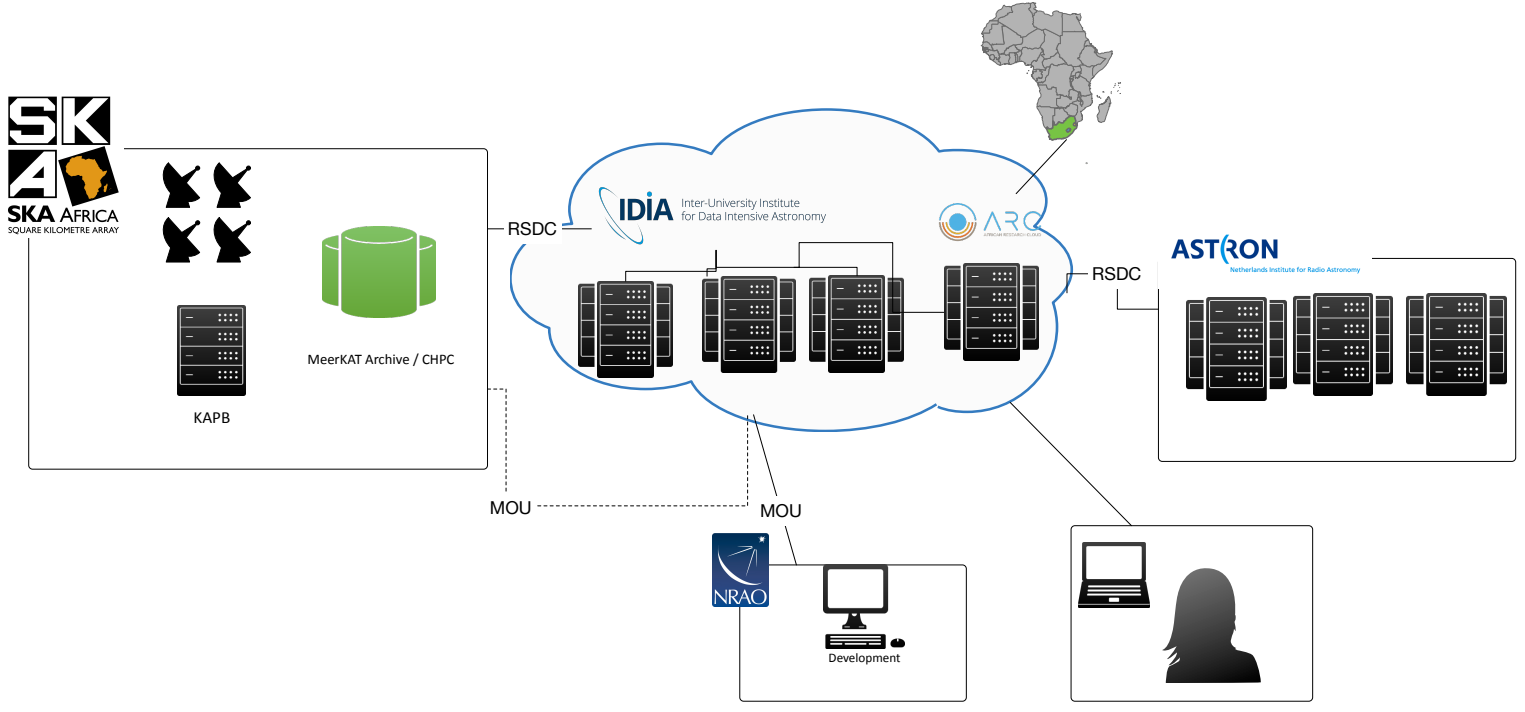
- Scientists working on SKA data will be distributed
  - (so are the funding agencies)
  - Minimize latency to data being worked on
- Further processing of products required before in final state for analysis
  - SDP throughput must keep up with input data
  - Scope limits prevent certain types of derivative data products being created
  - Projects will produce more data than could be stored in the archive
- Data volumes are too large to send direct to end users

Why schedule data transfers?

- Allows priorities to be set on which data is moved next
  - Adhere to user/project resource allocations
  - Avoid starvation
- Manage network to maximize performance
  - Handle congestion – particularly on long-distance links (ASTRON)
  - Ensures that WAN is kept busy by keeping data in flight
  - Use efficient WAN data transfer protocols
- Allows checks to see if data is available at other locations
- Support subscriptions to datasets

MeerKAT

DOME project: IBM / IDIA / ASTRON / SKA-SA



Getting data

- *Query*
  - *CADC supplying packaged set of IVOA services*
- Request: Get a (set of) data product(s)
  - Also includes subscription support
- Prepare: Preprocessing – e.g. region extraction, averaging
- Deliver: Ship the data

- Request / Prepare / Deliver:
  - Focus of the DOME Delivery Service

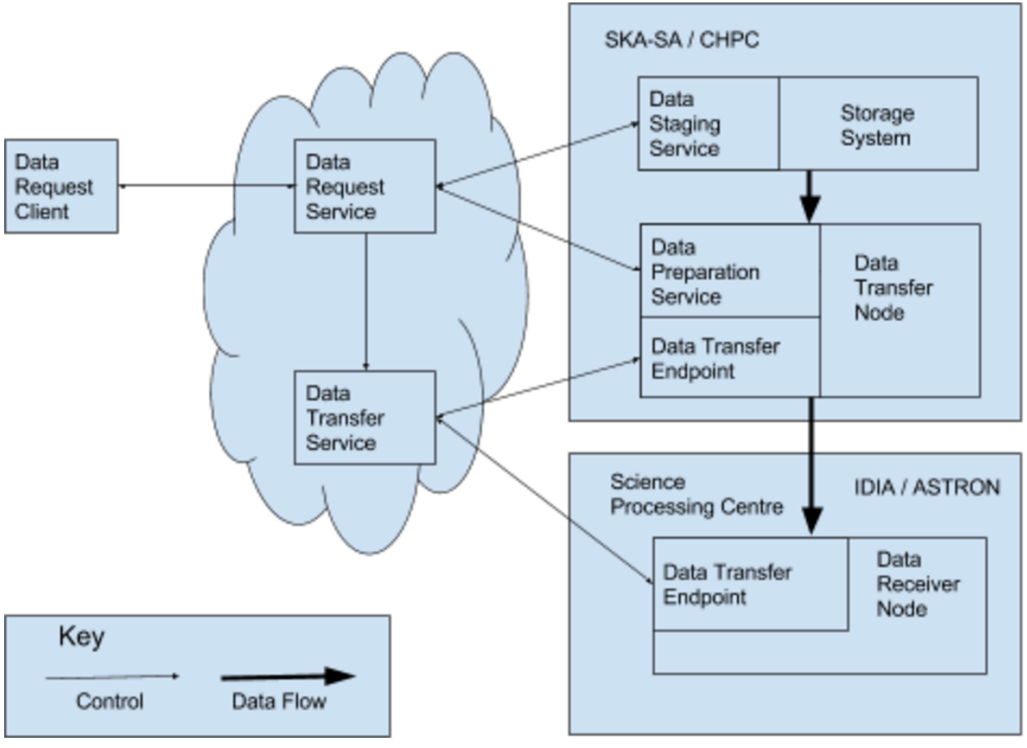
Transfer Components

- Transfer Service: The central service
- Stager: Copy product from archive to scratch space on a transfer node
- Node Agent: Perform preprocessing; Assist in managing data

- Integrated:
  - FTS – using GridFTP for transfers

- Security: X.509

Transfer Tool Overview



Status

- Operational prototype of the transfer service
- Query service: expecting the delivery of CADC Query service docker images in near future
  - Integration required
- Testing
- Expansion to meet the needs of the SKA


Questions?
David Aikema - david.aikema@uct.ac.za