

# SKA: Data products...and what to do with them

Rosie Bolton

SKA Science Data Processor consortium and  
SKA Regional Centre project scientist



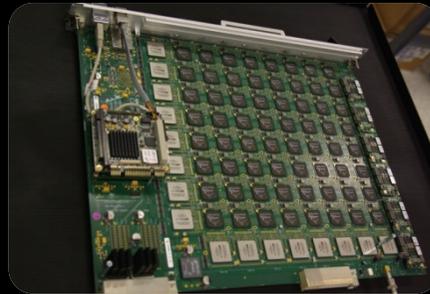
# SKA: A Leading Big Data Challenge for 2020 decade



## Antennas



## Digital Signal Processing (DSP)



Transfer antennas to DSP  
2020: 5,000 PBytes/day  
2030: 100,000 PBytes/day

Over 10's to 1000's kms

**HPC Processing**  
2020: 300 PFlop  
2028: 30 EFlop

To Process in HPC  
2020: 50 PBytes/day  
2030: 10,000 PBytes/day

Over 10's to 1000's kms



**High Performance Computing Facility (HPC)**

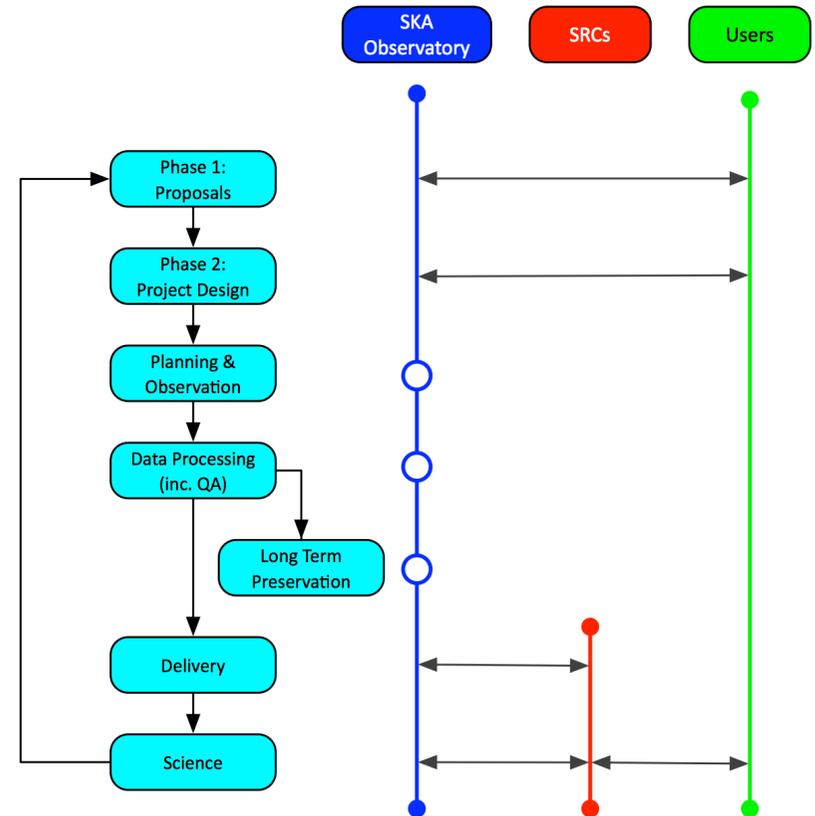


UNIVERSITY OF CAMBRIDGE

# SRC and SRC-Alliance Requirements and Goals



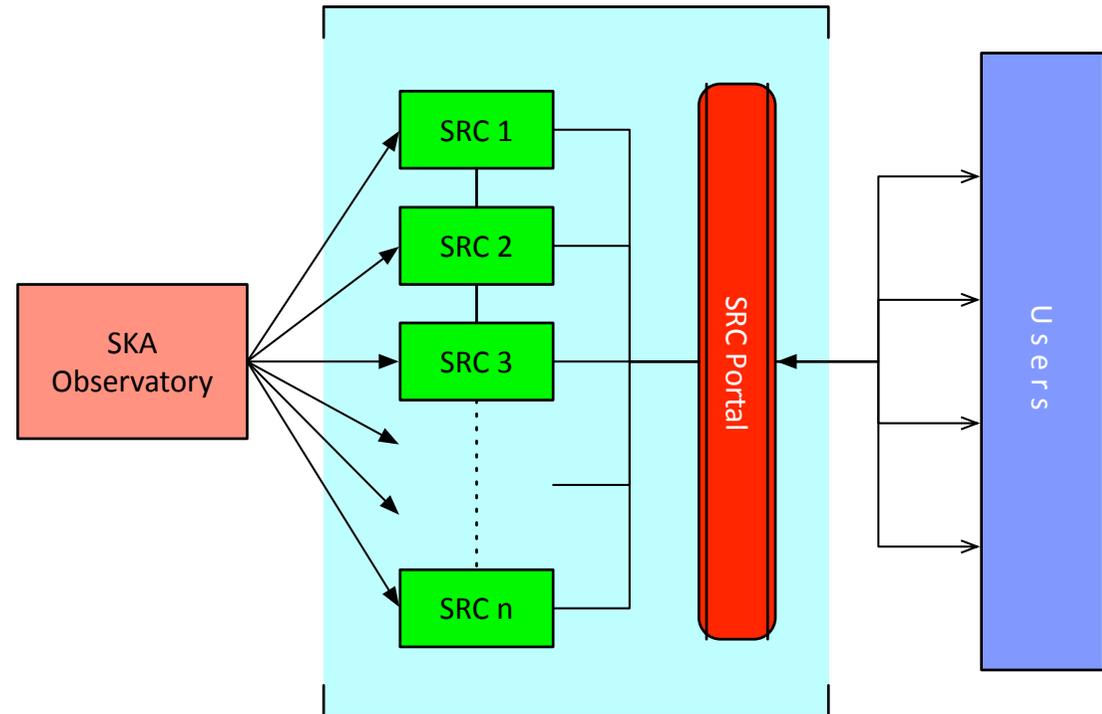
- Provide users across the globe with access to SKA data products.
- Provide capability to generate (and archive) advanced data products.
- Preserve and extend provenance information.
- Strongly evolving landscape over 50 year SKA Observatory lifetime



# What do SKA regional centres look like to SKA Observatory?



- What responsibilities can the alliance of SKA Regional Centres commit to?
- Not a subservient relationship: the principal responsibility of the SKA Regional Centres is to serve the astronomy community, not the observatory.
- Governance structure of the Alliance of Regional Centres (“WSKACG”) not yet determined.



# What are SKA regional centres? What do they “look” like to funding agencies?



- How are these funded?
- What is the physical infrastructure needed for compute and storage?
- What is the software infrastructure needed to manage this system (account logging, security, data permissions).
- What is the relationship to other science areas (WLGCC, other ESFRI projects, general science users)

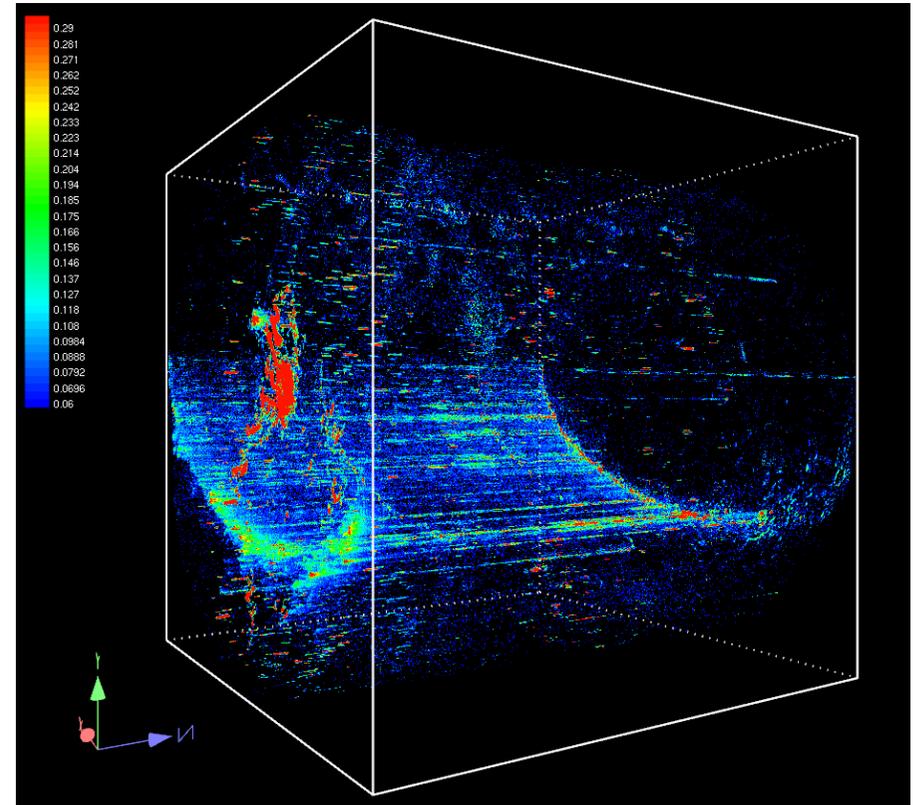


UNIVERSITY OF  
CAMBRIDGE

# What do SKA regional centres “look” like to users?



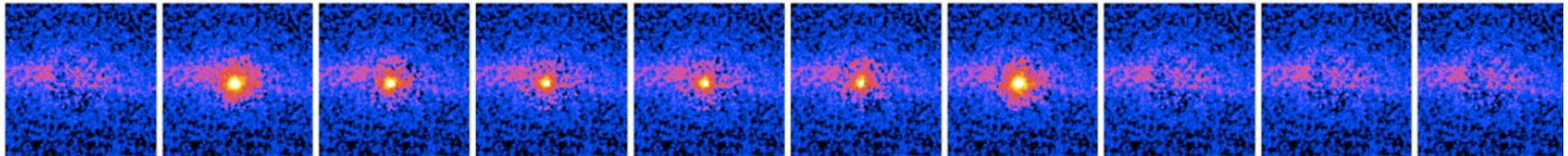
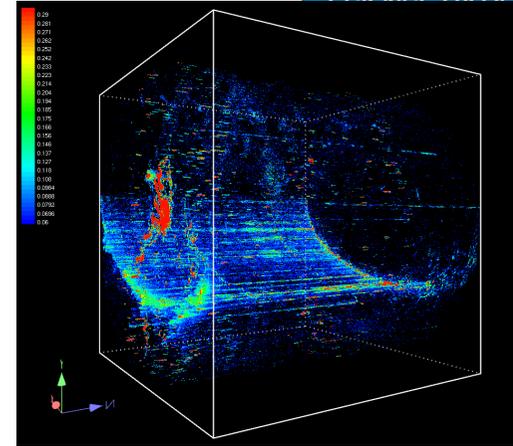
- Can we have a federated system supporting global user base in a location-agnostic way?
- What does the Science Gateway need to do?
- What is the mixture between interactive visualisation work and batch processing?



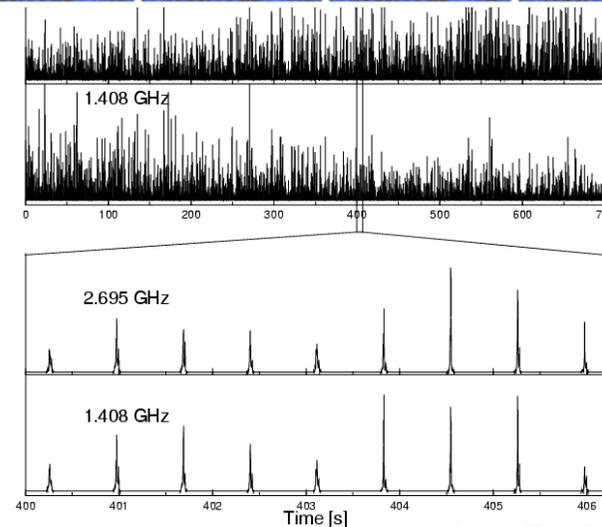
# SKA Observatory data products – many types



- Image cubes (2 spatial dimensions, plus radio spectral frequency, polarization)
  - Each can be huge, typically minutes-to-hours integrated together
  - High speed image plane searches



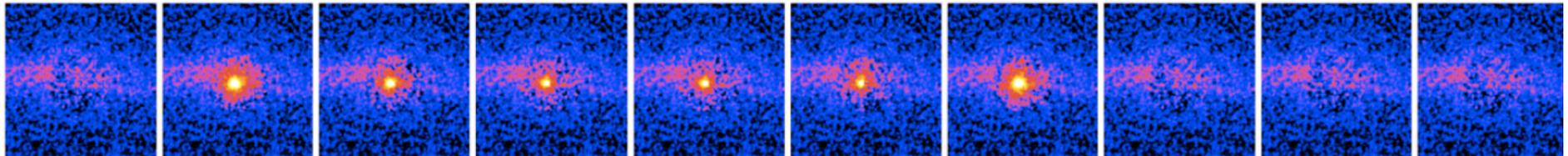
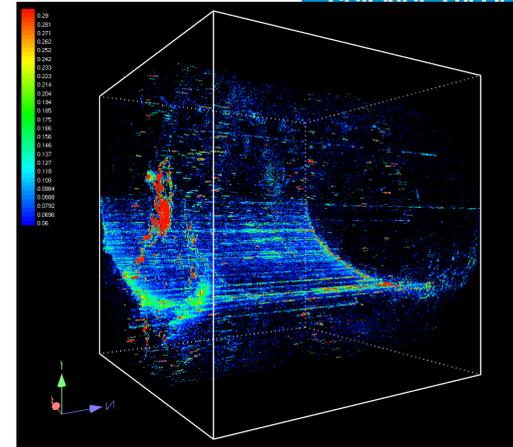
- Correlated visibility samples
  - Few seconds to Sub-second time resolution
- Time-series data for pencil-beams: search for milli-second to second periodic variability in pulsars



# SKA Observatory data products – data rates



- Image cubes (2 spatial dimensions, plus radio spectral frequency, polarization)
  - Each can be huge, typically minutes-to-hours integrated together
  - High speed image plane searches

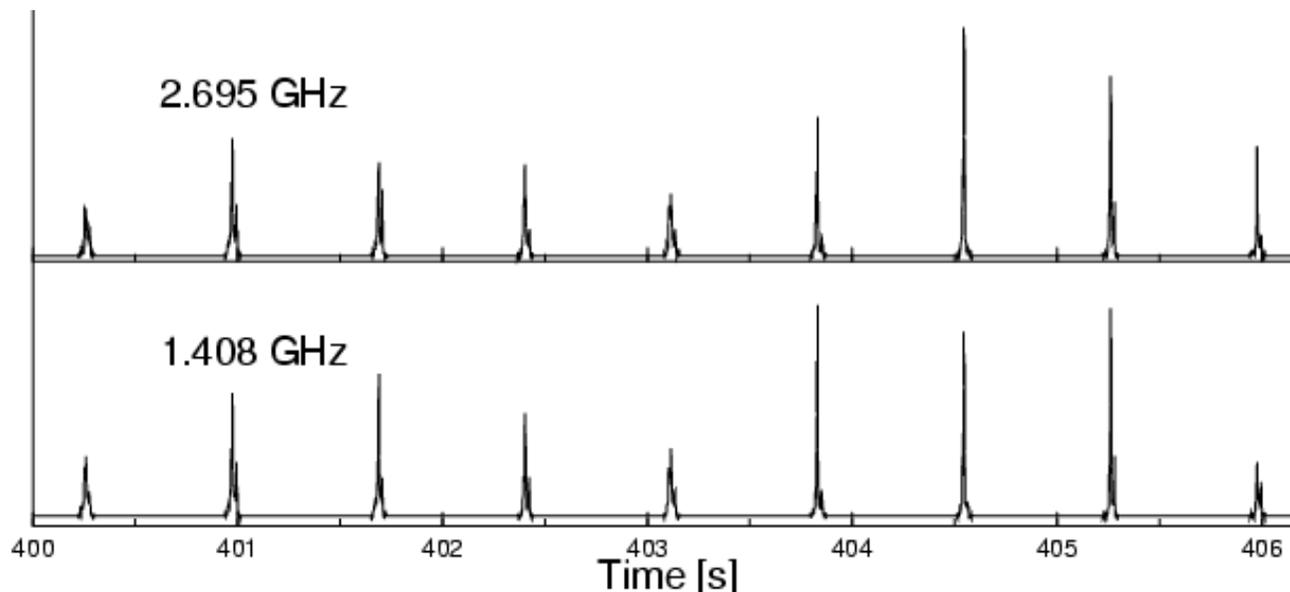


- Deep-cube: per 6 hours integration,  $O(50k \times 50k)$  pixels, 50k channels, 4 polarisations: 5 Petabytes. 1.85 Tbits/s (20 x 100gbit/s links)
- Image plane searching: per 1 second,  $O(5k \times 5k)$  pixels, 10 channels, 1 polarisation: each cube 25 Gbytes, 200 gbit/s

# SKA Observatory data products – data rates



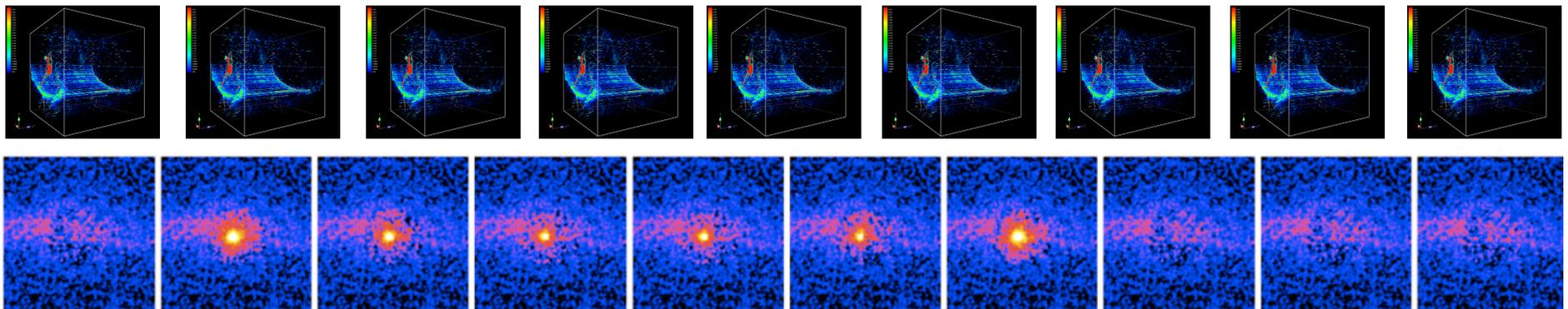
- Correlated visibility samples
  - Few seconds to Sub-second time resolution
  - Data rate out of SKA *depends on averaging* but could be between a few terabits/s and  $\sim 200$  gbit/s
- Time-series data for pencil-beams: search for milli-second to second periodic variability in pulsars
  - Data rates out of SKA are much lower than image-plane or correlation data products. However, might want to always run transients search mode alongside other modes.



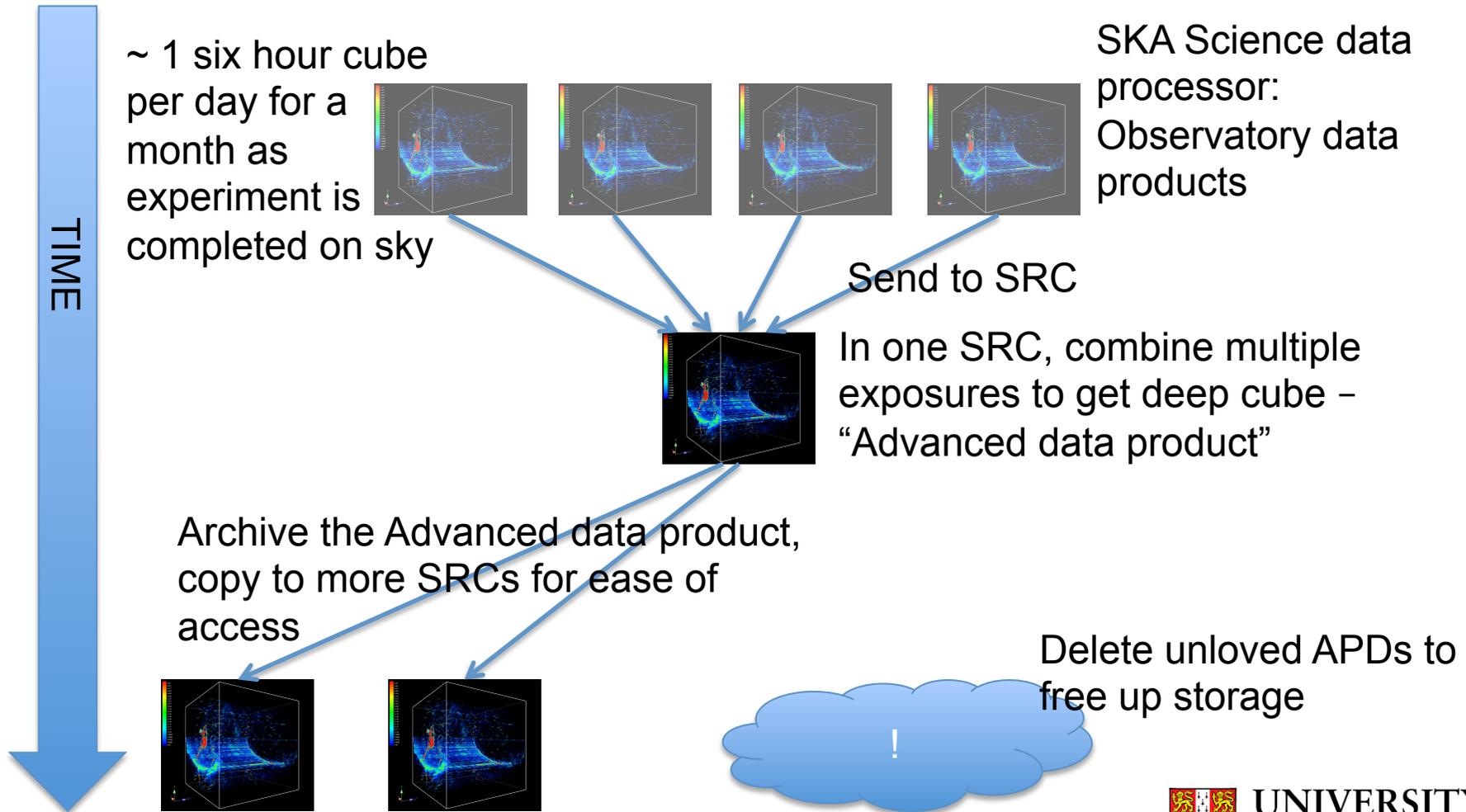
# SKA Observatory data rates



- Data rates out of the observatory are Highly tuneable
- Our estimates based on simple model of experiments are  $\sim 20$ gbits/s continuous
- But we could easily imagine scenarios ten or hundreds of times faster than this
- Ultimately, we'll be limited by the SKA Regional centres' capacity to take more data, and by the (cost) of networking around the globe. (Or by the number of astronomers ready to write papers.)



# SKA Data lifecycle





UNIVERSITY OF  
CAMBRIDGE

END