PHYSICAL SCIENCES

Big Data Research Infrastructure Collaboration Toward the SKA (BRICSKA)

RUSS TAYLOR, FABIO PORTO, CHENZHOU CUI, YOGESH WADADEKAR
& OLEG MALKOV

Abstract: Astronomy is entering an era of mega-data that will render conventional research methods as well as data and visual analytics tools ineffective. The Square Kilometre Array (SKA) drives one of the most significant big data challenges of the next decades. South Africa, China and India are partners in the global SKA collaboration and host recently completed, next generation radio astronomy facilities. South Africa, Brazil, China and India are involved in the Large Synoptic Survey Telescope (LSST), which represents a complementary mega-data challenge, vastly increasing the current data volume of optical surveys, and providing critical multi-wavelength data set for SKA analytics. Russian researchers are also engaged in radio astronomy and multi-wavelength, multi-messenger projects driving increasing volumes of observational data. This project brings together teams leading programs in data innovation in each partner country to collaborate on the development of new technologies and systems to meet the big data challenge of SKA pathfinder facilities and the multi-wavelength projects that are critical to the advance of astronomy. In so doing we will prototype and demonstrate scalable big data technologies for the new big data era, and establish a BRICS multinational federated data intensive cloud network for collaborative programs in data intensive astronomy.

Key words: BRICS, SKA, big data, large surveys.

INTRODUCTION

Modern science faces several challenges that are transformative in how science operates and in how innovations and new knowledge are generated and exploited. The challenges are (i) Big Data, how to record, transport and deliver it, (ii) Big Compute, how to process, analyse, and visualise big data, (iii) Big Science; by asking big questions as a global community facilitated by a suite of unique international facilities, science can only progress in leaps through large international collaborations, and (iv) Human Capacity Development; to make sure all people are active participants in the scientific endeavour, and that leadership in science worldwide is inclusively distributed across the globe.

Astronomy is a science at the nexus of these challenges and the BRICS nations have all made the strategic choice to invest in astronomy and in the technologies underpinning its practice; each country with their specific advantages. This project aims to build a strong international partnership to jointly develop technologies and systems that address the Big Data challenge posed by astronomy Big Science projects, of which the participating countries are strong members. Seizing the opportunity

presented by large international astronomy collaborations, the project will significantly contribute to Human Capacity Development in BRICS to ensure participation and leadership in global science, and by linking those science and technology skills to society and industry, contribute to stimulating innovation.

With BRICS countries home to some of the best astronomy facilities in the world, this project aims to facilitate, through development of people and technology, the synergies between these facilities, thereby addressing some of the big challenges facing science today, and helping answer some of the big scientific questions of our time, which have mobilised international resources into large scientific projects like the SKA and the LSST.

The SKA drives one of the most significant big data challenges of the coming decade (An 2019). New technologies under development on the pathway to the SKA are creating commensurate advances in the data capacities of radio telescopes. The data rates from existing radio telescopes to the researcher are 10,000 times larger than what was typical only a few years ago. At the same time the enterprise of observational radio astronomy is changing. Projects undertaken by large global collaborations, in which large amounts of observing time are devoted to major key science programs that create vast data sets, is becoming the new paradigm. This mode of observing combined with the new instrumental capacities is driving an exponential growth in the rate of data confronting researchers (see Figure 1).

To rise to the scientific opportunity opened up by this new generation of instruments requires the research community to develop new infrastructure, algorithms, software systems and platforms to manage, process, analyse and mine these data sets. The enterprise of science in this new data intensive age requires

new sets of skills and technologies; skills that combine knowledge of astronomy with computer science, information science and statistics, and technologies that adapt the innovations of the fourth industrial revolution to the service of scientific research. The international scope and scale of these next-generation astronomical endeavours, and the multi-national nature of the collaborations on large science programs means that we must respond to these challenges as a global community.

Three of the BRICS partner countries are part of the international SKA project organisation. All partner countries are engaged in data-intensive multi-wavelength and multi-messenger programs providing critical ancillary data to realise the scientific potential of SKA pathfinders and the SKA itself. The LSST is the flagship of upcoming optical facilities that will create vast data sets that must be mined both for transient phenomena and to create catalogues of multi-wavelength data for billions of objects that are interoperable with radio data. Four of the BRICS partner countries are international collaborators of the LSST project. The impact of MeerKAT and the SKA will be dramatically increased by incorporating multi-wavelength data, and the LSST dataset will be the widest, deepest and fastest optical survey ever done. Combining these datasets is however a non-trivial task. This proposal brings together scientific leads of large programs on SKA pathfinders, the LSST and other multi-wavelength facilities along with computer scientists, data scientists and experts in 4th industrial revolution technologies to build solutions for big data astronomy that will enable the full scientific potential of significant investment in major facilities, leverage investment in HPC systems for data intensive research, and develop the expertise to adapt and innovate new technologies to meet
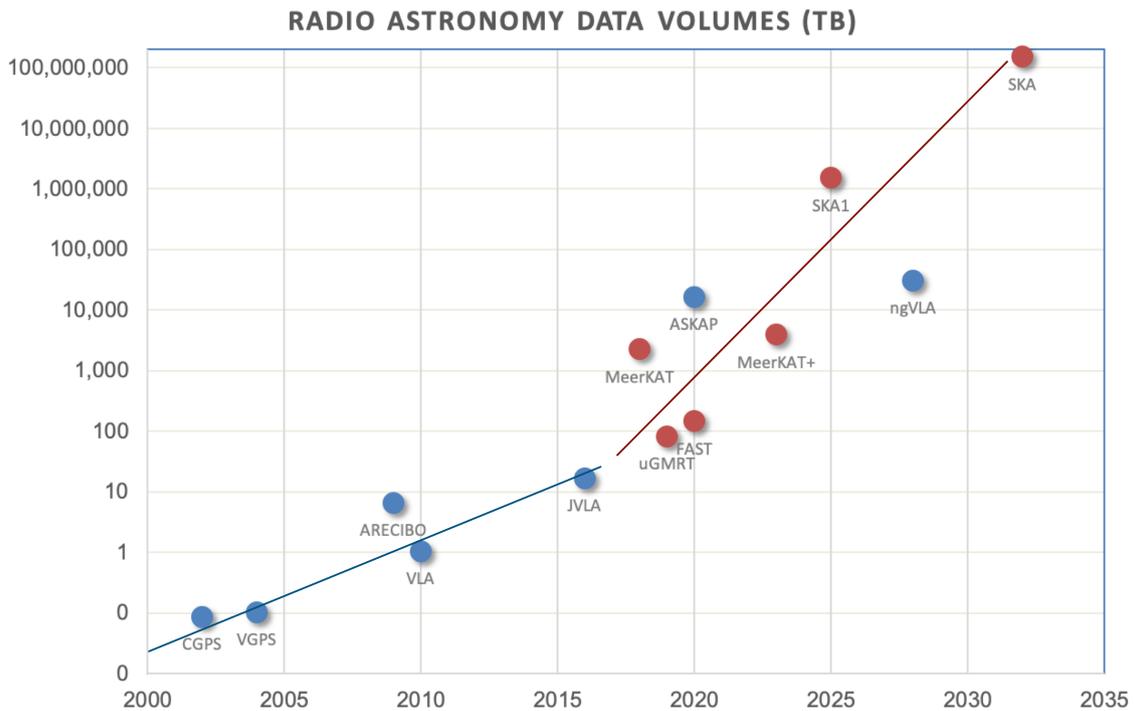
Figure 1. Data volumes for large projects on radio astronomy facilities as a function of time. Current and planned BRICS facilities are indicated in red. BRICS researchers are being confronted with a data deluge in the coming decade characterized by exponential growth that is much faster than has been experienced by the rest of the world to this point.

the exponential growth in astronomical data in the next decade.

Enabling access to globally competitive research for smaller institutions previously isolated from research; revealing and developing untapped talent in the BRICS countries; embedding young researchers into the global networks of science and industry; increasing direct economic benefits of science through interfaces with the private sector; increasing the awareness of, and the support for science among the general public with outreach; supporting development projects through interdisciplinary projects using the research infrastructure; and developing communities through the development of astro-tourism at astronomical facilities in the BRICS countries are some of the potential development benefits of this flagship project that are detailed below.

Realizing the HCD and development potential of the flagship naturally calls for a coherent effort across the BRICS nations. The proposal team and their respective national networks have extensive and complementary experience in all the topics mentioned above. Working jointly on those topics prevents duplication of effort in developing such platforms and promotes sustained collaboration between partners countries beyond astronomy research.

SCIENTIFIC RATIONALE

The concept of the Square Kilometre Array (SKA) emerged in the 1990s. Currently, it is one of the largest international science projects in the world, as it is set to answer fundamental open questions about our universe. Using the SKA and other telescopes described in Table I the participants in this project are research leaders

in most of the SKA key science projects (KSP) as detailed below.

Science Goals

Fundamental Physics

The discovery of gravitational waves has opened up a new window to the universe. SKA will be crucial to look for counterparts in radio to the sources of those gravitational waves. For example, the uGMRT has already demonstrated this with meaningful radio follow-up observations of gravitational wave events such as the GW170817 which was the merger of two neutron stars (Kim et al. 2017). Combining the SKA with the Five-hundred-meter Aperture Spherical Telescope (FAST) in China (Gibney 2019) may also allow us to study the low-frequency universe using high precision pulsar timing, massive black holes, interacting galaxies in galaxy pairs and in groups of galaxies. Some of the fastest and most energetic phenomena hold the key to our understanding of fundamental laws of physics in conditions of extreme energy. The detection of powerful bursts of gravitational waves, gamma rays, X-rays, or radio waves provide evidence for this; however, we still know very little about their appearance in optical light and therefore combining data from large modern facilities into multi-wavelength data sets is key to understanding what is happening in those conditions of extreme energy. Contributions to this science objective are expected directly from the ThunderKAT project on MeerKAT (Fender et al. 2016), with the MeerLICHT telescope, FAST, transient detection in the era of the Large Synoptic Survey Telescope (LSST), uGMRT follow-up of LIGO events, and more.

Magnetism

Magnetic fields are an important element of astrophysical phenomena that is yet to be well understood. They contain energy and influence physical processes on all scales, from star formation to galaxies and to the largest structures known in the universe. We don't yet know how these magnetic fields formed and we are not in a position to explain how strong they are. What SKA will be able to do is map magnetic fields in great detail. For this question and others, the MIGHTEE project on MeerKAT (Jarvis et al. 2016) will reach similar depth to that reached by the larger area SKA sky survey, and thus will provide a pilot to the experiments that will be carried by the SKA on a much larger survey volume. Similarly, when combined with uGMRT data, the SuperMIGHTEE data set (Taylor 2019) will become the premier data on the GHz radio properties of the deep sky thanks to its unique full spectral coverage and ultra-broad band nature.

The Hydrogen Universe

Neutral hydrogen is a powerful probe of the evolution of the universe. Neutral hydrogen is found everywhere and can therefore be used to probe anything from the early distribution of matter to the detailed structure of evolved galaxies themselves. Radio telescopes are the best instruments to observe spontaneous emission in radio waves of neutral hydrogen and therefore to map the distribution of neutral hydrogen in the universe. The SKA precursor and pathfinder telescopes are enabling detailed imaging of neutral hydrogen with unprecedented depth and detail, laying the groundwork for the SKA. From detecting galaxies in the early universe in emission and absorption as proposed by LADUMA (Blyth et al. 2016) and MALS (Gupta et al. 2016) to investigating galaxy formation and

evolution with MIGHTEE, with these projects on the MeerKAT telescope, the evolution of neutral hydrogen in the Universe will be studied. The low-frequency uGMRT receivers will complement these by enabling detections at even higher redshifts, and many new results are already being reported from systematic surveys of HI in emission and absorption, including some of the furthest detections to date of HI in emission. The science programs of FAST also include a large scale neutral hydrogen survey. Matching detected galaxies with objects and signals in other wavelengths will enable new discoveries, and new science. These ambitious research projects will help us understand how this neutral gas eventually turns into stars, and even how supermassive black holes at the centres of galaxies are fueled.

The Transient Universe

One of the key capacities of the new generation of radio telescopes and of the SKA is the ability to measure and detect transient objects. While the universe may appear to evolve only over scales of billions of years, astrophysical phenomena are all dynamic and some can occur over incredibly short timescales. We have only gained an awareness of this as the fields of view of telescopes and the ability to observe on short time scales have improved with advances in technology. Time domain astronomy, where repeated observations can be undertaken over a range of timescales (from sub-second to years), is key to obtaining data of sufficient cadence to unlock the nature of transient objects and understand the dynamics of astrophysical events.

Transient radio emission is associated with essentially all explosive phenomena and high-energy astrophysics in the universe. It acts as a locator for such events, and a measure of their feedback to the local environment. Because of this, radio transients are invaluable probes for subjects as diverse as stellar evolution, relativistic astrophysics and cosmology. ThunderKAT is a MeerKAT large survey program to detect and study such phenomena using the high sensitivity and wide area imaging capability of the MeerKAT. As well as performing targeted programs, ThunderKAT will co-observe with other MeerKAT large survey projects and search their data for transients. The uGMRT also provides for significant capabilities for detections of transient phenomena in the Universe.

In addition, optical observations of transients may lead to important discoveries in other fields of astronomy: Near-Earth Objects, nearby asteroids potentially dangerous for Earth; exoplanets; variability of super massive black holes in galaxies; stellar flares; gravitational lenses; and probably several other yet-unknown types of fast-changing phenomena that have not yet been observed.

To enable such studies, there is a need for Big Data and Big Compute facilities within the BRICS group to fully exploit the data resources following from transient science observations obtained with the facilities within the BRICS countries. The BRICS member countries' participation as global leaders in astrophysical transient research is expected to accelerate greatly over the next decade with the development of new facilities. A cornerstone of this will be the emergence of the LSST as a generator of transient discoveries. Two of the BRICS countries (Brazil and South Africa) are now actively participating in LSST transient science collaborations and more may still do so in the future.

The Continuum Universe

Radio continuum surveys were identified as a scientific priority for the SKA for several

reasons, including galaxy evolution, magnetism, transients, and more. Among those priorities are the understanding of the star formation history of the universe - to which MIGHTEE and SuperMIGHTEE will contribute by studying the evolution of star-forming galaxies, and of active galactic nuclei (AGN), where supermassive black holes reside, and how this affects the evolution of galaxies and their environment. Another important aspect is the potential for serendipitous discoveries. Unplanned discoveries often lead to fundamental insights into new physics. A technique called Very Long Baseline Interferometry (VLBI) is key for this last point. The technique combines signals from individual radio observatories to form a single, Earth-sized radio interferometer with an angular resolution several orders of magnitude finer than offered by connected-element interferometers. Many important results have resulted from the use of the VLBI technique, a recent example being the imaging of a black hole by the Event Horizon Telescope (Kohler 2019). One of the science goals of the FAST telescope is provide the largest aperture in international cm-wave VLBI networks, enabling unprecedented ultra-sensitive imaging. Similarly, MeerKAT will add an extremely sensitive antenna to global networks, paving the way towards SKA-VLBI (Paragi et al. 2015).

Cosmology

The SKA precursor and pathfinder telescopes are making possible new types of observations; neutral hydrogen intensity mapping surveys that do not detect individual galaxies. Combined with radio continuum surveys that detect radio galaxies out to very high redshift as proposed by LADUMA and to a certain extent MIGHTEE, and some of the ongoing and proposed programs with the uGMRT, and with LSST and optical observations able to measure cosmic distances,

cosmology is entering a new era where the large-scale structure in the universe can be mapped in three dimensions. Moreover, science collaboration between FAST and MeerKAT on the intensity mapping of neutral hydrogen will contribute to a specific cosmological measurement; that of the so-called Baryonic Acoustic Oscillation (BAO). This is fundamental, as the large-scale structure of the universe and how it came to be, of which the BAO is a keystone measure, is one of the pillars of our current understanding of cosmology.

Large projects on next-generation BRICS facilities

The scientific journey between now and SKA1 will be charted through large science programs with new SKA pathfinder and precursor facilities. These programs map directly onto the key science goals of the SKA, and prototype not only the technical and scientific capabilities of the SKA but also the growth of big data in astronomy and the new modality of global collaboration around very large projects generating vast amounts of data. At the same time developments and new projects in optical and infrared astronomy such as the LSST are set to create a rapid growth of multi-wavelength data sets.

The proposal team includes leaders of large programs on SKA pathfinder facilities and in multi-wavelength astronomy, most of which involves collaborations among researchers in the BRICS partners. The data solutions to be developed will thus advance key science projects that address the fundamental questions driving multi-national investment in megascience projects in the coming decade.

This flagship program and the complementary flagship program on transients are seen as synergistic, where together they can both advance the science and also the associated development and spin-off benefit goals. This

Table I. Current and Planned Facilities providing big data challenges for BRICS Astronomy Research.

Telescope	Location	Wavelengths	Data rate
SKA1-mid	South Africa	Radio: 350 MHz - 14 GHz	1000 TB/day
MeerKAT	South Africa	Radio: 580 MHz - 3.5 GHz	10-100 TB/day
FAST	China	Radio: 70 MHz - 3 GHz	690 TB/day
uGMRT	India	Radio: 0.1 - 1.45 GHz	10 TB/day
PRAO	Russia	Radio: 111 MHz	0.1 TB/day
AVN, VLBI	Global	Radio: 1 - 100 GHz	10 TB/day
LSST	Chile	Optical - near Infrared	20 TB/night
MeerLICHT	South Africa	Optical	0.1 TB/night

will leverage existing and planned new facilities within the BRICS countries and will also draw on the opportunities presented by other space- and ground-based facilities that exist within the BRICS group. As shown in FIgure 1 the new generation of BRICS facilities that drive the data challenge in the next decade are MeerKAT, the upgraded Giant Metrewave Radio Telescope, the Five-hundred-metre Aperture Spherical Telescope and the SKA1-mid. Parallel developments in wide-field Very Long Baseline Interferometry and mult-wavelength astronomy with the LSST provide complementary data challenges that are critical to our overall science objectives.

MeerKAT

The MeerKAT telescope is the precursor to the SKA mid-frequency array. MeerKAT was inaugurated in July 2018, and science operations have begun. MeerKAT will operate as a stand-alone South African facility for approximately five years until its 64 antennas are merged with 135 SKA antennas to form SKA1-mid. The large number of antennas, high sensitivity, and wide-field of view of MeerKAT make it the premier imaging telescope in the world for

radio frequencies in the GHz range. The MeerKAT science program includes large survey projects that will occupy about 70% of the observing time. Each program requires typically thousands of hours of observing and will, over the course of 5 years, create Petabyte scale data sets. These data sets must be converted to science products and then visualised, analysed and mined for the answers to the scientific questions. Large programs relevant to this proposal include:

■ The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) project is being undertaken by an international collaboration of researchers. The MIGHTEE observations will provide radio continuum, spectral line and polarisation information. MIGHTEE, along with multi-wavelength data, will allow a range of science to be achieved, as listed above.

■ The LADUMA (Looking At the Distant Universe with the MeerKAT Array) project will study evolution of the neutral hydrogen gas in galaxies over cosmic time spanning over two-thirds the age of the Universe. The survey strategy is to observe a single area on the sky rich in existing multi-wavelength data. The primary science goals of LADUMA

include investigating, as a function of environment and look-back time: the distribution of mass of neutral hydrogen of galaxies and how it depends on stars and dark matter in their environment and the evolution of the cosmic neutral gas density. LADUMA will be the deepest HI survey before the SKA comes online, thereby acting as a pilot project for future wider-field deep observations of neutral hydrogen to be carried out with the SKA.

■ The MeerKAT Absorption Line Survey (MALS) will carry out the most sensitive search of HI and OH absorption lines at $0 < z < 2$, the redshift range over which most of the evolution in the star formation rate density takes place. The key science themes of the survey are: (1) Evolution of atomic and molecular gas in galaxies and relationship with star formation rate density, (2) Fuelling of active galactic nucleus (AGN), AGN feedback and dust-obscured AGNs, (3) Variation of fundamental constants of physics, (4) Evolution of magnetic fields in galaxies, and (5) Physical modeling of the ISM, Astrochemistry and Cosmology. Due to the excellent sensitivity of the MeerKAT telescope, MALS will also deliver an extremely sensitive HI 21-cm emission, radio continuum and polarization survey to address a wide range of issues at the forefront of galaxy evolution research.

■ ThunderKAT is MeerKAT large survey program to detect transient radio signals. ThunderKAT comprises a comprehensive and complementary program of surveying and monitoring both galactic (within the Milky Way) and extragalactic transients such as microquasars, supernovae and possibly yet unknown transient phenomena. ThunderKAT will both survey directly and detect transients in other MeerKAT large survey projects. This commensal use of the other surveys means that the combined MeerKAT large survey projects will produce by far the largest GHz-frequency radio transient program before the SKA.

The associated MeerLICHT optical telescope will simultaneous take large field optical images for all night-time MeerKAT observations and create a large optical data set for unique and powerful multi-wavelength studies of transient phenomena.

The Upgraded Giant Metrewave Radio Telescope

The National Centre for Radio Astrophysics of the Tata Institute for Fundamental Research operates the Giant Metrewave Radio Telescope (GMRT) on the Deccan plateau north of Pune India. The Giant Metrewave Radio Telescope (GMRT) consists of 30 antennas each of 45-meter diameter, spread over a region of 25 kilometres. It is the world's largest and most powerful radio interferometer dish array in its frequency range (100 to 1450 MHz). The GMRT has undergone a major upgrade with new receivers, control, data transmission and wide band correlator systems that has substantially improved performance and increased the data rate from the array by over a factor of 10. The data rate can be as high as 100 TB/hour at the raw voltage recording mode. The upgraded GMRT is an SKA pathfinder and has the capability to demonstrate SKA kind of science with unprecedented sensitivity at low radio frequencies

There are several large programs underway with the uGMRT including search for pulsars and transients (both targeted and general surveys); precision timing of millisecond pulsars with

the aim to contribute to the global effort for the detection of gravitational waves; detailed imaging of selected target deep fields; detailed studies of galaxy clusters; large surveys in the neutral hydrogen line – both in emission and absorption; radio follow-up of supernovae, gamma-ray bursts, and also gravitational wave detections, etc.

With interferometer baselines almost four times larger than those of the MeerKAT, the uGMRT offers similar imaging angular resolution at as MeerKAT at longer wavelengths. Moreover, the increased bandwidth and receivers of the uGMRT provide similar sensitivity to MeerKAT. The uGMRT and MeerKAT together thus have tremendous synergy as complementary facilities for the study of the deep radio sky over an ultra broad band, opening a powerful new approach to studies of the distant universe.

The SuperMIGHTEE project forms part of a bilateral Indo-South African Flagship program in Astronomy, for which we have developed a technical and scientific collaboration to expand MIGHTEE to a joint MeerKAT and uGMRT project. The unique full spectral coverage and ultra-broad band nature of the superMIGHTEE data set will make it the premier data on the GHz radio properties of the deep radio sky, and will not be surpassed until well into the SKA phase 1 era.

The Hz-MALS project uses uGMRT to extend the redshift coverage of MALS to $2 < z < 5.2$. This redshift coverage is unique to uGMRT and the outcomes from the survey make it highly relevant to design future surveys with SKA.

The Five-hundred-meter Aperture Spherical Telescope

The Five-hundred-meter Aperture Spherical Telescope (FAST) is a next-generation extremely large single dish radio telescope. It is funded by the National Development and Reform Commission (NDRC) and managed by the National Astronomical Observatories (NAOC) of the Chinese Academy of Sciences (CAS), with the government of Guizhou province as a cooperation partner. Being the largest filled-aperture telescope worldwide located at a radio quiet site, the FAST science impact on astronomy will be extraordinary, and has the potential to revolutionise many other areas of the natural sciences. Compared with its precursor Arecibo, FAST has an advantage of a factor of two in raw sensitivity and a factor of five to ten in surveying speed. FAST will also cover two to three times more sky area thanks to its innovative design of an active primary surface. A science instrument with an order of magnitude improvement in any of its capacities, which is also able to explore new dimensions in parameter space, is likely to generate unexpected discoveries.

FAST produced its first light in September 2016 and following a commissioning, phase normal operation to commenced in late 2019. The science programs of FAST mainly include a large scale neutral hydrogen survey, pulsar observations, leading the international very long baseline interferometry (VLBI) network, detection of interstellar molecules, detecting interstellar communication signals, and pulsar timing arrays. FAST is currently the largest single dish telescope, while MeerKAT is the largest aperture synthesis array telescope.

Science collaboration between the two could probe pulsars, exotic events, the low HI surface brightness, low mass galaxies and the HI intensity mapping for cosmological BAO measurement. The combined surveys may also allow us to study low-frequency gravitational waves using the high precision pulsar timing, massive black holes, interacting galaxies in galaxy pairs and in groups of galaxies. FAST is facing a similar situation of big data processing.

The required data transmission rate of FAST is at least 6GB/s, which represents a big challenge for data transfer and processing.

Very Long Baseline Interferometry

Among other things, VLBI enables the study of AGNs and black holes, magnetic fields, galaxy formation, or even cosmological dark energy models. In this new survey-driven era, it is vital that global VLBI networks continue to increase their sensitivity and observing capabilities to enable high resolution followup, discovery, and statistical analyses at radio wavelengths. This is of particular importance in the realization of SKA1-mid, at which point VLBI is in effect merged into a standard connected-element radio interferometer. SKA1-mid will thus carry out high resolution wide-field surveys by design. The increasing wide-field VLBI capability poses significant big data and big compute challenges.

VLBI has been at the heart of the SKA since its conceptualisation. One of the key science objectives of FAST is to lead the international VLBI network. Shanghai Astronomical Observatory, Chinese Academy of Sciences (SHAO) have been developing VLBI techniques since 1972. The longest baseline of the Chinese VLBI Network is 3249 km. With wide-field VLBI being so closely tied with developments towards the SKA, as well as VLBI itself being inter-continental in nature, it fits very naturally into the goals and aspirations of this BRICS proposal.

Large Synoptic Survey Telescope

The full science goals of the SKA will only be achieved in synergy with other instruments. The LSST will create a data set of multi-wavelength information for a vast number of cosmic radiation sources that can be mined jointly with radio survey data. The LSST will also transform time domain astronomy by detecting millions of transient and variable phenomena each night. LSST will be the next big data generator of optical transients, estimated to be of order several million detections of new transients or variability per night. Analysing this enormous quantity of data to allow rapid follow-up of transients with other telescopes will require significant compute infrastructure, as well as novel algorithms. The parallel proposal for a dedicated BRICS-wide Flagship program on transients shares many of the Big Data and Big Compute challenges discussed in this proposal in relation to the SKA. The parallel proposal aims to develop a network of ground-based optical telescopes for the follow-up of astrophysical multi-wavelength and multi-messenger transient objects. The demands for data handling associated with transient studies amongst the BRICS collaboration will be met by the implementation of this Big Data Flagship program.

PROJECT OBJECTIVES

The objectives of the proposed research are to develop and foster collaborations between researchers among the five BRICS partner countries to address the data challenges of the coming era. Specifically we have the following goals:

■ Establish a prototype network of federated cloud-based infrastructure and e-science tools to support the next-generation data intensive radio astronomy to serve our national research communities, and exploit synergies between partners and opportunities for enhanced research through joint big data astronomy science programs.

■ Create new technologies and modalities for visualisation and machine learning

enhanced visual analytics of cloud based remote big data, as well as cloud-based tools for collaborative exploration by distributed research teams.

■ Develop cloud-based provisioning of HPC for automated processing workflows for use by science teams. Provisioning flexible containerised analytics environments such as Jupyter notebook allowing interactive engagement with the data for user applications. Such platform and systems will empower research teams to work with big data. Such workflow systems will address the challenges of reproducible science with big data.

■ Develop cloud-enabled systems and environments for the fusion of multi-wavelength data sets such as the LSST with radio data, and for multi-messenger and transient science, including a cloud-based LSST data capacity in South Africa to host and analyse LSST transient events. An objective here is to develop and incorporate machine learning to analyse data and enable us to characterise, classify (and discover) the unknown and prioritise rapid response programs.

■ Develop scalable knowledge bases and data lakes to provide access to integrated and correlated knowledge transforming raw data into scientific knowledge on a global scale.

Data intensive cloud architectures: Teams within IDIA (South Africa) will work closely with researchers based in India (NCRA-TIFR), China (Tianjin University, SHAO, Alibaba Cloud) and Brazil (LNCC) to build open-source based architecture and tools to enable deployment of

HPC clusters within clouds systems. Areas being worked on include:

■ Elastic scheduling and workflow tools for OpenStack and CloudStack

■ Distributed role based authorization to resources.

■ Scheduling of data movement between sites.

■ Scheduling of compute jobs to sites.

■ Use of containerisation to support movement of applications and reproducible science.

Teams in South Africa (IDIA) and China (NAOC, SHAO, Alibaba Cloud) are also working on science gateway and portal technologies for user access to data products, to cloud processing and analytics.

Algorithms and Machine Learning for Big Data Processing and Analytics Perhaps the largest area of collaboration lies in the development of novel algorithms and HPC software for processing and analysis of big (radio) astronomy data, and we include the following subsections for collaborative purposes:

■ processing pipelines for raw data to science products, including development and integration of workflow systems, research data management and reproducible science technologies (South Africa: IDIA, Rhodes University; India: NCRA-TIFR, IUCAA, IISER; Brazil: LNCC, LIneA; China: NAOC, Tianjin University, Guangzhou University, Russia). We aim to leverage the scientific and technical skills from our partners to further develop workflows and recipes for calibration and imaging pipelines.

In addition, we will test and develop cloud interfaces to allow for seamless interaction

with radio astronomy software, with the aim to facilitate collaborative development for pipeline design and data inspection.

■ algorithms for mining big data for information, including machine learning (South Africa: IDIA, SARAO; Brazil: LNCC, LIneA; India: NCRA-TIFR, IUCAA, SPPU; China: NAOC, Tianjin University, Guangzhou University, Russia). We will use the techniques developed by e.g. the Spitzer Data Fusion and the HELP EC-FP7-SPACE projects to merge multi-wavelength surveys and develop machine learning tools to best exploit the wealth of multi-wavelength data available in upcoming radio deep survey fields and enable new discovery space.

■ data systems and distributed knowledge bases for management for fusion and mining of diverse data sets. This work will be lead by LNCC advancing their developments for LSST.

■ The novel processing challenges and extraordinary image sizes emerging from the field of wide-field VLBI presents a new type of problem that foreshadows the SKA (South Africa: IDIA, SARAO; China: SHAO, FAST; India, Russia). We will interface disparate pieces of custom-built software into a more general, multi-purpose VLBI survey toolkit. A component that will require new development, based on an existing pilot project, will be the area of VLBI source-finding. At present, this does not include Bayesian inference and machine-learning that are relatively straightforward extensions of the pilot source-finding project.

Cloud-based Visual Analytics and Exploration of Big Data Cloud-based visual analytics of remote big data (South Africa: IDIA; China: SHAO, Alibaba Cloud, India: IUCAA). This will expand on an initial framework being developed at IDIA for low-latency cloud server-client visualization of remote big data including 2D and 3d (virtual reality) interfaces. We will expand on a new hdf5 image data set framework that is optimized for big data and develop cloud provisioned HPC server-side rendering technologies and tools that interface to multiple remote clients. We will design interfaces and tools sets for visual analytics of big data sets from the cloud from MeerKAT, uGMRT and FAST in response to use cases that meet the needs of the BRICS data intensive programs.

HCD and Development
Human capital development (HCD) is a key aspect of the project. An integral part of the new approach to science driven by big data, big compute and large international collaborations is to create and embed the platforms within the research communities so that researchers may work collaboratively on these large, diverse data sets. Such platforms are best developed through an extensive, international program such as the one proposed here: it promotes sustained collaboration between partners from the BRICS countries, and it prevents duplication of effort in developing such platforms in individual countries.

The project will develop the required training resources that can run in the cloud and give access to large data sets. The HCD and education platform development and maintenance needs to be supported by staff members embedded in the partner organisations that develop and maintain such infrastructure for research. Material for running workshops will be made openly available, and

referenced extensively through the proposal participants' global networks.

By developing a dedicated platform to run training workshops, data science hackathons and research schools that operates in the cloud, the HCD and skills development efforts of this project will be greatly enhanced. It will be possible to multiply this model across BRICS countries, in smaller institutions as described above, and beyond. Relevant experience of proposal participants includes the support by IDIA of DARA Big Data Africa research schools and hackathons (https://idia.ac.za/BigDataAfrica-2018/), the Astronomy Data Science Toolkit developed by the IAU OAD (http://datascience.astro4dev.org), and the expertise of our Chinese collaborators (NAOC, Alibaba Cloud).

ALIGNMENT WITH NATIONAL STRATEGIES

South Africa

The South African government White Paper on Science Technology and Innovation (STI) envisions STI enabling inclusive and sustainable South African development in a changing world. This proposal addresses two of the three high-level goals spelled out in the White Paper, namely to take advantage of opportunities presented by megatrends and technological change, and to contribute to a more inclusive economy at all levels. The grand data challenge in astronomy is part of the magatrend characterised as the fourth industrial revolution, and the integrated HCD strategy is specifically aimed at making the development of skills high in demand more inclusive, and to facilitate the translation of highly skilled young scientists into the economy. Recognising that STI can contribute to the Sustainable Development Goals, the White Paper highlights the importance of:

exploiting the pivotal role of ICT and harnessing the potential of big data; exploiting the full potential of scientific knowledge and improving the performance of historically disadvantaged universities; focusing on inter-disciplinary research; expanding research infrastructure, including cyber-infrastructure; and exploiting the potential of STI for African development and continental integration.

This research project is closely aligned with the SA National Strategy for Multiwavelength Astronomy, building on the large investment South Africa has made in radio astronomy (through MeerKAT and the SKA), as well as at optical wavelengths (SALT, LSST). This proposal seeks to build capacity in the development of solutions in processing, analysing, mining, visualising and maximising the science coming from increasing large, rich and complex multi-wavelength data sets. By bringing the LSST data to the SKA-mid at an SA LSST data centre, we will enable multi-wavelength studies and increase the impact of SKA science. In addition, ensuring access to the LSST alert stream will allow the full use of the rich range of telescopes within South Africa for transient follow-up.

Development of the hardware, software and human capacity within the field of VLBI is a key part of SARAO and DST?s strategy to develop the African VLBI Network and the expertise within the SKA African partner countries to become world-class users of this instrument. Furthermore, this is directly aligned with the strategic goal of the AVN to form the backbone of the SKA Phase 2 to be built across the African continent.

This project aligns with the National Strategy for Human Capacity Development for Research, Innovation and Scholarships through the proposed collaborations on globally competitive research and innovation. These strategic international engagements will

enhance the research and technological capacity for innovation within South Africa, as well as ensure the establishment and maintenance of research infrastructure and platforms. In this way, the project is also aligned to the Strategic Research Infrastructure Roadmap.

This project helps advance the SA National Integrated Cyberinfrastructure System (NICIS). It leverages the investment in the DIRISA Tier 2 data intensive research facility to work with international partners on data intensive challenges that will be applicable to cyberinfrastructure solutions in SA for astronomy and other disciplines such as bioinformatics.

Brazil

Brazil is engaged in establishing data centers in support to eScience. The initiative includes HPC infrastructures, with the prominent availability of the Santos Dumont supercomputer services available for the whole scientific community and currently hosting more than 100 scientific projects. In a complementary way, the RNP (National Research Network) is fostering the improvement of links within Brazil and with Africa, Europe and the US. In particular the Monet cable has been installed linking Chile to Miami, and passing by São Paulo enabling the transfer of data produced by the LSST project. Additionally, the SAIL and SACS cables now link Brazil to Africa through Angola and Cameroon, respectively, improving the connection with the African continent countries, including South Africa. Brazil is intensively involved in important astronomy surveys. The LIneA laboratory hosts a tertiary site of the Sloan SDSS III collaboration providing a replica of the SDSS portal with access to published data from the Latin America Community. Brazil additionally, supports the Dark Energy Survey (DES) project, through the DES-Brazil consortium, making available a

scientific portal for the publication of DES data and the execution of scientific pipelines. Brazil is also participating in the LSST project, through the Brazilian Participation group (BPG), led by LIneA and the National Laboratory of Astronomy (LNA). Through these different initiatives, Brazil is contributing to the development of astronomy in South America through the development of Big Data software and the availability of HPC and network infrastructure. Moreover, it is investing in the education of multi-disciplinary human resources founded in basic science and computer science.

China

China is one of the prominent participants in SKA. SKA global collaboration is an important task for building a community of common destiny for mankind, as proposed by the Chinese government. The Chinese government put large investment in technology research, data center and human capital development. Chinese scientists have been working closely with members from other countries preparing for the technology and science using SKA data, and are involved in 13 associated science working groups and focus groups of SKA. The China SKA science team work together with the information, communication and computer industries aiming at tackling the challenges associated with the SKA big data, which will not only promote major original scientific discoveries, but also apply the obtained technological achievements for stimulating the national economy. This project is perfectly in step with the goals of China SKA team.

Exploring the synergy between FAST and the MeerKAT/SKA projects is part of the research content of this project, and will open up new research areas for scientific users of both FAST and SKA data. With the improvements in sensitivity and resolution by the combination

of FAST and MeerKAT, the survey area of their common sky is still quite large. The science cooperation between the largest single dish telescope, FAST, and the largest aperture synthesis array telescope, MeerKAT/SKA, in the world, will result in detection of hundreds or even thousands of new pulsars and some exotic events, as well as allowing the study in details of the exotic events and the HI gas in nearby galaxies. This is also highly consistent with the scientific goal of FAST.

National Astronomical Data Center (NADC) of China is recently established based in National Astronomical Observatories, CAS (NAOC) on the purpose of managing astronomical data at the national level and providing data services and technologies for the whole life cycle of data management, including outreach and education. This project works on cloud-based computing environments and scientific platforms for data users, and on data techniques including machine learning, visualisation and data mining, which is aligned with NADC's goal.

India

The Giant Metrewave Radio Telescope is a SKA pathfinder facility. The recently completed upgrade of the GMRT has further enhanced the front-line capabilities of the GMRT, and several exciting new discoveries and results are being produced. It has also resulted in an order of magnitude increase in the data rates from the observatory, making it imperative to adopt big data intensive techniques, and improved, efficient data analysis pipelines to process the data. In order to maximise the science returns from the uGMRT, NCRA plans to implement imaging pipelines to create science ready products for the benefit of users who do not have the resource to handle large data. There is significant scope for new developments, and these are on the road-map going forward.

India is a full partner in the SKA, and NCRA-TIFR is the nodal institute leading the Indian participation in the SKA. Indian scientists and engineers lead by NCRA, with significant participation from Indian industry, are involved in several work packages of SKA. In particular, NCRA is leading the Telescope Manager (TM) work package, which is the brain and nervous system of the SKA. In addition, researchers from India are involved in significant roles in projects on SKA precursor facilities like MALS on MeerKAT in South Africa and Solar science on MWA in Australia, many of which involve the need to handle large volumes of data. India also plans to set-up a SKA regional data centre to cater to the specific needs of storing SKA data and developing specialised analysis pipelines that are of benefit to Indian researchers involved with key science projects of the SKA.

Finally, much of the above activity envisaged in India is being carried out in collaboration with industry partners, many of whom bring new ideas to the table and are willing to (and are well equipped to) take on the challenges of big data analytics that are entailed in these ongoing and planned activities. This includes the exciting new field of machine learning which is finding increased role in the plans of researchers from India.

Given the above, this project is very strongly aligned with both institutional and national level strategies and plans. Indian researchers and institutions will be able to contribute meaningfully to this initiative, and also stand to benefit significantly in their efforts on existing and planned Indian and other BRICS member facilities. It will also allow Indian industry to engage and grow in these high-tech areas with a view to becoming a major player in the global arena. The goals and activities of this collaborative project also align well with India's plans for effective participation in

the construction phase of the SKA, including expediting the work on setting up of a SKA regional data centre in India, and enhancing meaningful partnership with BRICS members participating in the SKA.

Russia

In the future of multi-wavelength, multi-messenger astrophysics, it is important to develop the ability to organise timely, coordinated observations across multiple facilities with ever increasing volumes of observational data, as well as the technical capabilities and expertise to efficiently sort, search and analyze large sets of heterogeneous observations. Russia operates, or is involved with, many observational facilities, including numerous ground-based optical and radio telescopes, the BDUNT and BNO neutrino telescopes, the Russian-German Spektr-RG space-based X-ray observatory, two international networks of optical telescopes (MASTER and ISON), and the proposed BRICS-based network of optical telescopes. Besides the technical requirements outlined above, the fusion of these many astrophysical data sets also requires intensive training of new specialists and human capital development, with a particular focus on the role of exchange between the BRICS countries. Besides direct application to the field of astrophysics, the knowledge and techniques developed throughout the course of the project are a high priority for other fields faced with similar challenges, including geophysics and climate research, telecommunications, energy efficiency, material sciences and medicine, among others.

BENEFITS BEYOND ASTRONOMY

Economic impact: collaborations with industry

Building and managing a SKA regional centre is a formidable challenge due to its unprecedented

size and scale. The development of the research infrastructure and algorithms needed to tackle the big data and big compute challenges addressed in this proposal lends itself to joint research and development projects with industry for the local, regional and global markets. A multi country collaboration, involving industry partners is critically needed for this activity. A successful collaboration will lead to the development of human and technical capacity within academia and industry in all BRICS countries. For example, India has an ongoing engagement with industry in several areas. NCRA is already in collaboration with some industry partners like Thoughtworks India Ltd, Persistent Systems Ltd and TCS Research on handling large data, machine learning and other technical aspects with regard to SKA regional data centers. In Brazil Petrobras and DELL Brazil have expressed interest in the project. In South Africa, there are ongoing conversations with IBM Research - Africa and SAC, a local Space Engineering company. Underpinning the growth of industrial and economic activity is also the need for trained talent. The inspirational aspect of astronomy attracts young people to science and technology fields but not all will become astronomers. Engaging in those studies, however, leads to the acquisition of skills that are in high demand in industry. This flagship will play a key role in making that talent available for the growth of industry in the BRICS nations.

Revealing and developing untapped talent

BRICS countries are characterised by a growing young population hungry for opportunities. HCD will take place through graduate study support in the form of scholarships and participation in the research and innovation that this flagship will drive among other things. As part of their training, young researchers supported by this network will learn communication skills and the

international aspect of the project will equip them with multicultural experiences critical to being competitive in a global workforce. The joint experience of the proposal partners ensures a successful implementation of this approach to HCD.

The benefits of an associated HCD program will be felt wider than astronomy and computing. Students and early career researchers have the technical and problem-solving skills that are vital to many sectors of a knowledge-driven economy, and that are essential to using big data to address some of the most pressing development issues. The exponential growth in data science and technology careers internationally means that research and collaboration in the areas of cloud-based and high performance computing have high potential for impact on human capital development. This needs to be proactively nurtured as proposed in this flagship project.

Access to research for smaller institutions

In BRICS countries, many smaller research institutions and universities are not currently competitive on a global scale of scientific research. Many new researchers and students will have access to big data and big compute, through the gateway technologies and cloud-based toolkit we propose to develop in this proposal. For example, researchers based at the IAU/OAD, University of Zululand and IDIA (South Africa) will collaborate with researchers in China (NAOC), India (NCRA-TIFR, IUCAA) and Brazil (LNCC). Access will be established through training workshops at institutions and through the support of participation of students from those institutions in research activities of the project. In addition to this, it offers the opportunity to train IT staff and researchers in new technology.

The inclusive participation of all research institutions in the BRICS countries needs to be promoted. The international nature of this flagship guarantees that such efforts are not in isolation and offer collaboration opportunities previously not easily accessible for those institutions. In South Africa for example, this means access for historically disadvantaged institutions with the support of well established institutions. In China, it may mean that smaller universities in that vast country can join the endeavour. The experience of the IAU Office of Astronomy for Development (OAD), the South African Radio Astronomy Observatory (SARAO) and the Inter-University Institute for Data Intensive Astronomy (IDIA) will ensure effective inclusion of smaller institutions.

Embedding young researchers in global networks

Young researchers engaged in this broad collaboration among leading institutions in BRICS partner countries will experience exciting career opportunities. The will enhance the BRICS context as sources of scientific and technical talent, and will strengthen the BRICS countries' leadership in global science through investment in strategic areas where BRICS displays particular strength, such as radio astronomy. By developing big data and big compute expertise and ensuring participation of early-career researchers in this flagship, this investment is guaranteed to pay off considering the roles those play in the development of economies in the context of the 4th industrial revolution.

Enabling young scientists to work on interdisciplinary projects on research infrastructure is an important element to facilitate the beneficial dissemination of trained big data and big compute scientists across other sectors and raises the participants' awareness of how they can contribute to development. Such

experiences are also highly valuable for young scientists to gain employment in industry as it develops soft skills that cannot be acquired strictly through formal astronomy or computer science training.

Public Awareness of Science

This flagship will contribute to megascience projects like the SKA and the LSST, in which BRICS countries have made significant investments. The benefits of those investments need to be effectively communicated to the public as they are the major source of support. Outreach is widely recognised as a key development instrument to: create awareness of, and popular support for science programs and scientific research; stimulate pride and develop trust in the national science systems; constructively engage stakeholders at their level; empower young scientists with strong communication skills for multifaceted careers; influence young people's study and career choices. Outreach efforts will pay particular attention to future scientists: undergraduate students and high-school learners, and be relatable to them and their immediate circles.

Initiatives are planned to develop outreach resources showcasing the facilities, science collaborations and technology developments in the project; to promote young scientists as role models and their studies as leading to viable and opportunity-rich career choices; to offer practical outreach and communication training for the students and researchers involved

The proposal members have strong experience in public outreach and well established dissemination networks to work in partnership with. Planetariums can immerse the public in real scientific data on digital domes and help generate public support for science, as carried out in South Africa at the Iziko Planetarium and Digital Dome. The Chinese

efforts to develop planetariums and science museums like the special planetarium and science museum themed "Astronomy & Society Development" in Xinchang attract large numbers of visitors. In India, both IUCAA and NCRA-TIFR carry out extensive public outreach activities aimed at students and the general public with tens of thousands of visitors on the National Science Day.

PROJECT MANAGEMENT AND COSTING

Governance and Management Structure

The overall responsibility for the success of the project, oversight of its programs, development of policies and budget allocations will be through a Project Executive Committee comprised of the project leads from each country. The Project Executive will be the reporting body to the BRICS funding authority. The Executive Committee will meet quarterly or more often as necessary to provide effective oversight and direction.

We will form working groups in thematic areas, including for example:

■ cloud computing developments and platforms

■ AI and machine learning

■ ontologies and knowledge bases

■ data intensive VLBI

■ Outreach/Education and Development

■ data processing and analytics pipelines and workflows

■ visualization, visual analytics and exploration.

Each thematic working group will have two co-chairs from different participating countries. The working group chairs will coordinate

activities and foster collaborations among the participants in each area.

The working group chairs and the Project Executive together will constitute a project management team to provide oversight for the overall management of the project and programs and advises the Project Executive.

The day-to-day management of the project will be carried out by a dedicated project manager who will report to the management team.

Project and Program Development and Tracking

At the core of the project is the development of a federated cloud platform to enable and support distributed data intensive research and global collaboration on big data. Since this platform is critical to other program elements, we will focus on implementation of the federated cloud as an early priority with intent to have a test system in place in 2021 for early use. This system will continue to develop and evolve over the course of the program in response to the big data needs of the project. A dedicated technical team consisting of one person at the location of cloud node in each country will be responsible to the central project and support development of the cloud technology and platforms, and software and systems to support the data intensive projects and users.

The majority of data intensive research and development work to be carried out under this project will be in the form of collaborative initiatives between partner countries within thematic working groups using the BRICS federated data intensive astronomy cloud. We will support working group workshops for formulation and detailed planning of collaborative programs. To advance the work and reinforce collaborations we will provide postdoctoral and postgraduate bursaries for

high priority programs. The projects to be supported with bursaries will be identified by the working group co-chairs and recommended to the Executive.

We will support hosting and travel to an annual project meeting involving all participants. This meeting will rotate among the partner countries, so that over the first phase of the project each country will have an opportunity to host. This meeting will include progress reports, detailed planning, exploration of new developments and directions, workshoping, new collaborative opportunities, cross-fertilization and individual networking, and discussion of annual priorities and objectives. We will hold an associated cloud-based big data hackathon, training and outreach event.

Budget

The budget for the BRICSKA program consists of funds to support:

■ *BRICSKA annual project meeting.*

Cost for hosting an annual project meeting which will rotate among the member countries.

■ *Travel for research collaboration meetings and targeted workshops*

One international and one national trip per annum per fellowship.

Travel for co-investigator, postdocs, students and technical staff to attend the annual project meeting.

Travel for extended visits by individual researchers to collaborators.

Travel for small local workshops in thematic areas.

■ *BRICS post-doctoral and post-graduate fellowships.*

We propose six BRICS postdoctoral fellowships, one for each of the thematic areas. Plus twelve postgraduate PhD fellowships. These individuals will work on astronomy-driven data science projects that advance use of the cloud platforms and software systems developed as part of the BRICSKA collaboration.

These positions will serve to not only further the data-intensive astronomy programs but also to reinforce collaboration among the partners. The postdoctoral fellowships will be joint appointments at institutions from at least two participating countries. The postgraduate fellows will be jointly supervised by co-investigators from at least two partner countries.

■ *Equipment and Infrastructure*

Minor equipment budget for postdoctoral and postgraduate fellows. Major computing will be supported by federated cloud resources at participating data centers. Each will be provided with a laptop.

In 2023 and 2025 we request minor infrastructure funds to support LSST data storage (2 PB each) and analysis at the South Africa data intensive research cloud.

■ *Project management and software and system development.*

A project manager will be appointed at the Inter-University Institute for Data Intensive Astronomy in South Africa.

We propose five technical appointments, one at a participating data centers from each partner country. This individual will be responsible to the central project to support development of the federated cloud environment, and support deployment of environments, platforms

and software systems for use of the federated cloud for collaborative BRICSKA science programs.

■ Operations

A 50% appointment to support administrative and logistical needs of the project.

■ Outreach and Astronomy for Development Support

A 50% position for an outreach and development program coordinator. One technical support position to develop and support cloud-based platforms for the diverse outreach and development programs and to serve as a technical lead for implementation at the federated nodes, and support of projects and programs among the BRICS partners.

The budget amounts for each category by year for the first six years are provided in the Table II. Amounts are in currency of Euros. The growth of data in astronomy will only continue to increase exponentially for the foreseeable future. We thus propose that this project be established for an initial 10-year period. The budget requests start-up funding in the first year to launch the project followed by a project budget for the first five year phase starting in 2022. The budget is inflated by 5% per annum starting in 2023 to account for inflation. The total amount for the first 6 years is 4.42M Euro.

For the first year of the project, 2021, we budget funds only for travel, for hosting the BRICSKA project meeting, and for 50% salaries of a project manager and administrative and logistics support. These appointments would be made mid-year in time for a start up meeting. At the start-up meeting we would implement the project governance and management structure, create detailed plans and agreements for

Table II

Budget (Euros)						
	2021	**2022**	**2023**	**2024**	**2025**	**2026**
Host Annual Meeting	6,000	6,300	6,615	6,946	7,293	7,658
Travel						
Annual Meeting	91,200	127,200	133,560	140,238	147,250	154,612
Collaboration visits & thematic Workshops		24,000	25,200	26,460	27,783	29,172
HCD						
Postdoc Fellowships		111,995	117,595	123,475	129,649	136,131
Graduate Fellowships		104,723	109,960	115,458	121,230	127,292
Staff						
Project Manager	23,240	46,480	48,804	51,244	53,806	56,497
Technical		232,400	244,020	256,221	269,032	282,484
Admin & Logistics	5,810	11,620	12,201	12,811	13,452	14,124
Outreach & Development						
Coordinator		23,240	24,402	25,622	26,903	28,248
Technical		46,480	48,804	51,244	53,806	56,497
Interventions		11,000	11,550	12,128	12,734	13,371
Infrastructure		26,145		95,284		60,424
Total	**126,250**	**771,584**	**782,711**	**917,130**	**862,938**	**966,509**

programs and activities in the thematic areas, and finalize projects for initial post-doctoral and post-graduate appointments. Funds for the full projects, including fellowships and appointments of technical staff would be required beginning in the second year.

The staff identified in the budget could be employed by institutions in the BRICS partner countries. The project management and administrative support staff would be hosted at IDIA partner universities, while the technical staff would be hosted at data centers participating in the BRICS federated astronomy cloud. These budget elements may be most effectively funded by direct allocations within the applicable partner countries.

REFERENCES

AN T. 2019. Science opportunities and challenges associated with SKA big data. Science China

Physics, Mechanics, and Astronomy 62(8): 989531. doi:10.1007/s11433-018-9360-x.

BLYTH S ET AL. 2016. LADUMA: Looking at the Distant Universe with the MeerKAT Array. In: MeerKAT Science: On the Pathway to the SKA, p. 4.

FENDER R ET AL. 2016. ThunderKAT: The MeerKAT Large Survey Project for Image-Plane Radio Transients. In: MeerKAT Science: On the Pathway to the SKA. p. 13.

GIBNEY E. 2019. Gigantic Chinese telescope opens to astronomers worldwide. Nature 574(7776): 16-17. doi:10.1038/d41586-019-02790-3.

GUPTA N ET AL. 2016. The MeerKAT Absorption Line Survey (MALS). In: MeerKAT Science: On the Pathway to the SKA, p. 14.

JARVIS M ET AL. 2016. The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. In: MeerKAT Science: On the Pathway to the SKA, p. 6.

KIM S ET AL. 2017. ALMA and GMRT Constraints on the Off-axis Gamma-Ray Burst 170817A from the Binary Neutron Star Merger GW170817. ApJ, Letters 850(2): L21. doi:10.3847/2041-8213/aa970b.

KOHLER S. 2019. First Images of a Black Hole from the Event Horizon Telescope. AAS Nova Highlights.

PARAGI Z ET AL. 2015. Very Long Baseline Interferometry with the SKA. In: Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 143.

TAYLOR AR. 2019. High Sensitivity Imaging with MeerKAT and uGMRT: Exploring the Deep GHz Universe. In: 2019 URSI Asia-Pacific Radio Science Conference (AP-RASC), p. 1-1. doi:10.23919/URSIAP-RASC.2019.8738702.

RUSS TAYLOR[1,2]
https://orcid.org/0000-0001-9885-0676

FABIO PORTO[3]
https://orcid.org/0000-0002-4597-4832

CHENZHOU CUI[4]
https://orcid.org/0000-0002-7456-1826

YOGESH WADADEKAR[5]
https://orcid.org/0000-0002-1345-7371

OLEG MALKOV[6]
https://orcid.org/0000-0002-1558-4198

[1]University of Cape Town, Inter-University Institute for Data Intensive Astronomy, Department of Astronomy, Private Bag X3, 7701, Rondebosch, Cape Town, South Africa
[2]Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Rd, Bellville, 7535, South Africa
[3]National Laboratory for Scientific Computing, Av. Getúlio Vargas, 333, Quitandinha, 25651-075 Petrópolis, RJ, Brazil
[4]National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing, 100101, China
[5]National Centre for Radio Astrophysics, Post Bag 3, Ganeshkhind, Pune, India
[6]Institute of Astronomy, 48 Pyatnitskaya, St., Moscow, 119017, Russia

Correspondence to: **Russ Taylor**

E-mail: russ@idia.ac.za

Author contributions

As proposed for the BRICSKA project, this paper is a collaboration among the BRICS collaborators. The content was co-ordinated by R. Taylor based on country-specific contributions from the co-authors.