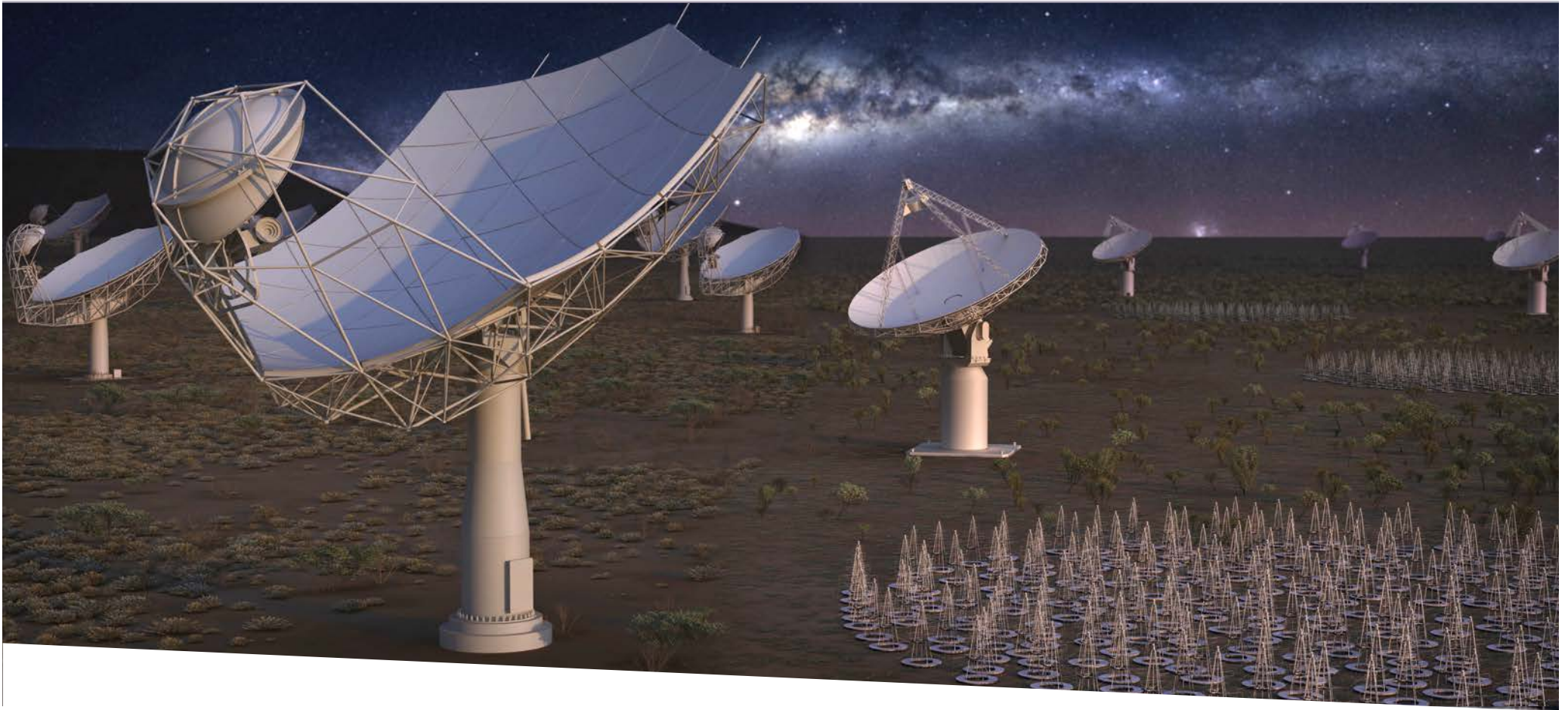


SKA SWG Update



SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Robert Braun, Science Director

15 Sept 2020

Science Activity Updates

- Next Science Meeting (Anna)
- Next Science Data Challenge (SDC2) (Anna, Philippa)
- AOB

Next SKA Science Meeting

- SKA Science Meeting, March 15 – 19, 2021
 - Fully virtual conference
 - Learnings from many precedents over past months
 - Also “CSIRO Symposium: The Future of Meetings” 15 – 17 Sept
 - Title: “The Precursor View of the SKA Sky”
 - Both Live and Pre-recorded (but with Live Q&A) sessions being planned to maximise time-zone coverage and accessibility
 - Parallel sessions based on multiple-SWG science topics, total number to be confirmed

Next SKA Science Data Challenge (SDC2)



- SDC2 goals:
 - Get the community familiar with an SKA HI imaging data product (TB scale)
 - Develop efficient methods for source finding and source characterization -> SWG and SRC applications
 - Encourage the community towards best practices in coding (containerised challenge submissions)
 - Expose/train the community in use of proto-SRCs
 - Test proto-SRC resources on a real use case, with real users

Next SKA Science Data Challenge (SDC2)



Task	Status
Dataset	Prototype HI and continuum data shared for feedback through HI SWG chairs
Computational resources	Progressing agreements with HPC facilities to provide computational resources for the challenge
Supporting website	Dedicated website now live containing links to the test data, instructions, details of the HPC facilities, leaderboard, links to discussion forum, etc.
Scoring code	SDC2 scorer to be available from the start of the challenge, for teams to self-assess their performance

SDC2: Simulated HI Survey

Survey Properties

- $z = 0.24 - 0.5$
- 20 deg^2 FoV
- $7 \text{ arcsec} / 30 \text{ kHz}$ resolution
- 2000^{h} simulated integration
- 1 TB Data Product size

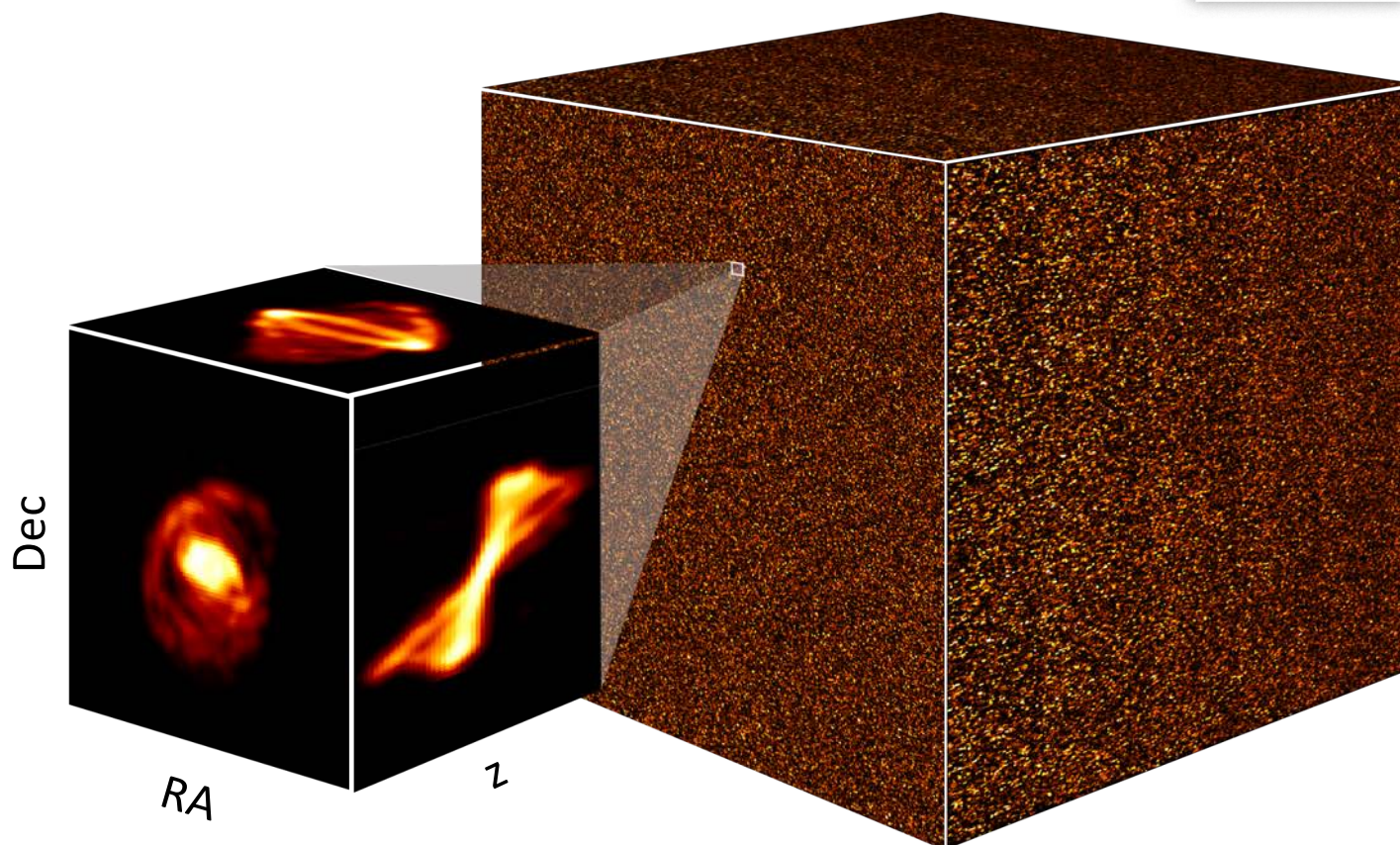


Illustration of simulated HI sky prior to sampling with the telescope beam and inclusion of system noise

- Almost 10^6 simulated neutral hydrogen galaxies and 10^7 continuum sources (not shown here)
- Expect up to 10^5 HI detections with more than 10^3 well-resolved

SDC2 Web Page:

Now live on sdc2.astronomers.skatelescope.org



Welcome to the second SKA Science Data Challenge. Our latest challenge will see participants analyse a simulated datacube 1 TB in size, in order to find and characterise the neutral hydrogen content of galaxies across a sky area of 20 square degrees.

Neutral hydrogen – or HI – exists in large quantities beyond the visible edges of most star-forming galaxies. Emitting light at a fixed radio wavelength during occasional electron ‘spin-flips’, HI traces the rotation of galaxies, allowing astronomers to infer the amount of mass – both visible and dark – contained within. The unprecedented sensitivity of the SKA will be used to map HI out to the formation of the first galaxies, just 380,000 years after the Big Bang. This period, known as “Cosmic Dawn”, began some 13.5 billion years ago. The challenge dataset will be a simulation of an SKA HI observation up to a distance of 4 billion light years.

In order to provide such a large dataset for analysis, we have teamed up with high performance computing facilities around the world. Participants will be invited to compete in teams and create accounts at one of those facilities, on which the data will be accessed and processed directly.

With the challenge launching soon, we now invite teams and individuals to find out more and register their interest in participating, using the link below.

REGISTER YOUR INTEREST





SDC2 Scoring Code:

SDC1 scorer distributed via PyPI. SDC2 scorer under development

ska-sdc 1.0.0

`pip install ska-sdc`



Latest version

Released: Aug 13, 2020

A package providing tools for the SKA Science Data Challenges.

Navigation

Project description

Release history

Download files

Project links

Homepage

Statistics

View statistics for this project via [Libraries.io](#), or by using [our public dataset on Google BigQuery](#)

Project description

Science Data Challenge Scoring Code API

The SKA Science Data Challenge #1 (<https://astronomers.skatelescope.org/ska-science-data-challenge-1/>) tasked participants with identifying and classifying sources in synthetic radio images.

In addition to the synthetic images, participants were provided with a section of the 'truth catalogue' of sources used to generate the artificial data. Comparing the truth catalogue with the 'submission catalogue' produced by a participant's solution would provide a means of determining the success of the solution.

To evaluate the accuracy of the results, a program was developed to cross-match sources between the submission and truth catalogues, and calculate a 'score' based on the result of this cross-match.

This is an open-source implementation of the program used to score and rank the submissions for the first SKA Science Data Challenge (SDC). A number of improvements have been made, most notably the use of a more performant cross-match algorithm. As such it is not possible to make a direct comparison between the scores produced by this package and the original program. The original IDL code is available at: <https://astronomers.skatelescope.org/ska-science-data-challenge-1/>

SDC2 access and processing



Description

Provide access to different cluster facilities to assist the users on solving the challenge

Main goals

- Test and put in place strategies that take a step forward to deal with the real SKA data
- Exploring prototyping centralized data capabilities

Benefits

Users

- Provides Computing resources for the users
- No need to download the data
- Prepare the scientific community for future SKA practices

SKA

Tests SRC Prototyping

- Data access and transfer
- Containerization
- Access and security
- Protocols

Possible outcome with the challenge: pipelines



Clusters

Specialized in different tasks



Supporting

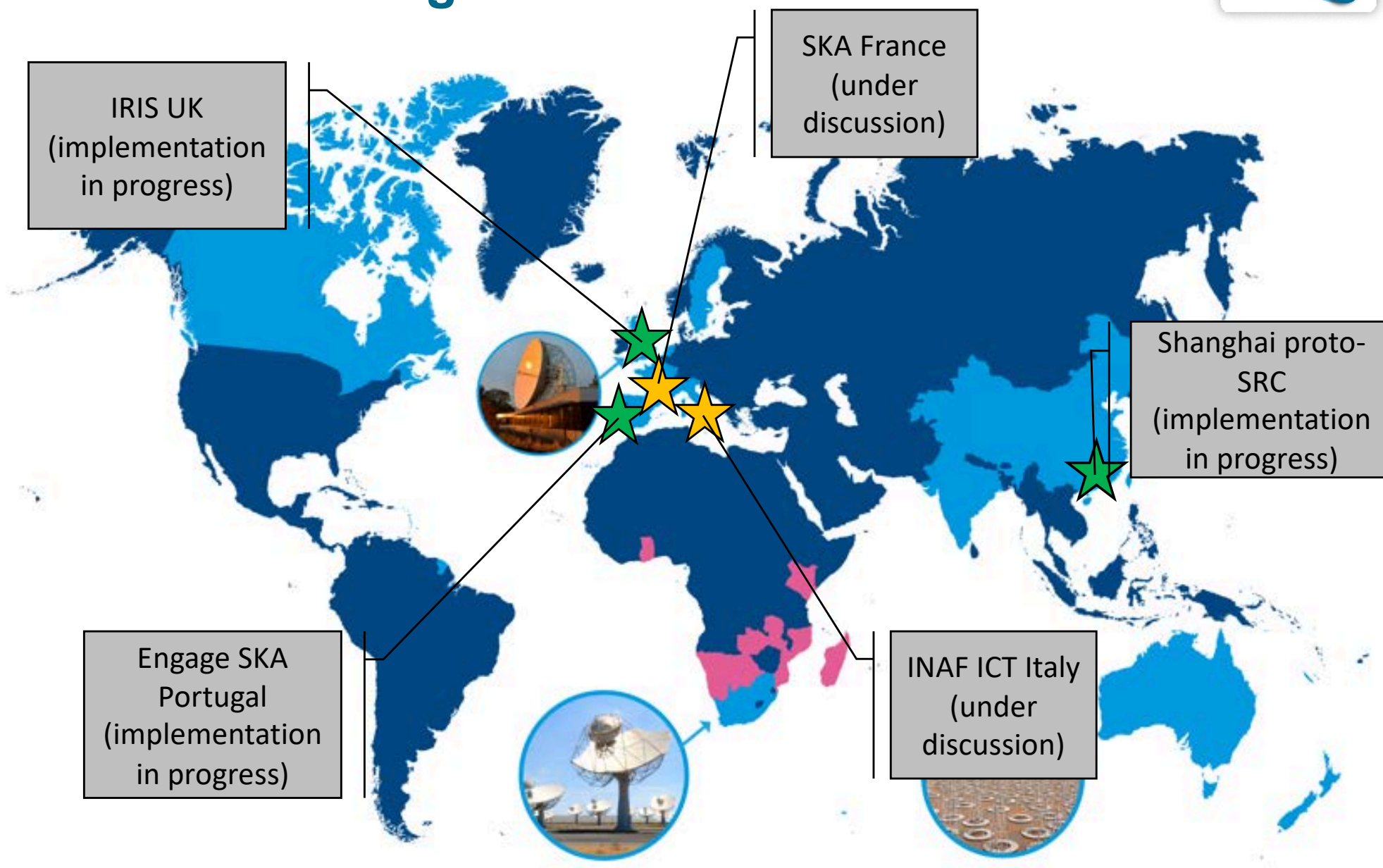
Access and usage instructions



Teams

allocated to different clusters

SDC2 Processing:



SDC2 Processing Centres

- How many do we need?
 - Registration of interest opened on SDC2 website
 - Computational cost per team to solve the challenge estimated
 - Each facility to indicate a maximum number of teams they can support
 - Once computational needs are met, balance between including participation of proto-SRCs and timely delivery of SDC2
 - Facilities engagement can continue for future data challenges

With the challenge launching soon, we now invite teams and individuals to find out more and register their interest in participating, using the link below.

REGISTER YOUR INTEREST

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

