



SKA - CRISP WP19

17th March 2013

Dr. Rahim Lakhoo (a.k.a Raz)

P.I – Dr. David Wallom

Outline

- Plan overview
- Requirements – Data distribution and access
- Requirements – Network services
- Metadata investigation
- Survey of tools and technologies
- Prototype infrastructure
- Summary

Plan overview

- Gather requirements for data distribution and access
 - Square Kilometre Array
- Gather requirements for network services
 - Global fibre networks
- Metadata investigation
 - Data continuum
 - Open access data
- Survey of tools and technologies
 - Requirements based
 - Create a shortlist
- Prototype infrastructure
 - Controllable testbed
 - Emulate network/systems

Requirements – Data distribution and access

- Gather data distribution and access requirements from current phase 1 and phase 2 plans for Square Kilometre Array (SKA)
 - SKA phase 1 baseline specifications were released last week
 - Request for proposals (RfP) may introduce new requirements
- SKA has two sites planned for phase 1 - South Africa and Australia
 - SKA phase 2 uses the same sites, but on a larger scale
- SKAs Science Data Processor (SDP) is envisioned to develop a significant amount of data
 - Assess performance parameters for data distribution
 - Map data distribution requirements to existing e-Infrastructures and other ESFRI clusters
- Gather data distribution and access requirements from existing pathfinder and related projects (i.e. LOFAR, KAT-7, ASKAP, ALMA)

Requirements – Network services

- Build on the data distribution requirements from SKA plans to ascertain network service requirements
- Data from different geographical sites – South Africa and Australia
 - Possibility of merging of data from the two sites
- Work with network providers (i.e. GEANT/NRENS) to ensure connectivity and bandwidth to RIs
 - Assess currently installed global fibre network
 - Develop a roadmap for any required new global fibre connections
 - Take into account the expected changes in the models of data access until RI services are online

Metadata investigation

- Supporting data continuum from raw data to publication
- SKA plans to support two modes of operation that will need metadata associated
 - Public - All sky surveys with open access to data
 - Dedicated - PI requests for time on the telescope
- Metadata investigation to enhance data identification and access, with links to
 - Operational data
 - Configuration data
 - Workflow representations
 - Etc...

Survey tools and technologies

- There are a number of systems that are currently available for data distribution, which may perform the functions needed. For example:
 - PELICAN framework (Artemis)
 - GridFTP/Globus Online
- Survey current high performance technologies and tools for data distribution
- Select ~3 technologies/tools that best match the requirements found
- Test selected technologies/tools to be tested on a prototype infrastructure

Prototype infrastructure

- Develop and implement a prototype infrastructure for testing
- Infrastructure aims to emulate a network, its end points and resources
- Provides a controllable and configurable environment for testing:
 - Different types of data
 - Resilience
 - Reliability
 - Performance
- The prototype will be built using HPC resources

Summary

- Align efforts with SKA pathfinder and related projects
 - Some are already dealing with similar issues
- Align efforts with current operational large scale infrastructure projects
 - Create a strategy to support suitable technologies/tools
- Develop a prototype infrastructure to evaluate technologies/tools with different scenarios and use cases
- All investigation outcomes and results are planned to provide input for CRISP and SKA documentations