



Compute and Storage for SKA Regional Centres

A. M. M. Scaife*⁽¹⁾, Rohini Joshi ⁽¹⁾, Therese M. Cantwell ⁽¹⁾, Alex O. Clarke⁽¹⁾ and the AENEAS Project⁽²⁾

(1) Jodrell Bank Centre for Astrophysics, School of Physics & Astronomy, University of Manchester, Manchester M13 9PL, U.K., e-mail: anna.scaife@manchester.ac.uk; rohini.joshi@manchester.ac.uk; therese.cantwell@manchester.ac.uk; alex.clarke-3@manchester.ac.uk

(2) www.aeneas2020.eu

The Square Kilometre Array (SKA) will be the world's largest radio telescope. The pre-construction design effort for the SKA started in 2012 and involves approximately 100 organizations in 20 countries. The scale of the project makes it a huge scientific endeavour, not only in terms of the science that it will ultimately undertake, but also in terms of the engineering and development that is required to design and build such a unique instrument. This design effort is now drawing to a close and early science operations are anticipated to begin in ~2025.

The scientific priorities for the SKA have been defined, in collaboration with the observatory, by eleven science working groups (SWGs). These priorities are summarized in a list of thirteen High Priority Science Objectives (HPSOs) for the SKA. In addition to the HPSOs, the observatory will also reserve a fixed amount of time for *open proposals*, which will be accepted on a competitive basis.

Primary data products from the SKA telescopes will be generated by two High Performance Computing (HPC) sites, known as Science Data Processors (SDPs), within the observatory itself. The SDPs will be connected onto the global backbone using one (bi-directional) 100 Gbps data link per SDP. Without the observatory-level processing performed in the SDPs, the output data rates might be 5-10 times higher than can be accommodated by such links. On average, the SDPs will compress the raw data from the telescopes by a factor of 40.

The SDPs will produce fifteen standard Observatory Data Product (ODP) types, specified by the observatory. These data products are designed to reduce the absolute data volume from the telescopes, but are not suitable for immediate scientific publication. Due to both the volume of SKA data products and the processing requirements they entail, both SWG members and the principal investigators (PIs) of open proposals will receive their SKA data products through one of the worldwide SKA Regional Centres (SRCs).

SKA Regional Centres exist outside the observatory and will be provided by the member countries of the SKA project for the benefit of the scientific community. These regional centres are expected to provide both storage and processing facilities for SKA data so that secondary, or post-, processing can be performed. The secondary ('advanced') data products produced by this post-processing will also be stored at the regional centres.

Unlike other scientific facilities such as the Planck satellite, LSST or the LHC, the SKA is not a single experiment or set of experiments but is an observatory. As a consequence it is not possible to define a finite set of compute models for an SRC in the same way as has been done to define the scientific computing requirements for (e.g.) the experiments of the Large Hadron Collider (LHC). An SKA regional centre requires the flexibility to enable a broad range of single user development and processing, as well as providing an infrastructure that can support efficient large-scale compute for the standardized data processing of reserved access key science projects.

1. A. M. M. Scaife & the AENEAS Project, "Analysis of compute load, data transfer and data storage anticipated as required for SKA Key science," *AENEAS Deliverable D3.1*, <https://www.aeneas2020.eu/project-deliverables/>

2. A. M. M. Scaife & the AENEAS Project, "Initial System Sizing for the European SKA Data Centre," *AENEAS Deliverable 3.3*, <https://www.aeneas2020.eu/project-deliverables/>