



Communication and Marketing Department
Isebe loThungelwano neNtengiso
Kommunikasie en Bemerkingsdepartement

Private Bag X3, Rondebosch 7701, South Africa
Welgelegen House, Chapel Road Extension, Rosebank, Cape Town
Tel: +27 (0) 21 650 5427/5428/5674 Fax: +27 (0) 21 650 5628

www.uct.ac.za

16 November 2015

Key SA-Dutch SKA data science partnership to be signed

The visit to South Africa by Netherlands Prime Minister Mark Rutte will include an important SKA data science partnership agreement between South African and Dutch companies and research organisations.

A key to unlocking the secrets hidden in the immense volume of data generated by the world's biggest radio telescope – SKA (the Square Kilometre Array telescope) – will be realised on Tuesday, 17 November, with the signing of a data science partnership between South Africa and the Netherlands.

The agreement is part of the visit to South Africa by the Prime Minister of the Netherlands, Mr Mark Rutte, and his trade delegation of 75 companies. SKA South Africa and the University of Cape Town, through the Inter-University Institute for Data Intensive Astronomy (IDIA), will sign a Memorandum of Understanding with fellow research institutions in the Netherlands, IBM and ASTRON to collaborate in a ground-breaking research project entitled: "Precursor Regional Science Data Centres for the SKA (SKA-RSDC)". The MoU will be celebrated as part of 'House of the Future' – a programme of workshops, seminars, presentations and round tables with South African and Dutch stakeholders, taking place from 16 to 20 November in Turbine Hall, Johannesburg.

The South Africa-Dutch agreement on data science aims to establish national and regional data centres – in both countries – to tackle one of the most significant challenges presented by SKA: how to manage, process, and make accessible the immense amount of data the telescope will generate. The data centres will provide astronomers around the world with access to the large-scale data infrastructures and associated high performance computing (HPC) needed to make sense of the data.

"We assume that there will be at least two astronomy-focused sites, one each in South Africa and Netherlands", says Professor Russ Taylor, IDIA founding director and joint UCT/UWC SKA Research Chair.

"The initial focus of the centres will be to service the current and future data archiving, distribution and science exploration needs of the MeerKAT and LOFAR radio telescopes in SA and the Netherlands, respectively. The activity, combining both operational and research

components, is an important step on the path towards being able to efficiently extract major science value from the massive astronomical datasets which will be collected by the SKA", says Dr Jasper Horrell, General Manager: Science Computing and Innovation at SKA South Africa.

The techniques developed can, in turn, be applied in other fields such as big data analytics, high performance computing, green computing, and visualisation analytics.

ENDS

NOTES FOR EDITORS

- How big is the data challenge
- How the SA-Dutch data centres will work
- More about the collaboration

1. The Data Challenge for SKA

SKA (Square Kilometre Array) will be the world's largest radio telescope – a hundred times bigger than any current radio telescope; it will revolutionise our understanding of the Universe. SKA will be built in two phases – SKA1 and SKA2 – starting in 2018. SKA 1 will include two instruments – SKA1 MID (to be built in South Africa) and SKA1 LOW (to be built in Australia); they will observe the Universe at different frequencies.

As astronomy has developed it has become increasingly clear that the old ways of working with data no longer apply. Precursors to SKA (telescopes on one of the two SKA sites) and Pathfinders (SKA-related technology, science and operations activities) have ushered in an era of data-intensive astronomy. One such Pathfinder – LOFAR (the Low Frequency Array telescope built by ASTRON, the Netherlands Institute for Radio Astronomy) has a data collection exceeding 20 petabytes. As a frame of reference, it would take about 2 000 years to play only one petabyte of average-length MP3-encoded songs.

With South Africa's MeerKAT and the Netherlands' Apertif (APERture Tile in Focus) telescopes both expected to come online in 2016, the scale of such data collection is poised to increase significantly. The large scale of the datasets and the requirements of the astronomers to perform complex scientific analyses, which are often compute-intensive, demand innovative approaches. Data at these scales present unique challenges not just for managing the collection, but also for how researchers extract their science.

In all of these Precursor and Pathfinder facilities, the data gathering and initial processing is done onsite – close to the instruments themselves – under the control and development of the core telescope project teams. However, to make as much sense of, and derive as much value from, the data as possible, this first level of data must be made available to a broader scientific community; hence the need to develop innovative ways to access, manage and process the data. This is what the SA-Dutch regional science and data centres (RSDCs) hope to realise.

2. How RSDCs could assist, including more technical detail

With previous traditional radio facilities, the majority of this subsequent analysis, including providing the required processing and storage resources, has been the responsibility of the individual researcher or science team. However, at the data scales of the SKA, this is unfeasible. One way to address this issue is to establish national or regional centres to provide users with access to the large-scale, High Performance Computers (HPC) infrastructure they will require to extract the full range of SKA science.

The main aims of these Regional Science and Data Centres (RSDCs) will be:

- to ensure that the data collected by the instruments is well-curated and made available in an easily accessible way to downstream science processing centres and institutes;
- to maintain long-term archives of science data products;
- to provide sufficient extra compute and storage for researchers to be able to reprocess data, perform customised analysis, and visualise results without having to necessarily move all the data to their local facilities;
- to provide expert support to users with their specific analyses;
- to develop and maintain new tools and functionality to increase scientific exploitation of the data collections;
- to provide mechanisms for security and federation.

The SA-Dutch collaboration will consider a multi-tier model, similar to CERN, where the core SKA telescope facilities would produce the initial data streams, while the RSDCs would provide sufficient resources to store subsets of the SKA archive, support significant processing and post-processing capability, and further distribute data to users and smaller sites. An alternative could be a more dynamic multi-cache approach where data is distributed in a demand driven flexible way. Given their likely scale and range of function, it is natural that these RSDC facilities may hold data from more than a single instrument, or even multiple disciplines.

3. South Africa, Dutch collaboration to establish RSDCs

Where possible the work will build on the knowledge and holistic models already developed as part of the DOME Project – a Dutch-government funded project between ASTRON and IBM. Precursor instruments like MeerKAT and Pathfinder facilities like LOFAR will provide an opportunity to tackle specific, realistic problems.

“We propose to establish a federated system that links the MeerKAT, LOFAR, and Apertif science archives with distributed RSDC facilities. We assume that there will be at least two astronomy-focused sites, one each in South Africa and Netherlands”, says Professor Russ Taylor, IDIA founding director and joint UCT/UWC SKA Research Chair.

4. More about the collaborating organisations

The Inter-University Institute for Data-Intensive Astronomy (IDIA) was launched on 3 September 2015. This partnership between the University of Cape Town, the University of the Western Cape and the North-West University will develop crucial capacity for big data management and analysis, a spin-off of the SKA project. The R50-million, five-year IDIA partnership will integrate researchers in astronomy, computer science, statistics and eResearch technologies to create data science capacity for leadership in the MeerKAT SKA precursor projects, other precursor and pathfinder programmes and SKA key science.

ASTRON, the Netherlands Institute for Radio Astronomy is a division of the Netherlands Organisation for Scientific Research (NWO). ASTRON operates two well-known observatories in the Netherlands – the Westerbork Synthesis Radio Telescope and LOFAR, the Low Frequency Array.

SKA South Africa

The SKA project is an international effort to build the world's largest radio telescope with a square kilometre (one million square metres of collecting area). The first phase construction of SKA is being built in South Africa and in Australia. In South Africa, the SKA site is located in the Karoo near Carnarvon in the Northern Cape Province. SKA South Africa is building the MeerKAT, 64-antenna array radio telescope, which serves as a pathfinder instrument to SKA and will be integrated into SKA Phase 1.

The DOME project investigates approaches in exascale computing (which refers to computing systems capable of at least a billion billion calculations per second) at the ASTRON & IBM Centre for Exascale Technology in Drenthe, the Netherlands. The research is targeted at the specifications of SKA. SKA South Africa joined the DOME project in December 2012.

IBM-NL and ASTRON have been working together since 2012 in a 5-year collaboration totalling 32.9 million EURO to research exascale computer systems that will be needed by SKA.

MoU signatories: Mike Garrett, representing NWO and ASTRON; Alexander Brink, representing IBM; Jasper Horrell, representing SKA-SA/NRF; and Russ Taylor, representing IDIA/UCT.

ENDS

Issued by South Africa Square Kilometre Array (SKA SA) and the University of Cape Town, through the Inter-University Institute for Data Intensive Astronomy (IDIA)

ENQUIRIES:

Lorenzo Raynard
SKA South Africa
Cell: 071 454 0658

Professor Russ Taylor
Director: Inter-University Institute for Data Intensive Astronomy
SKA Research Chair: University of Cape Town & University of the Western Cape
Cell: 060 803 9133
Email: russ@ast.uct.ac.za

Professor Frikkie van Niekerk
Deputy Vice-Chancellor: Research, Innovation & Technology
North-West University
Cell: 083 676 7236
Email: frik.vanniekerk@nwu.ac.za