



SKA SWG Update

Robert Braun, SKAO Science Director

17 January 2023



SKA Science Update

- Data Challenges Update (Philippa, Simon, Anna)
- Science Meetings (All)
- AOB



Science Data Challenge 2 results paper

- High level findings:
- **Complementary** methods
- Mix of **new and existing** techniques; **machine learning and non-machine** learning
- **SoFiA package** very popular thanks to excellent documentation and ease of use
- Analysis of **biases** and **HI mass** recovery with redshift

SKA Science Data Challenge 2: analysis and results

P. Hartley^{1*}, A. Bonaldi^{1,2}, R. Braun¹, J. N. H. S. Aditya³, S. Aicardi⁴, L. Alegre^{1,5}, A. Chakraborty⁶, X. Chen⁷, S. Choudhuri^{8,9}, A. O. Clarke¹, J. Coles¹⁰, J. S. Collinson¹, D. Cornu¹¹, L. Darriba¹², M. Delli Veneri¹³, J. Forbrich¹⁴, B. Fraga¹⁵, A. Galan¹⁶, J. Garrido¹², F. Gubanov¹⁷, H. Håkansson¹⁸, M. J. Hardcastle¹⁴, C. Heneka¹⁹, D. Herranz²⁰, K. M. Hess^{12,21,22}, M. Jagannath²³, S. Jaiswal³, R. J. Jurek²⁴, D. Korber¹⁶, S. Kitaef²⁵, D. Kleiner²⁶, B. Lao³, X. Lu¹¹, A. Mazumder⁶, J. Moldón¹², R. Mondal²⁷, S. Ni²⁸, M. Önnheim¹⁸, M. Parra¹², N. Patra^{6,29}, A. Peel¹⁶, P. Salomé¹¹, S. Sánchez-Expósito¹², M. Sargent^{16,30,31}, B. Semelin¹¹, P. Serra²⁶, A. K. Shaw³², A. X. Shen^{33,34}, A. Sjöberg¹⁸, L. Smith¹⁰, A. Soroka¹⁷, V. Stolyarov^{10,35}, E. Tolley¹⁶, M. C. Toribio³⁶, J. M. van der Hulst²², A. Vafaei Sadr³⁷, L. Verdes-Montenegro¹², T. Westmeier²⁵, K. Yu⁷, L. Yu³⁸, L. Zhang^{39,40}, X. Zhang²⁸, Y. Zhang³, A. Alberdi¹², M. Ashdown¹⁰, C.R. Bom¹⁵, M. Brüggen¹⁹, J. Cannon⁴¹, R. Chen³⁸, F. Combes^{11,42}, J. Conway³⁶, F. Courbin¹⁶, J. Ding³⁹, G. Fourestey¹⁶, J. Freundlich⁴³, L. Gao²⁸, C. Geller²⁶, Q. Guo⁷, E. Gustavsson¹⁸, M. Jirstrand¹⁸, M. G. Jones⁴⁴, G. Józsa⁴⁵, P. Kamphuis⁴⁶, J.-P. Kneib¹⁶, M. Lindqvist³⁶, B. Liu³⁸, Y. Liu⁷, Y. Mao⁴⁷, A. Marchal⁴⁸, I. Márquez¹², A. Meshcheryakov⁴⁹, M. Olberg³⁶, N. Oozeer⁴⁵, M. Pandey-Pommier⁵⁰, W. Pei⁷, B. Peng³⁸, J. Sabater⁵, A. Sorgho¹², J.L. Starck¹⁶, C. Tasse^{51,52}, A. Wang³, Y. Wang⁷, H. Xi³⁸, X. Yang³, H. Zhang³⁹, J. Zhang²⁸, M. Zhao²⁸, S. Zuo⁴⁷

Affiliations can be found after the references

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The Square Kilometre Array Observatory (SKAO) will explore the radio sky to new depths in order to conduct transformational science. SKAO data products made available to astronomers will be correspondingly large and complex, requiring the application of advanced analysis techniques in order to extract key science findings. To this end, SKAO is conducting a series of Science Data Challenges, each designed to familiarise the scientific community with SKAO data and to drive the development of new analysis techniques. We present the results from Science Data Challenge 2 (SDC2), which invited participants to find and characterise 233245 neutral hydrogen (HI) sources in a simulated data product representing a 2000 h SKA MID spectral line observation from redshifts 0.25 to 0.5. Through the generous support of eight international supercomputing facilities, participants were able to undertake the Challenge using dedicated computational resources. Alongside the main challenge, ‘reproducibility awards’ were made in recognition of those pipelines which demonstrated Open Science best practice. The Challenge saw over 100 participants develop a range of new and existing techniques, in results which highlight the strengths of multidisciplinary and collaborative effort. The winning strategy – which combined predictions from two independent machine learning techniques to yield a 20 percent improvement in overall performance – underscores one of the main Challenge outcomes: that of method complementarity. It is likely that the combination of methods in a so-called ensemble approach will be key to exploiting very large astronomical datasets.

Key words: methods: data analysis – radio lines: galaxies – techniques: imaging spectroscopy – galaxies: statistics – surveys – software: simulations

1 INTRODUCTION

The Square Kilometre Array (SKA) project was born from an ambition to create a telescope sensitive enough to trace the formation

* E-mail: philippa.hartley@skao.int



Reproducibility awards

SDC2



Reproducibility:

Is the software:

- Well-documented
- Easy to install
- Easy to use

Reusability:

Does the software:

- Use an open licence
- Have findable code
- Use code standards
- Use built-in tests

In partnership
with the Software
Sustainability
Institute



www.software.ac.uk

Reproducibility of the solution		Can the software pipeline be re-run easily to produce the same results? Is it:	
	Well-documented	<ul style="list-style-type: none"> • Well-documented Research software documentation best practice • Easy to install Top tips for packaging software • Easy to use Top tips for documentation 	
Well-documented	High-level description of what/who the software is for is available		Do you encourage other people to develop new projects? Does it:
	High-level description of what the software does is available		Using an open source licence
	High-level description of how the software works is available		Do you have findable code Choosing a repository for your project
	Documentation consists of clear, step-by-step instructions		Writing readable source code
	Documentation gives examples of what the user can see at each step e.g. screenshots or command-line excerpt		Writing readable source code
	Documentation uses monospace fonts for command-line inputs and outputs, source code fragments, function names, class names etc		Do you use a licence, BSD 3-Clause
Easy to install	Documentation is held under version control alongside the code		Do you use a repository
	Full instructions provided for building and installing any software		Do you have a reader
	All dependencies are listed, along with web addresses, suitable versions, licences and whether they are mandatory or optional		Do you have a repository available online
	All dependencies are available		Do you use a sustainable third-party repository Introduction to GitHub
Easy to use	Tests are provided to verify that the installation has succeeded		Do you use operators
	A containerised package is available, containing the code together with all of the related configuration files, libraries, and dependencies required. <i>Using e.g. Docker/Singularity</i>		Do you use well
	A getting started guide is provided outlining a basic example of using the software <i>e.g. a README file</i>		Do you use images or packages
	Instructions are provided for many basic use cases		Do you use image and variable names
		Reference guides are provided for all command-line, GUI and configuration options	Do you refer to the architecture or design
	Testing	Source code has unit tests	
		Software recommends tools to check conformance to coding standards e.g. A 'linter' such as PyLint for Python	



Reproducibility awards

SDC2



Reproducibility:

Is the software:

- Well-documented
- Easy to install
- Easy to use

Reusability:

Does the software:

- Use an open licence
- Have findable code
- Use code standards
- Use built-in tests

In partnership
with the Software
Sustainability
Institute



www.software.ac.uk

Results

Team name	Reproducibility award	Pipeline
EPFL	Bronze	https://github.com/epfl-radio-astro/LiSA
FORSKA-Sweden	Silver	https://github.com/FraunhoferChalmersCentre/ska-sdc-2
HI-FRIENDS	Gold	https://github.com/HI-FRIENDS-SDC2/hi-friends
NAOC-Tianlai	Bronze	https://github.com/kfyu/SDC2-tianlai
SHAO	Bronze	https://github.com/astrosumit/SDC2-SHAO
Team SoFiA	Silver	https://github.com/SoFiA-Admin/SKA-SDC2-SoFiA

Award announcement to be featured in next edition of Contact



Reproducibility awards

SDC3

- Revised award system
- Reproducibility 'badges'
- Based on Software Sustainability Institute's six steps to reproducibility
- Simpler for teams to follow and achieve

How to make your script ready for publication



www.software.ac.uk/research-software-camps



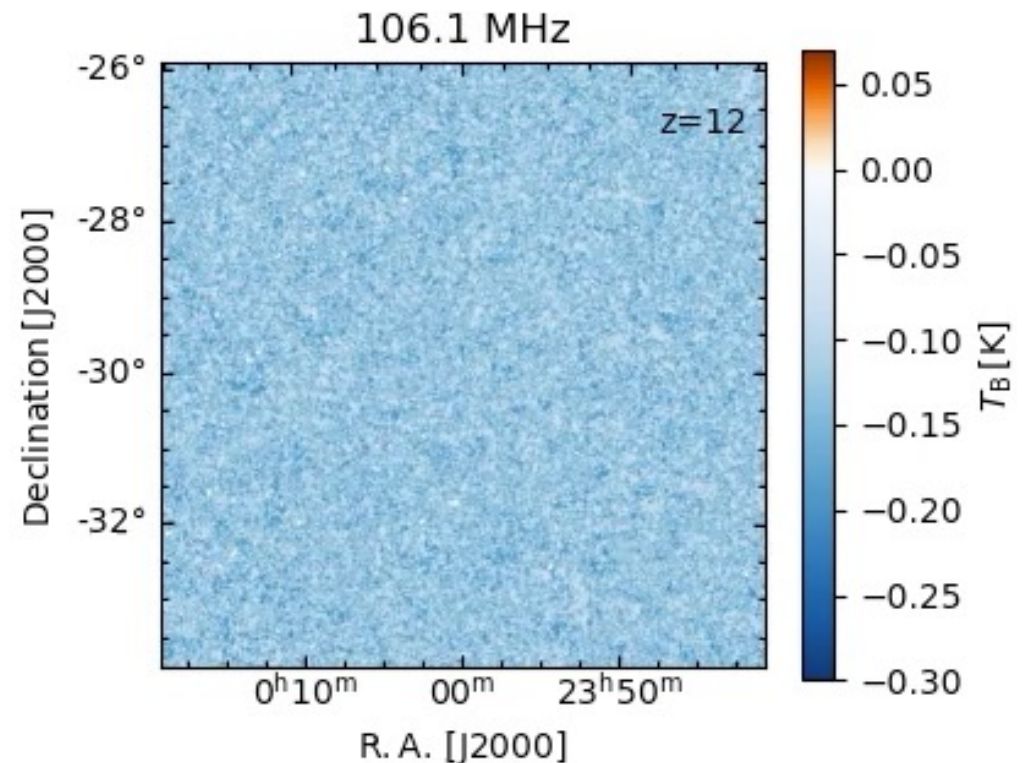
www.software.ac.uk



Science Data Challenge 3

Developed in collaboration with SKA EoR SWG members

- SDC3a "**Foregrounds**" (SDC3a; SWG Coordinators: C. Trott, V. Jelic)
 - **Foreground removal** exercise
 - SDC3a launching February 2023
- SDC3b "**Inference**" (SDC3b; SWG Coordinators: A. Mesinger, G. Melema)
 - Extraction of **cosmological parameters**
 - SDC3b launching Q3 2023



SDC3a Foregrounds

- Foreground Subtraction + 21cm Power Spectrum Extraction (SWG contacts: Trott & Jelic)
- Targets SWGs like CD/EoR, Cosmology, Continuum, etc.
- Dataset is calibrated visibilities and high fidelity image
- Challenge will involve:
 - a) Removal of point source + diffuse foregrounds
 - b) Extraction of the cylindrical power spectrum
 - c) Comparison with the original input signal PS



SDC3a Dataset

Visibilities

- 7.5 TB
- MS and UVFITS
- 106 – 196 MHz
- 100 kHz channels
- 10s integration times
- 4h track, 1000h integration

Image Cube

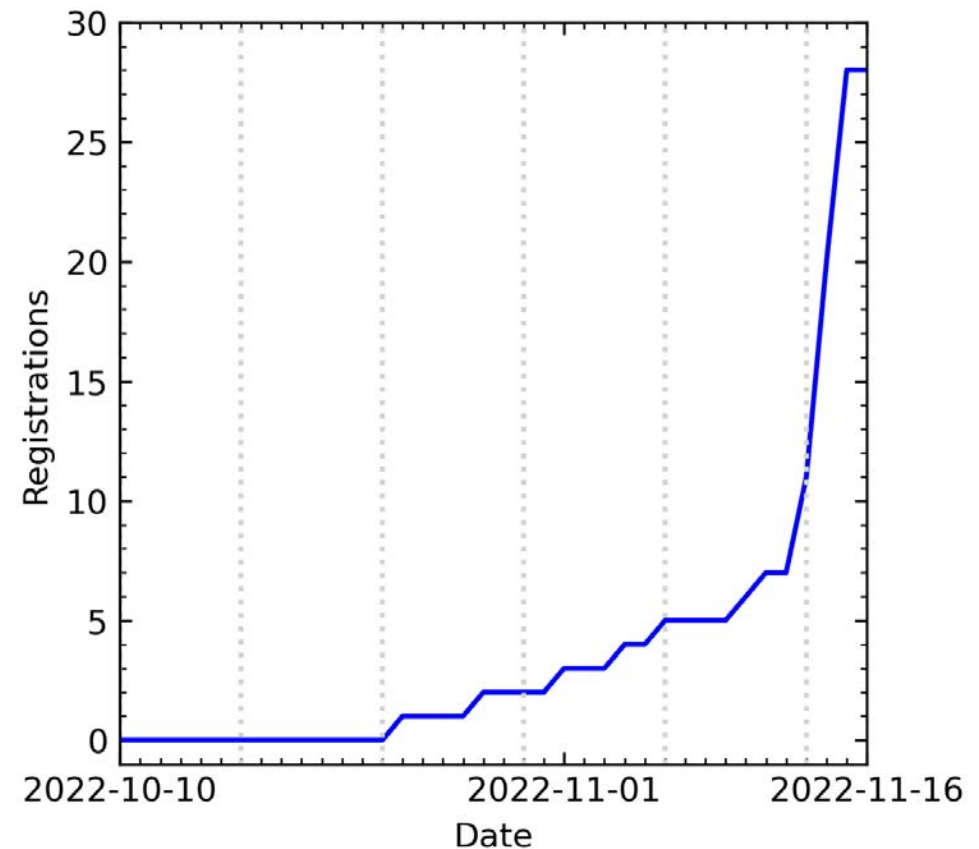
- 15 GB
- 2048 x 2048 pixels
- Pixel size of 16 arcsec
- 901 frequency channels
- Uniform and natural cubes
- PSFs for both weightings

→ *Dataset generated on SKAO server in ~4 weeks*



SDC3a Registrations

- Registration ran from 10th October 2022 until 15th November 2022
- A total of 28 registrations from 13 countries were recorded

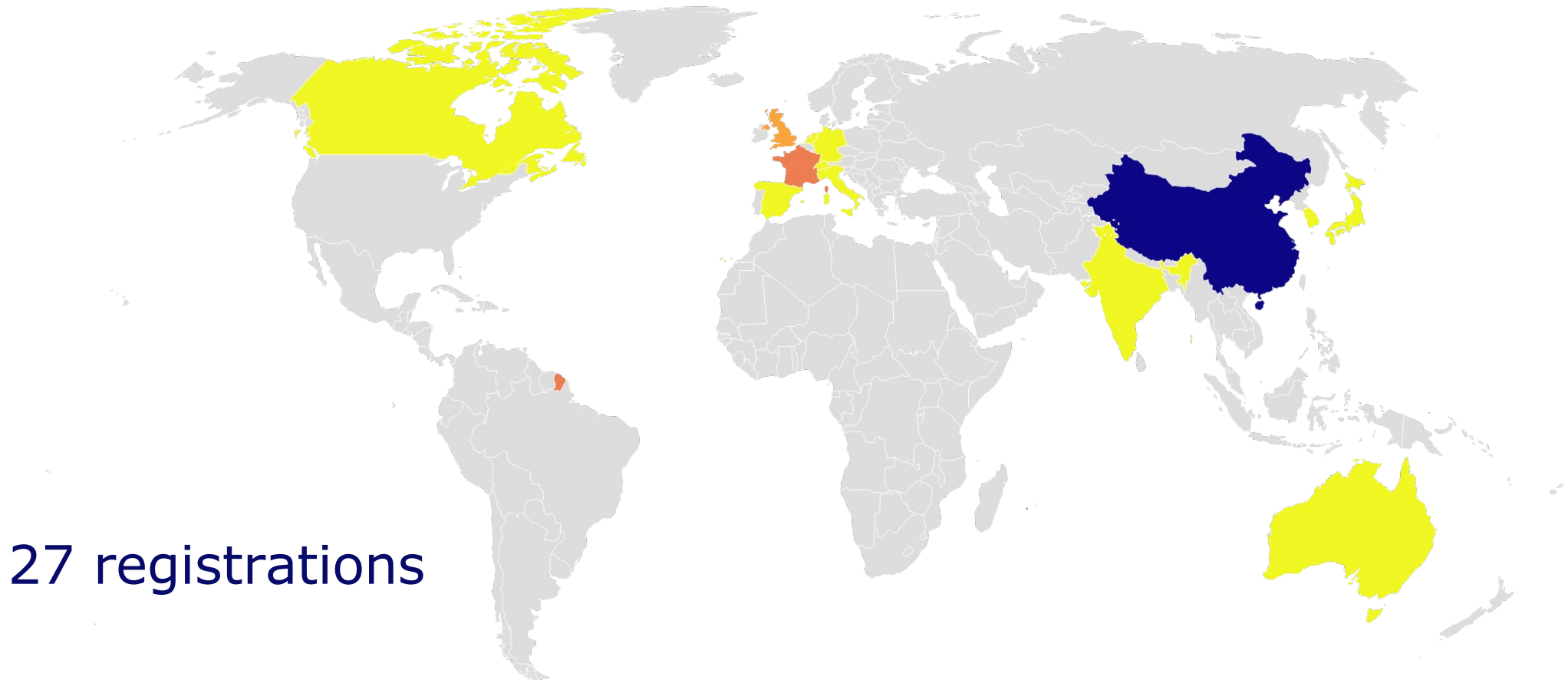


Number of registrations over time, starting at the start of the registration period.



SDC3a Registrations

Team Leader Affiliations

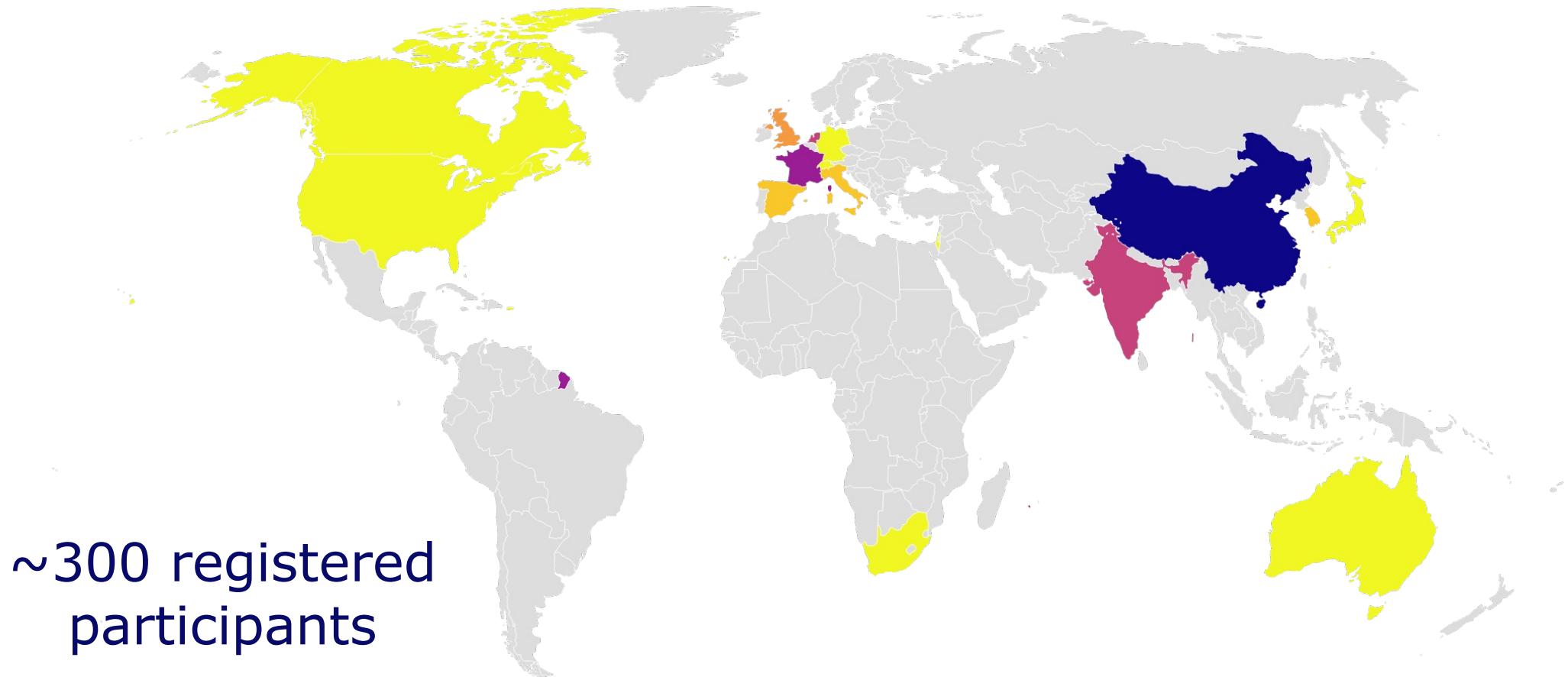
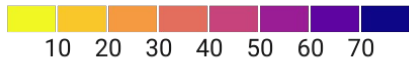


27 registrations



SDC3a Registrations

Team Member Affiliations

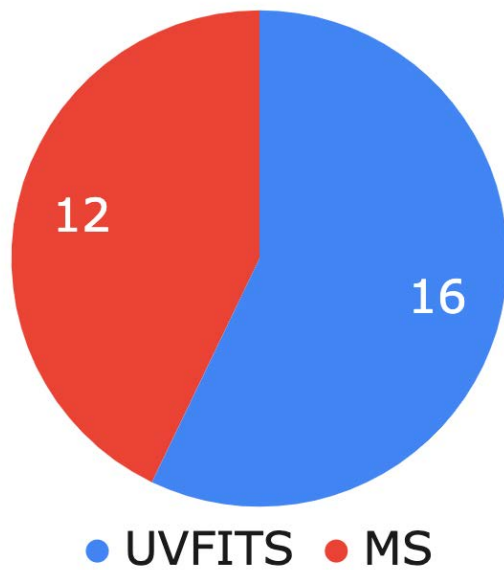


~300 registered participants

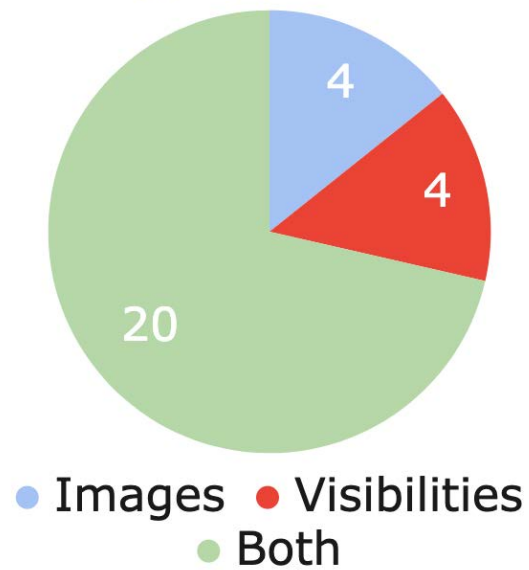


SDC3a Registrations

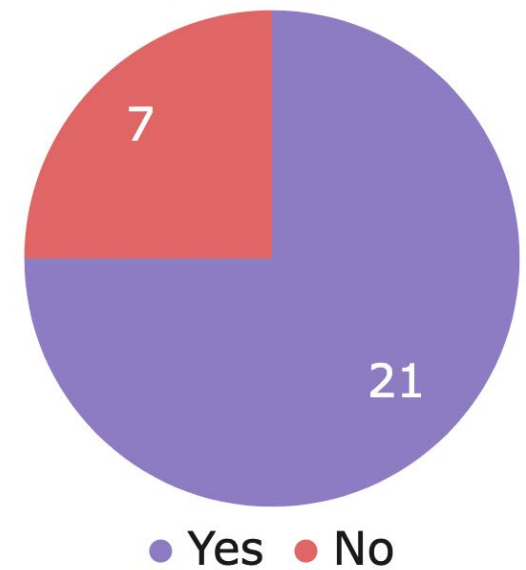
Data Format Preference



Data Type(s) Wanted



HPC Requested?



HPC Facility Partners

- ASTRON

- AusSRC

- CESGA

- ChinaSRC

- EngageSKA

- GENCI-IDRIS

- INAF

- IRIS-CAM

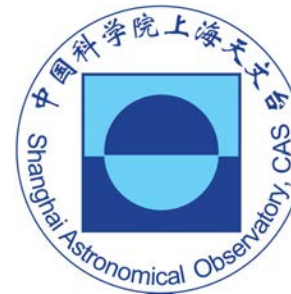
- IRIS-MAN

- JPSRC

- SPSRC

- Swiss SRC

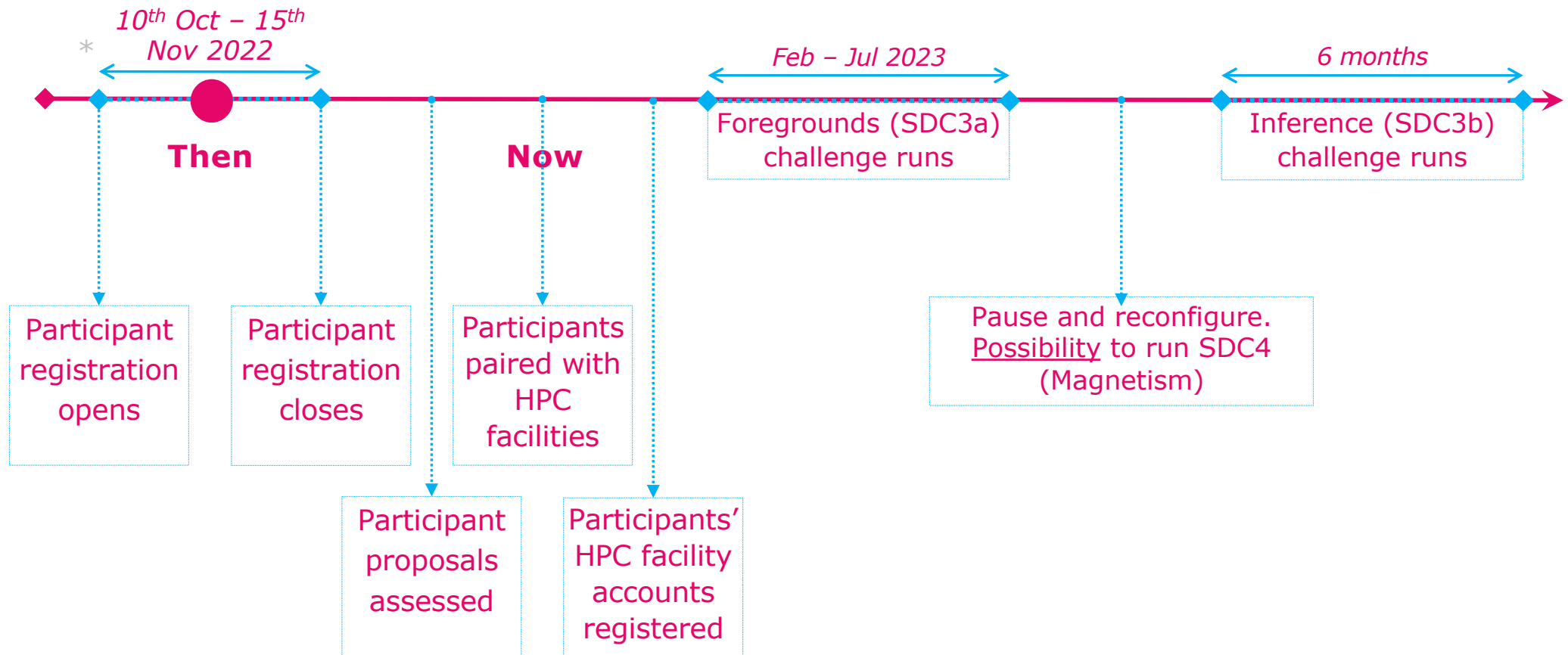
- UC-LCA



HPC Facility Partners



SDC3 Timeline – Important Events



**Not to scale*



Science Meetings

- MeerKAT Extension Science Workshop, 1 – 3 Feb. 2023, Univ. Witwatersrand, ZA
<https://www.wits-astro.org/mkplus2023>
- Joint ESO/SKAO meeting “Coordinated Surveys of the Southern Sky”, in Garching: week of 27 February 2023 (see next slides)
<https://www.eso.org/sci/meetings/2023/CSSS.html>
- PHISCC 2023 – “HI surveys in full swing”, Cape Town, 27 – 31 March 2023
<http://www.astro.rug.nl/phiscc2023>
- Joint SKAO/ngVLA Science Conference week of 30 April 2023, in Vancouver (see next slides) <http://go.nrao.edu/ngVLASKA>
- Community of European Solar Radio Astronomers (CESRA) Workshop, 3 – 7 July 2023, <https://star.herts.ac.uk/cesra/>
- EAS 2023, Krakow, 10 – 14 July 2023, SKAO Lunch Session (1.5 hour) approved, now being planned





Coordinated Surveys of the Southern Sky

February 27 - March 3, 2023

ESO Garching



- Remote participation possible
- Conference Mon afternoon - Thur morning
- Workshop (by invitation) Thur afternoon - Fri morning
- Discuss synergies and coordinated observations between ESO facilities and SKAO + precursors/pathfinders
- Forge collaborations between communities
- Registration open until end of Jan
- Preliminary program available
- 250+ registrations so far!
- For more info visit <https://www.eso.org/sci/meetings/2023/CSSS.html>



New Eyes on the Universe: SKA & ngVLA Vancouver 1-5 May 2023

Important Dates:

Abstracts

Dec 2, 2022 – Abstract submission open

Feb 10, 2023 – Abstract submission deadline (Oral)

Registration

Jan 16, 2023 – Opens (yesterday)

Mar 7, 2023 – early bird registration closes

Apr 7, 2023 – Closes

**** If planning to attend in person, please check if you need a Canadian visa, and the processing time in your country ****

Programme

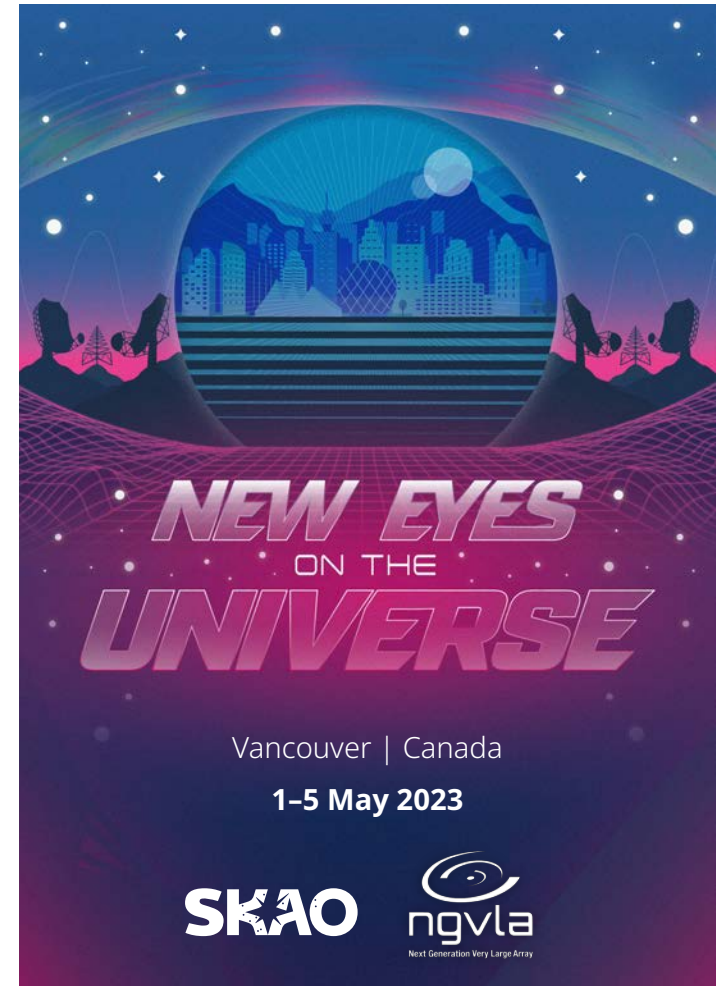
Mar 3, 2023 – Announced

(Draft block programme available on conference web page – next slide)

Hotel

Apr 7, 2023 – cutoff for conference rate. Please stay at the hotel if you can.

<http://go.nrao.edu/ngVLASKA>



New Eyes on the Universe: SKA & ngVLA Vancouver 1-5 May 2023

Program:

Special Session (Monday) – Broader Impacts

Monday – Wednesday: focus on science

Thursday – Friday: focus on operations
e.g., Joint Observations, Interoperability, Data Products, Data Processing, Archival Science, and Analysis Tools (but not limited to). And some science talks.

Invited Speakers:

Leindert Boogaard, Katherine de Kleer,
Nanase Harada, Kenda Knowles, James Miller-Jones, Leah Morabito, Sarah Burke Spolaor, Nienke van der Marel, Tony Beasley, Phil Diamond, Michael Rupen

	Monday May 1, 2023	Tuesday May 2, 2023	Wednesday May 3, 2023	Thursday May 4, 2023	Friday May 5, 2023
8:45 AM	Welcome				
9:00 AM	Talks	Talks	Talks	Talks	Talks
10:30 AM	Break	Break	Break	Break	Break
11:00 AM	Talks	Talks	Talks	Talks	Talks
12:00 PM	Poster Flash	Poster Flash	Poster Flash	Poster Flash	Poster Flash
	Lunch / SWG tagups	Lunch / SWG tagups		Lunch / SWG tagups	Lunch / SWG tagups
2:00 PM	Talks	Talks	Free Afternoon Optional Group Excursions	Talks	Talks
3:30 PM	Break	Break		Break	Conference Summary
4:00 PM	Special Session Talks	Talks		Talks	
5:30 PM	Opening Reception			Bar Happy Hour/Poster Session	
6:00 PM		Conference Dinner			

<http://go.nrao.edu/ngVLASKA>



Any Other Business

- News from SWG Chairs?
- ...



*We recognise and acknowledge the
Indigenous peoples and cultures that have
traditionally lived on the lands on which
our facilities are located.*

SKAO

www.skao.int