

Big Data and Digital Technology Workshop



Report of Contributions

Contribution ID: 1

Type: **not specified**

Registration

Contribution ID: 2

Type: **not specified**

Summary of previous workshops

Monday, 16 September 2019 11:30 (30 minutes)

Contribution ID: 3

Type: **not specified**

Ice-breaker activity

Monday, 16 September 2019 12:00 (45 minutes)

Contribution ID: 4

Type: **not specified**

Data science: a tool for social innovation

Monday, 16 September 2019 13:45 (30 minutes)

The world as we know it is changing at an exponential rate. The new world has placed a new demand for countries to keep up and adapt in innovative ways. Developing countries are increasingly being left behind and access to new skills and capabilities is ever more inadequate. The African continent is being presented with an opportunity for scientific participation with the global community via Astronomy and Big Data with the coming Square Kilometer Array (SKA) Telescope. African countries are now growing and joining the ranks of other emerging economies but there is a need to harmonize science into society for sustainable impact. The challenge with such a delicate stage of development is universal and equal access to opportunity and knowledge access. In this talk I will present how data science is being used in the SKA Africa partner countries to bring together experts in the field of Big Data, Machine Learning, Astronomy and Social sciences to seek how we can all work together for Social uplift.

Primary author: Dr OOZEER, Nadeem (South African Radio Astronomy Observatory (SARAO))

Presenter: Dr OOZEER, Nadeem (South African Radio Astronomy Observatory (SARAO))

Contribution ID: 5

Type: **not specified**

Developments in the Southern African Development Community (SADC) - Cyberinfrastructure Framework to Support Regional Collaborations on Big Data projects

Monday, 16 September 2019 14:15 (30 minutes)

This talk shares experiences around the development and implementation of the Southern African Development Community (SADC) Cyber-infrastructure Framework aimed at supporting multi-sectoral and multidisciplinary regional collaborations around big data projects of regional impact. The Cyber-infrastructure Framework envisaged a regional commons of high performance computing platforms over a fabric of National education and research networks to host multi-sectoral data repositories - facilitated by robust regional policies - utilised by trained users working on collaborative regional R&D projects. The Frameworks defines several pillars for implementation - ranging from harmonised regional policy through to infrastructure deployment, R&D, resource mobilisation, strategic partnerships, and human capital development. In the R&D pillar there are some regional projects e.g. in Weather and Climate and potential projects in geoscience, bioinformatics and spatial drone image data - all aimed at exercising the CI. The cyber-infrastructure also seeks to promote participation of citizens in science, promote education and innovation value chains.

Presenter: Dr MOTSHEGWA, Tshiamo (University of Botswana)

Contribution ID: 6

Type: **not specified**

Getting ready for the AVN in Madagascar: Big Data challenges & opportunities

Monday, 16 September 2019 14:45 (30 minutes)

Madagascar is one of the South African SKA partner countries on the AVN project. Located 50 km away from Antananarivo, the city of Arivonimamo hosts the old telecon dish to be converted into a radio telescope as part of the pan-African project. Despite the progress being made and various opportunities and support from the AVN and DARA community, there are still outstanding issues that need to be addressed such as the local ability to handle and analyse huge quantities of data collected well before successful conversion of the dish.

Presenter: Dr RANDRIAMANAKOTO, Zara (South African Astronomical Observatory)

Contribution ID: 7

Type: **not specified**

IDIA from big data to big ideas

Monday, 16 September 2019 15:45 (30 minutes)

Primary author: Dr HORRELL, Jasper (IDIA)

Presenter: Dr HORRELL, Jasper (IDIA)

Contribution ID: 8

Type: **not specified**

Discussion 1: Local communities and computing in developing countries

Monday, 16 September 2019 16:15 (45 minutes)

Panel members:

Nadeem Oozeer

Tshiamo Motshegwa

Zara Randriamanakoto

Primary author: Dr BRETON, Rene

Presenter: Dr BRETON, Rene

Contribution ID: 9

Type: **not specified**

Cardiac Remodelling Prediction using Deep Learning

Tuesday, 17 September 2019 09:30 (30 minutes)

Left Ventricular (LV) remodeling involves changes in the ventricular size, shape and function. Hence, analysis and prediction of LV remodeling is important to improve patient survival, emergency medical response and treatment strategies after acute myocardial infarction (AMI). As cardiac magnetic resonance imaging (MRI) is increasingly used in clinical assessment of cardiovascular diseases, the overwhelming size of a typical cardiac MRI image dataset poses a significant challenge for a time-efficient image quantification and interpretation for the cardiac expert. Furthermore, there is scarce study on classification and prediction of LV remodeling using multiple cardiac MRI derived data (oedema, infarct size and micro vascular obstruction at onset) and textual data. Therefore, it is critically important to propose an approach in mining meaningful patterns from multimodal data (image and text) from Sarawak Big Heart Data. We used a Deep Learning approach for the prediction of LV remodeling that can assist cardiac experts in LV remodeling prediction by analysing the images acquired from cardiac MRIs for faster decision making in clinical cardiac health management and prevention of cardiovascular mortality. This approach will then be realised in a proof of concept and will be compared with the existing manual images analysis by cardiac experts in term of sensitivity, specificity, positive and negative predictive value. In line with the Malaysia National Key Economic Area (NKEA), this research will address on a focus area in the Strategic Thrust 2 – Improving Well Being for All in the Eleventh Malaysia Plan and contribute to the coding of AMI, clinical management and treatment strategies for the improvement in system delivery for better health outcomes.

Primary author: Dr NURFATIMAH , Dayang (University of Malaysia, Sarawak (UNIMAS))

Presenter: Dr NURFATIMAH , Dayang (University of Malaysia, Sarawak (UNIMAS))

Contribution ID: 10

Type: **not specified**

Interpretable models for predicting early mortality in patients coronary artery disease

Tuesday, 17 September 2019 10:00 (30 minutes)

This work refers to the construction of models using machine learning algorithms for early prediction (during the first 24 hours of admission) of hospital mortality in patients with coronary artery disease through the use of clinical notes and structured clinical data (electronic health record - EHR). The aim is to effectively identify suitable models to predict early mortality and recognize risk factors. For unstructured EHR, n-gram models to extract feature in the clinical notes are explored. For structured data, different machine learning algorithms are evaluated and combined with different kind of information to identify risk factors. We also validate the model performance and compare its performance using reference scores. It is important to mention, that this study is focused on the analysis of the first 24 hours of admission, because during this time it is possible to identify invasive procedures or not, avoiding irreversible damage or sudden death.

Primary author: Dr GARCÍA NOCETTI , Demetrio Fabián (UNAM)

Presenter: Dr GARCÍA NOCETTI , Demetrio Fabián (UNAM)

Contribution ID: 11

Type: **not specified**

Astrostatistics: Big Questions in the age of Big Data

Tuesday, 17 September 2019 11:00 (30 minutes)

Astronomy has been one of the pioneering sciences in handling large volumes of data, which has historically meant that the bond between astronomy and statistics has been very strong. However, due to the advent of modern astrophysics in the 20th century, a significant estrangement between astronomy and statistics arose. Only recently the bond has begun to mend and to grow strong again thanks to astronomers acknowledging the limitations of “tried and true” statistical methods and becoming more willing to learn and apply modern statistical methods to their research. Incidentally, this has created an open science environment in which these methods have become available for anyone, lowering the barriers to entry for many astronomers in the Global South. In particular, I will mention how this has impacted my own research using a few examples from recent years.

Presenter: Dr CHAPARRO, German (ECCI)

Contribution ID: 12

Type: **not specified**

Computational Fluid Dynamics as a digital technology practice implementation

Tuesday, 17 September 2019 11:30 (30 minutes)

The talk focuses on Computational Fluid Dynamics as a digital technology practice implementation. These kind of techniques are not just useful in industry but in many different scientific research areas such as Astronomy and others.

Primary author: Mr SUAREZ, Mauricio (National Astronomical Observatory, Colombia)

Presenter: Mr SUAREZ, Mauricio (National Astronomical Observatory, Colombia)

Contribution ID: 13

Type: **not specified**

Discussion 2: Medicine and modern techniques

Tuesday, 17 September 2019 12:00 (45 minutes)

Panel Members:

Dayang Iskandar

Fabian Garcia Nocetti

Germán Chaparro

Mauricio Suarez

Primary author: Prof. THOMPSON, Mark (University of Hertfordshire)

Presenter: Prof. THOMPSON, Mark (University of Hertfordshire)

Contribution ID: 14

Type: **not specified**

Astronomy and Airlines: Expanding our horizons with Newton and GCRF

Tuesday, 17 September 2019 13:45 (30 minutes)

Since 2017, my Thai collaborators (NARIT, MFU) and I have worked together on various Newton/GCRF-funded projects. The first of these projects focussed on using data from GOTO - a new wide-field, high-cadence optical astronomical survey - to expose Thai students and researchers to vast amounts of digital data. Initial projects focused on the development of Machine Learning algorithms to analyse GOTO data, but our work has since expanded to include data management and front-end development. As well as astronomy-related research, however, our team has also been using our expertise in data analysis and management to help businesses and organisations across Thailand via two GCRF-funded projects. In these projects, students work closely with businesses across a diverse range of sectors - from supermarkets and hotels to airlines and cloud-computing providers - to help solve their data-related problems while also acquiring technical skills that are truly relevant for Thailand's growing technical economy. In this presentation, I will discuss the work we have done so far across all our projects, and outline our plans for the future of our collaboration.

Presenter: Dr MULLANEY, James (University of Sheffield)

Contribution ID: 15

Type: **not specified**

Recommender Systems - Mobile Phones

Tuesday, 17 September 2019 14:15 (30 minutes)

Recommender systems are a large part of Machine Learning Research. Relating to the Telecom industry, Antel (a Uruguayan Telecom Company) have developed a machine learning recommender system model to give mobile phones recommendations for telecom retail.

Primary author: LABORDE, Sebastian (ANTEL)

Presenter: LABORDE, Sebastian (ANTEL)

Contribution ID: 16

Type: **not specified**

BIOS, a case study of a public-private partnership for the promotion of biotechnology and data science in Colombia, Latin America

Tuesday, 17 September 2019 14:45 (30 minutes)

In 2007, the national government and public and private entities defined that it was necessary to take advantage of Colombian biodiversity through a high-level research center in the country that developed and harnessed its great potential in the fields of Bioinformatics, Biology Computational and Data Science. This idea pointed to the achievement of the objectives in innovation and economic development of the national government so that in 2008 its creation began, thanks to a cooperation agreement between a private public alliance. Additionally I want to talk about one of our main projects: ORIGEN, a national strategy of precision medicine for the genetic study of the Colombian population through the use of tools in artificial intelligence.

Presenter: RESTREPO, Eduardo Gomez (Centre for Bioinformatics and Computational Biology of Colombia (BIOS))

Contribution ID: 17

Type: **not specified**

Big Data Service and Network Infrastructure

Tuesday, 17 September 2019 15:45 (30 minutes)

Big Data and Digital Technology are new eras in the world, many activities in our lives have a strong relationship to each other in this new era, whether in the government, industry, business, or education sectors. In this talk I would like to talk about big data and digital services as a new big business opportunity in Indonesia and look at the network infrastructure as the main carrying capacity of these services.

Primary author: TAUFIK, Muhammad (Indosat Ooredoo)

Presenter: TAUFIK, Muhammad (Indosat Ooredoo)

Contribution ID: **18**

Type: **not specified**

Discussion 3: Machine learning and industry

Tuesday, 17 September 2019 16:15 (45 minutes)

Panel Members:

James Mullaney

Sebastián Laborde

Eduardo Gómez Restrepo

Muhammad Taufik

Primary author: SAWANGWIT, Utane (NARIT)

Presenter: SAWANGWIT, Utane (NARIT)

Contribution ID: 19

Type: **not specified**

Developing Big Data Analytics capacity in SKA Partner countries

Thursday, 19 September 2019 09:30 (30 minutes)

Developing Big Data Analytics capacity in SKA African Partner Countries involves the roll-out of infrastructure to help build the expertise in the different countries. At the core of the infrastructure are the skills required to perform data analysis, such as scientific programming and machine learning algorithms. Hence this talk will focus on these developments in the 8 partner countries in Africa.

Presenter: Dr SITHOLE, Happy (Council for Scientific & Industrial Research (CSIR) - NICIS)

Contribution ID: 20

Type: **not specified**

Big Data, HPC & Astronomical Numerical Simulations

Thursday, 19 September 2019 10:00 (30 minutes)

The 21st century can be the era of Big Data. Huge databases are generated by different scientific projects, industry and society in general. In the particular case of Astronomy, both ground (SKA, ALMA, SDSS, LSST, etc) and space telescopes (HST, Chandra, Spitzer, etc), as well as Astronomical Numerical Simulations (Millennium Simulation project, IllustrisTNG project, etc) are developed in supercomputers and generate databases of tens or hundreds of petabytes. To manipulate these databases, different data mining algorithms and specialized software for data analysis are used. In the case of numerical simulations data can be manipulated in two ways: first, the database is generated and then analyzed, or in-situ data analysis is done, i.e., data is analyzed as it is generated. In this presentation I will discuss the two possibilities for data analysis.

Primary author: Dr GONZALEZ, Alfredo Santillan (UNAM)

Presenter: Dr GONZALEZ, Alfredo Santillan (UNAM)

Contribution ID: 21

Type: **not specified**

Monitoring of an invasive weed from space

Thursday, 19 September 2019 11:00 (30 minutes)

In this talk I will present our work on the monitoring of an invasive weed called Parthenium, which spreads very quickly across Pakistan and can destroy a significant fraction of the crops if not actively controlled. Our research looks into implementing a new approach which relies on enabling large-scale monitoring using satellite imaging and taps on skills borrowed from astrophysics. We use extensive ground truth data in order to train machine learning algorithms to recognise the subtle signature of the plant as viewed from space at low resolution. Our work, still in progress, shows promising results demonstrating that we might indeed be able to track this weed from space.

Primary author: Dr BRETON, Rene (University of Manchester)

Presenter: Dr BRETON, Rene (University of Manchester)

Contribution ID: 22

Type: **not specified**

Do we need to invoke Machine Learning or Artificial Intelligence to add value to big data?

Thursday, 19 September 2019 11:30 (30 minutes)

Whenever we mention Big Data, we hear Machine Learning (ML) and/or Artificial Intelligence (AI). Is this required? Radio Frequency Interference (RFI) affects astronomical data that is measured by radio telescope such as MeerKAT. RFI is known to be dynamic in various aspects. There is often claims that the RFI environment radio facility around site is changing . How can we quantify this from the telescope perspective? Once given their science dataset, astronomers remove the RFI and hardly give feedback to the Operations, Commissioning and RFI team about the quality of their data, unless it is really bad. Furthermore, how confidently can one be to remove some potentially “affected” data points? In this talk I will show how we address the above 3 questions from a data science perspective.

Primary author: Dr OOZEER, Nadeem (South African Radio Astronomy Observatory (SARAO))

Presenter: Dr OOZEER, Nadeem (South African Radio Astronomy Observatory (SARAO))

Contribution ID: 23

Type: **not specified**

Discussion 4: HPC & cloud computing, Agri-food/local communities

Thursday, 19 September 2019 12:00 (45 minutes)

Panel member:

Happy Sithole

Alfredo J. Santillan

Rene Breton

Nadeem Oozeer

Primary author: Dr PREMADI, Premana (Bandung Institute of Technology)

Presenter: Dr PREMADI, Premana (Bandung Institute of Technology)

Contribution ID: 24

Type: **not specified**

China SKA Regional Centre Prototyping

Thursday, 19 September 2019 13:45 (30 minutes)

China has been taking part in SKA pre-construction since the 1990's; 21cm and FAST both have some relationship with SKA. In past years, we had involved many packages including DISH/CSP/SDP. We are highly involved with the SKA Regional Center construction from 2015 onwards. In this presentation, I will look at the progress of China SKA Regional Centre Prototyping, including the building, hardware, network and also some initial results using this platform, and the future plan of SRC.

Primary author: Mr GUO, Shaoguang (Shanghai Astronomical Observatory)

Presenter: Mr GUO, Shaoguang (Shanghai Astronomical Observatory)

Contribution ID: 25

Type: **not specified**

Ongoing Astronomical Projects in Algeria and Big Data Issues

Thursday, 19 September 2019 14:15 (30 minutes)

This talk will be about some of the most important astronomical projects in Algeria. I will speak mainly about the ongoing project of the Aures Observatory and the farm of telescopes for Transient Monitoring (GWs, GRBS etc), the Scheduler program with genetics algorithms, Site testing and Radioastronomy with an emphasis on Big Data requirements and needs (such as infrastructure and HPC).

Primary author: Dr SEGHOUBANI, Nassim (Centre of Research in Astronomy, Astrophysics & Geophysics)

Presenter: Dr SEGHOUBANI, Nassim (Centre of Research in Astronomy, Astrophysics & Geophysics)

Contribution ID: 26

Type: **not specified**

Bosscha Observatory Exoplanet Research

Thursday, 19 September 2019 14:45 (30 minutes)

The exoplanet research program was initiated in Bosscha Observatory since 2006 using the transit method. However, it was only since 2015 that the research program has been carried out more systematically and automatically. With the operation of the new telescope in 2014 and the implementation of the observation automation system, we have begun a new journey of exoplanets detection research. After spending the first few years learning, we can then concentrate on developing data servers and computing . The research now carries out transit photometry observation to produce a database of light curve set that will be identified of exoplanets. The MCMC fitting method is used to look for physical parameters afterwards. Other exoplanet detection research is conducted using surveys on stellar clusters. With the same observational data, we learn to apply machine learning to classified variable stars.

Primary author: Dr MUHAMMAD, Yusuf (Bandung Institute of Technology)

Presenter: Dr MUHAMMAD, Yusuf (Bandung Institute of Technology)

Contribution ID: 27

Type: **not specified**

Developing network of astronomical facilities in Indonesia and Asia-Pacific Region: Instruments, Data, and Knowledge Sharing

Thursday, 19 September 2019 15:45 (30 minutes)

Development of astronomy in South East Asian countries are gaining momentum mostly due to the success of human capacity building and increase in research funding. However, to contribute to the advancement of science at the forefront, in particular astrophysics, can be prohibitively costly for most developing countries. It is therefore imperative to construct collaborative strategies among astronomical institutions in countries in the same region that includes scientific goals, human resources, hardware and software facility, as well as cost and benefit among aspects to consider.

Primary author: Dr PREMADI, Premana (Bandung Institute of Technology)

Presenter: Dr PREMADI, Premana (Bandung Institute of Technology)

Contribution ID: 28

Type: **not specified**

Discussion 5: SKA and new facilities

Thursday, 19 September 2019 16:15 (45 minutes)

Panel members:

Shaoguang Guo

Nassim Seghouani

Muhammad Yusuf

Primary author: SANPA-ARSA, Siraprapa (NARIT)

Presenter: SANPA-ARSA, Siraprapa (NARIT)

Contribution ID: 29

Type: **not specified**

Big data and big images in galactic astronomy

Friday, 20 September 2019 09:30 (30 minutes)

TBC

Primary author: Prof. THOMPSON, Mark (University of Hertfordshire)

Presenter: Prof. THOMPSON, Mark (University of Hertfordshire)

Contribution ID: 30

Type: **not specified**

Computational Fluid Dynamics: A first approach to aerodynamics

Friday, 20 September 2019 10:00 (30 minutes)

In this talk the problem of the external fluid passing around an obstacle in 2-dimensions will be dealt. The stream-vorticity formulation will be used in order to understand the behaviour of the fluid after passing such obstacle. Although the geometry of the obstacle will be a fundamental one, this can be changed to more complex topologies.

Primary author: Mr SUAREZ, Mauricio (National Astronomical Observatory, Colombia)

Presenter: Mr SUAREZ, Mauricio (National Astronomical Observatory, Colombia)

Contribution ID: 31

Type: **not specified**

The implementation of big data architecture for radio astronomy and machine learning in a balloon to characterize the atmosphere in Ecuador

Friday, 20 September 2019 11:00 (30 minutes)

TBC

Primary author: MENA, Darwin (National Polytechnic School)

Presenter: MENA, Darwin (National Polytechnic School)

Contribution ID: 32

Type: **not specified**

Radioastronomy and Space Science: from a big data playground to societal spin-offs

Friday, 20 September 2019 11:30 (30 minutes)

Large sensor-based science infrastructures for radio astronomy will be among the most intensive data-driven projects in the world, facing very high demanding computation, storage, management, and above all power demands. The geographically wide distribution of the radioastronomical sensors and its associated processing requirements in the form of tailored High Performance Computing (HPC) facilities and cloud distributed environments require a Greener approach towards the Information and Communications Technologies (ICT) adopted for the data processing to enable operational compliance with strict power budgets. In addition, this convergence of Big data technologies with parallel computing opens new digital avenues, that may enhance Internet of Things (IoT) applications in smart farming, smart tourism and the ubiquitous smart cities frameworks. Here we outline major characteristics and innovative scenarios that may share common digital background and potential high societal impacts.

Primary author: Dr BARBOSA, Domingos (Instituto de Telecomunicações)

Presenter: Dr BARBOSA, Domingos (Instituto de Telecomunicações)

Contribution ID: 33

Type: **not specified**

Discussion 6: Technologies

Friday, 20 September 2019 12:00 (45 minutes)

Panel members:

Mark Thompson

Mauricio Suarez

Darwin Mena

Domingos Barbosa

Primary author: Dr BRETON, Rene

Presenter: Dr BRETON, Rene