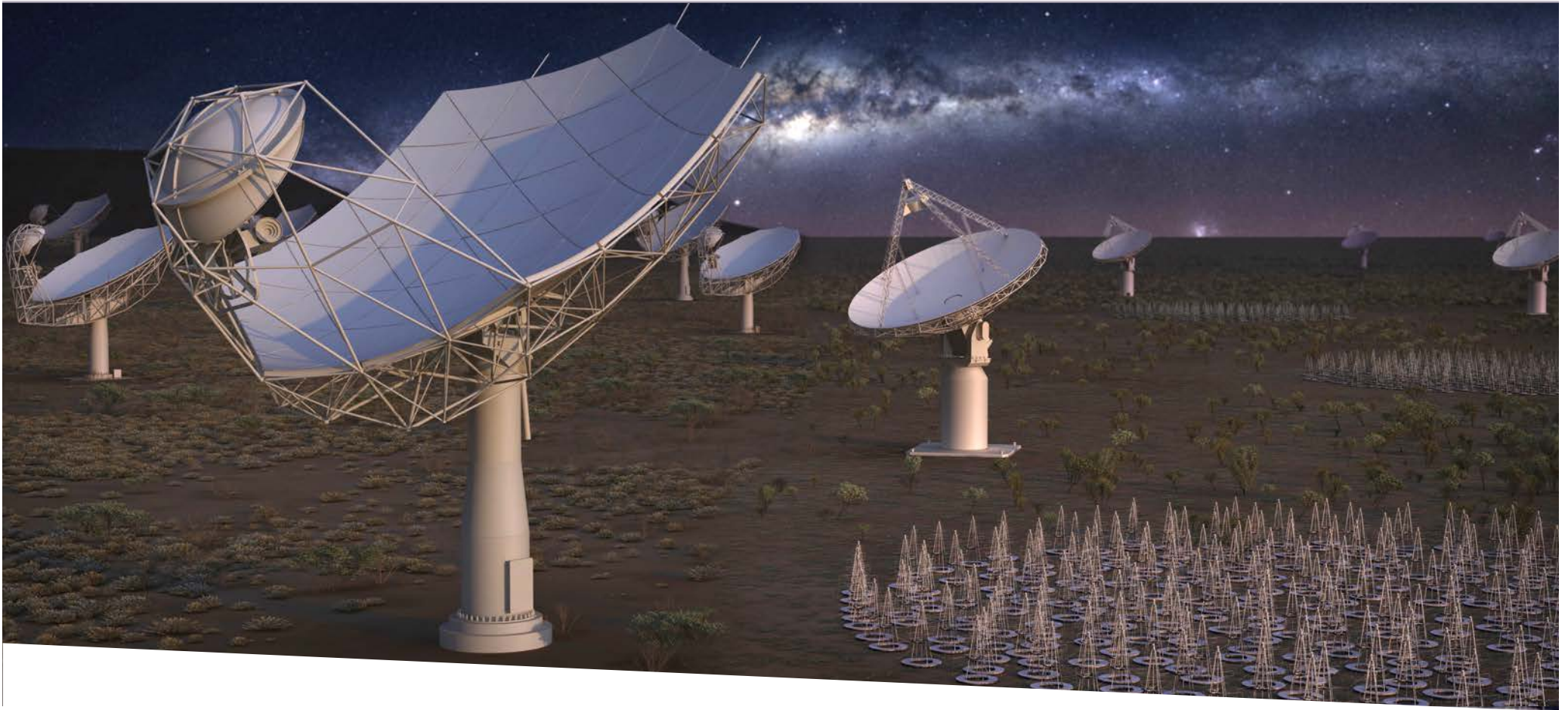


# SWG Science Update



**SQUARE KILOMETRE ARRAY**

Exploring the Universe with the world's largest radio telescope

**Robert Braun, Science Director**

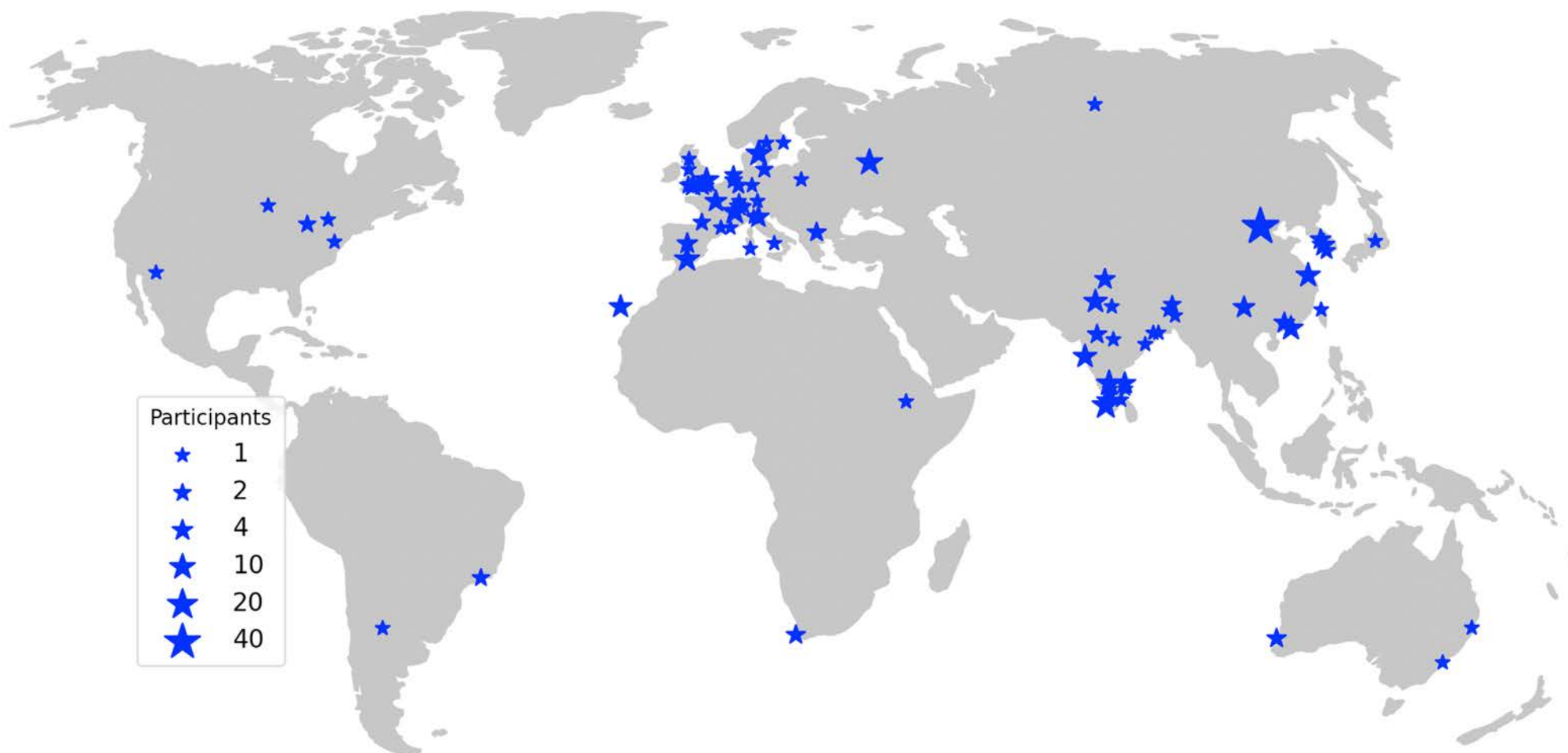
18 May 2021

# Science Activity Updates

- Science Data Challenges (Philippa, Anna)
- AOB

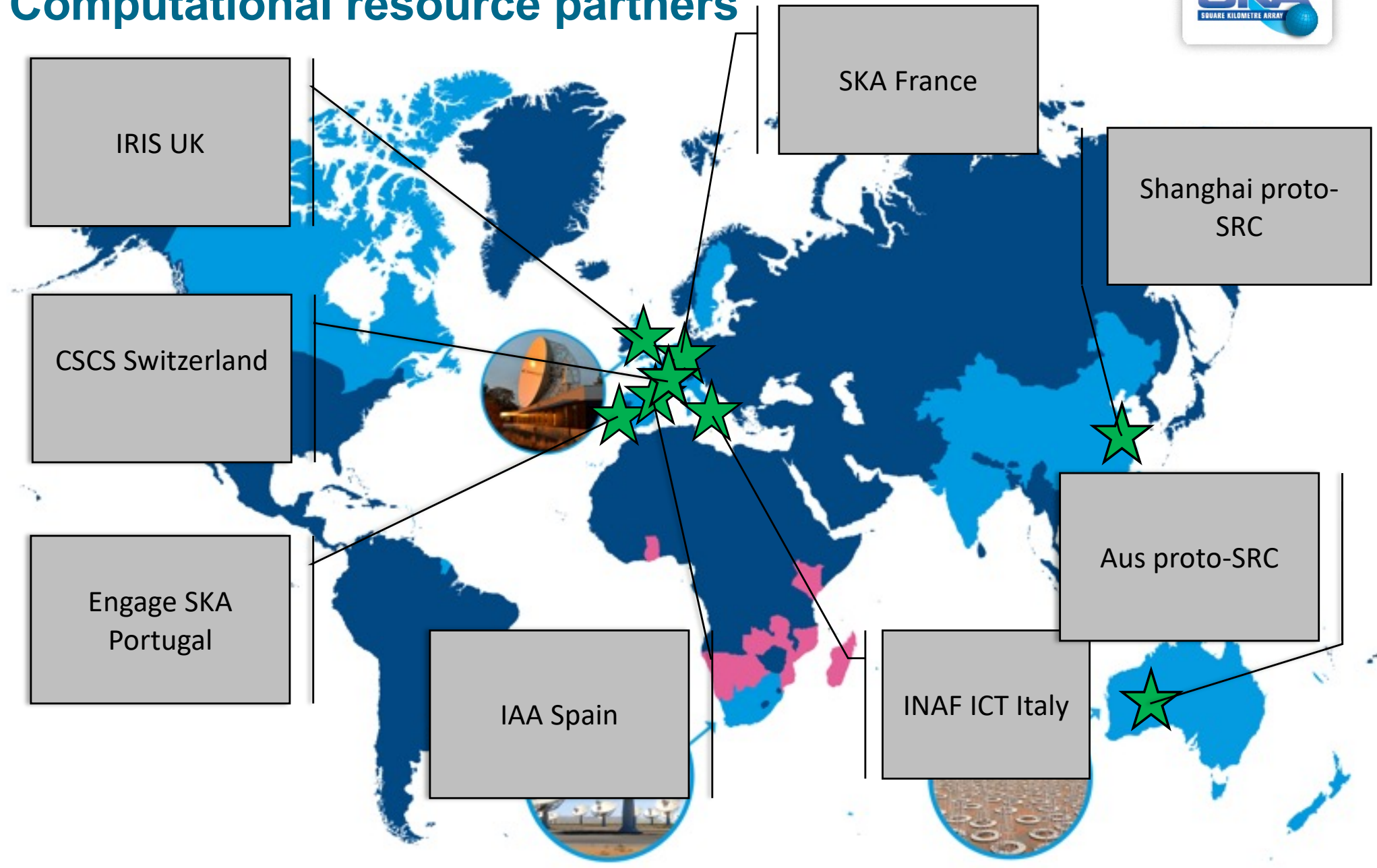
# Science Data Challenge 2

Duration Feb 1<sup>st</sup> – July 31<sup>st</sup>



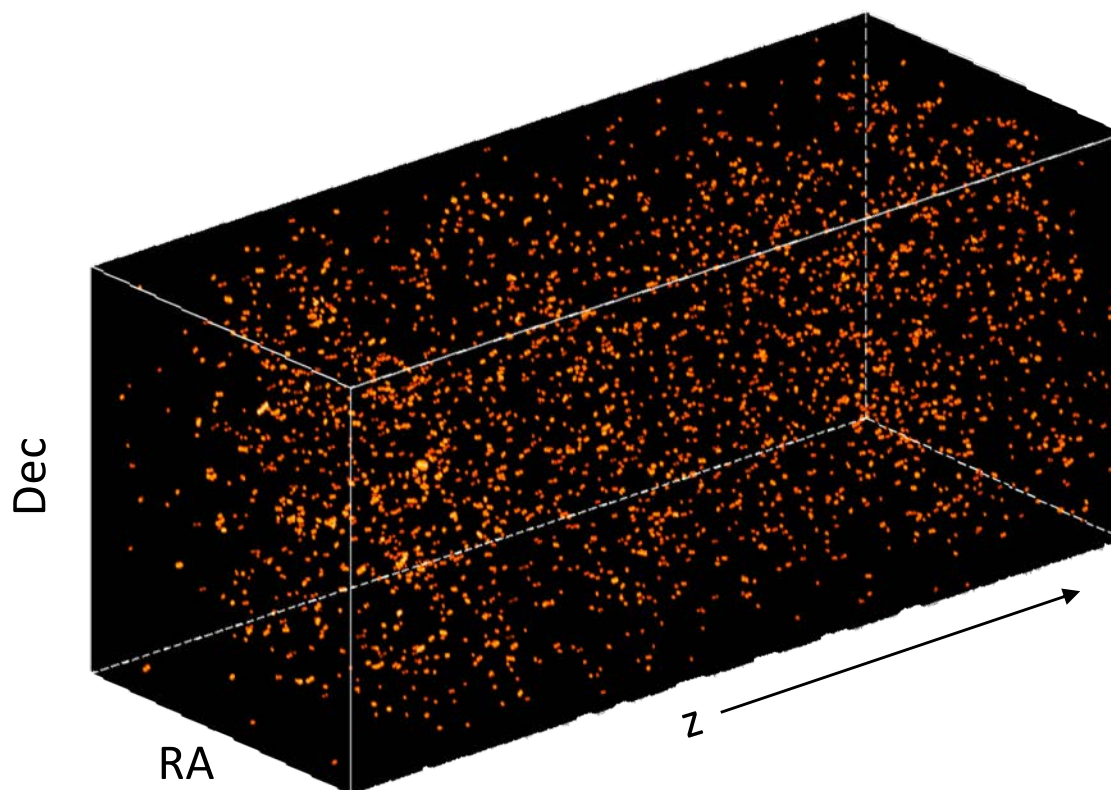


# Computational resource partners



# • (Fairly) big data

Proto-type model cube shown: 1% of full data product



- Integration time = 2000h
- Spatial/Frequency resolution = 7 arcsec / 30 kHz
- RMS per channel 13 – 18  $\mu$ Jy
- FoV = 20 square degrees
- Frequency = 950 MHz – 1150 MHz ( $z = 0.25$  to  $0.5$ )
- Data volume = 1 TB
- 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 survey



## SDC2 participant feedback

We would be very grateful to hear about how you are finding the challenge so far. This feedback will help us to identify how we can provide the best support during the challenge, and will also help inform the design of future challenges

**\* Required**

Please could you provide your team name \*

Your answer

### Data processing

We would like to understand your experience of processing the challenge datasets

On which SDC2 dataset(s) are you currently working? \*

☐ None so far

☐ Evaluation dataset

# Participant survey

- Survey sent to all participating teams
- Duration **6<sup>th</sup>-20<sup>th</sup> April**
- 13 responses (12 teams)
- Main findings:
  - Most planning to analyse full challenge dataset
  - Appetite for additional interaction
  - Some connectivity issues to HPCs
  - Some pipeline environment issues
  - Reproducibility [checklists](#) in use
- Actions
  - Updated facility [allocation record](#)
  - Individual follow-up where needed
  - Q&A telecons inviting all teams scheduled
    - Plus demo of new scoring service version and [leaderboard](#)

Reproducibility of the solution	
Can the software pipeline be re-run easily to produce the same results? Is it:	
<ul style="list-style-type: none"> <li>Well-documented <a href="#">Approach software documentation</a></li> <li>Easy to install <a href="#">See this for software</a></li> <li>Easy to use <a href="#">See this for documentation</a></li> </ul>	
Well-documented	<ul style="list-style-type: none"> <li>High level description of what/why the software is for is available</li> <li>High level description of what the software does is available</li> <li>High level description of how the software works is available</li> <li>Documentation contains of clear step-by-step instructions</li> <li>Documentation gives examples of what the user can do at each step e.g. screenshots or command line output</li> <li>Documentation uses <code>code</code> tags for command line inputs and outputs, source code fragments, function names, class names etc.</li> <li>Documentation is held under version control alongside the code</li> </ul>
Easy to install	<ul style="list-style-type: none"> <li>Full instructions provided for building and installing any software</li> <li>All dependencies are listed, along with web addresses, suitable versions, licenses and whether they are mandatory or optional</li> <li>All dependencies are available</li> <li>Tests are provided to verify that the installation has succeeded</li> <li>A compressed package is available, containing the code together with all of the external configuration files, libraries, and dependencies required (e.g. Docker/Conda)</li> </ul>
Easy to use	<ul style="list-style-type: none"> <li>A getting started guide is provided outlining a basic example of using the software e.g. a README file</li> <li>Instructions are provided for many basic use cases</li> <li>Reference guides are provided for all command line, GUI and configuration options</li> </ul>
Testing	<ul style="list-style-type: none"> <li>Source code has unit tests</li> <li>Software recommends tools to check conformance to coding standards e.g. A Tester such as Pylint for Python</li> </ul>



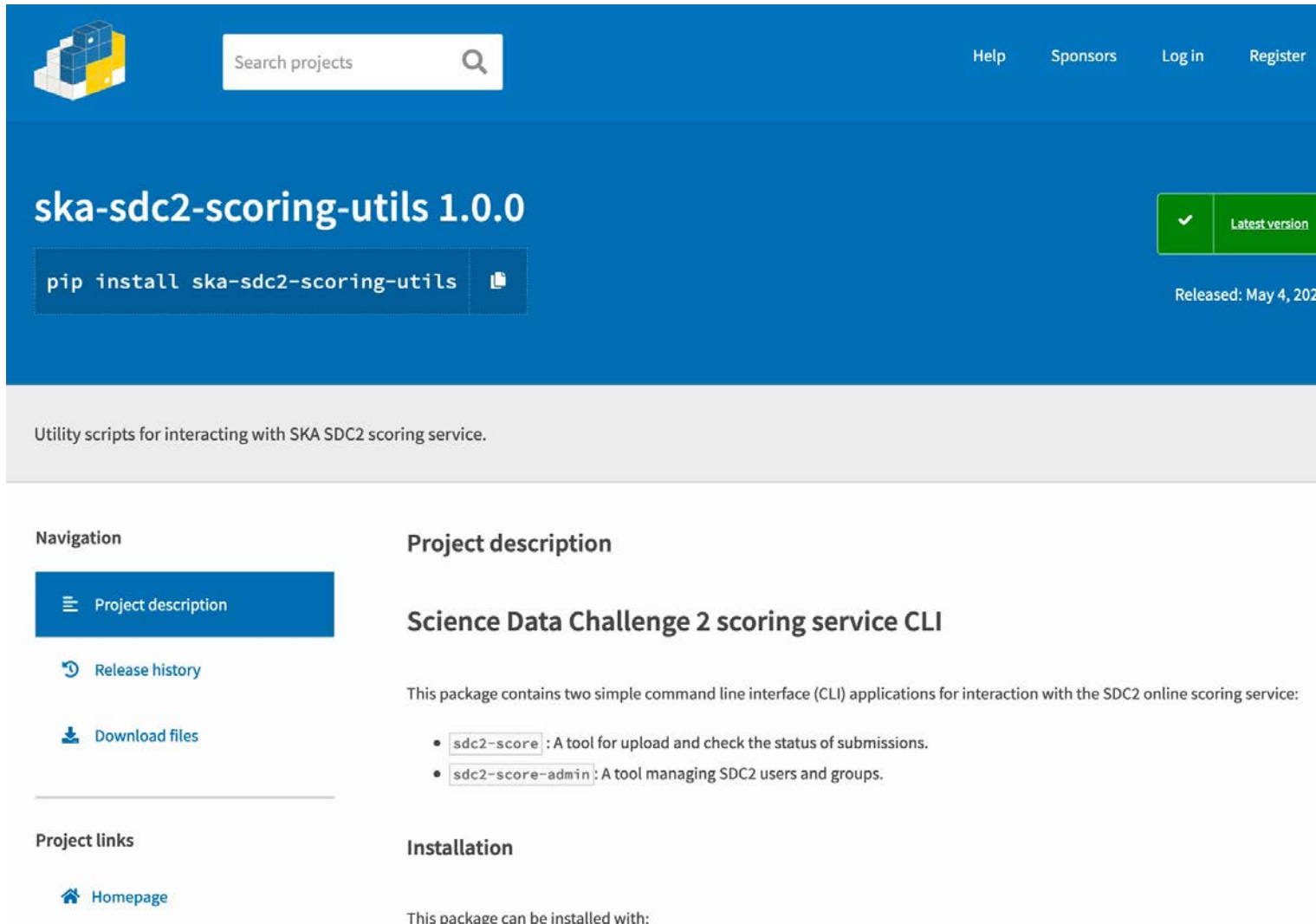
# SDC2 team telecons

- Main findings
  - Teams investigating varied solutions to challenge
    - Machine learning
    - Existing software
    - Bayesian
  - Many teams dividing work into source finding vs source characterization
  - Difficult to find time for some teams, but hope to ramp up soon



# Scoring service

<https://pypi.org/project/ska-sdc2-scoring-utils/>



The screenshot shows the PyPI project page for **ska-sdc2-scoring-utils 1.0.0**. The page has a blue header with a search bar and navigation links (Help, Sponsors, Log in, Register). Below the header, the project name and version are displayed, along with a green button indicating it is the "Latest version" and a release date of "Released: May 4, 2021". A code block shows the installation command: `pip install ska-sdc2-scoring-utils`. The main content area is divided into two columns. The left column contains a "Navigation" sidebar with links for "Project description" (selected), "Release history", and "Download files". The right column contains the "Project description" section, which states: "This package contains two simple command line interface (CLI) applications for interaction with the SDC2 online scoring service:". Below this, a bulleted list describes the tools: `sdc2-score` (for upload and status checks) and `sdc2-score-admin` (for managing users and groups). At the bottom of the right column is the "Installation" section, which states: "This package can be installed with:".

Utility scripts for interacting with SKA SDC2 scoring service.

**Navigation**

- Project description
- Release history
- Download files

**Project links**

- Homepage

**Project description**

### Science Data Challenge 2 scoring service CLI

This package contains two simple command line interface (CLI) applications for interaction with the SDC2 online scoring service:

- `sdc2-score`: A tool for upload and check the status of submissions.
- `sdc2-score-admin`: A tool managing SDC2 users and groups.

**Installation**

This package can be installed with:

# SKAO SDC2 team



Robert Braun



Anna Bonaldi



David Bastien



Philippa Hartley

Science



Lara Alegre



Rohini Joshi



James Collinson



Ben Mort



Rob Barnsley



Alex Clarke

Operations



Piers Harding

System

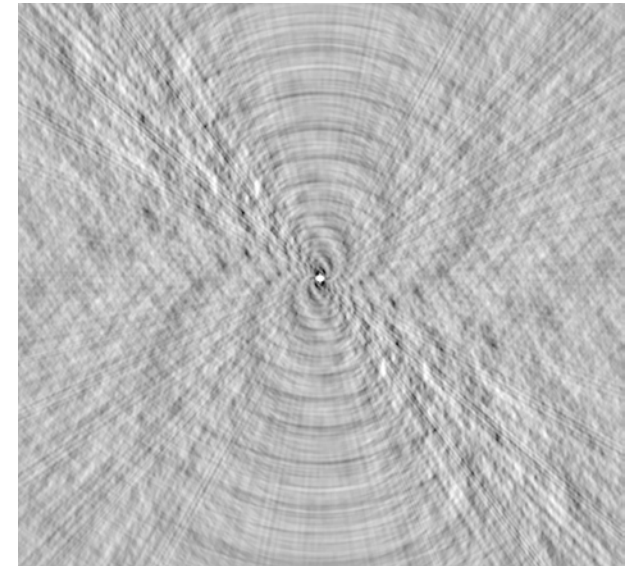
# Scoring service

[Demo of new version here](#)

- Allows self-scoring of challenge submissions
- Can be used for development datasets and full challenge dataset
- Automatically updates the leaderboard with full challenge scores
- The winning team will be the team with the highest position on the leaderboard at the time of the challenge closing (23:59:59 UTC+1 July 31<sup>st</sup>)
- Uses a web API to send catalogues to a remote host ([RESTful API](#))
- The remote host runs a Python script which performs cross-matching of source before determining accuracy of source characteristics
- Full description of scoring algorithm available on [SDC2 website](#)
- Additional modules can be added to the service so that it can be used for future data challenges

# Science Data Processor (SDP) alignment

- Potential to align with **SDP** work:
  - SDP currently working to understand and quantify effects of e.g. **imperfect calibration/RFI removal**
  - Future data challenges could incorporate these effects
  - Not yet computationally possible to generate at visibility level
  - Need to characterise for image plane
  - SDP timeline indicates next summer to investigate this
  - SDP keen to work with community to inform SDP software development



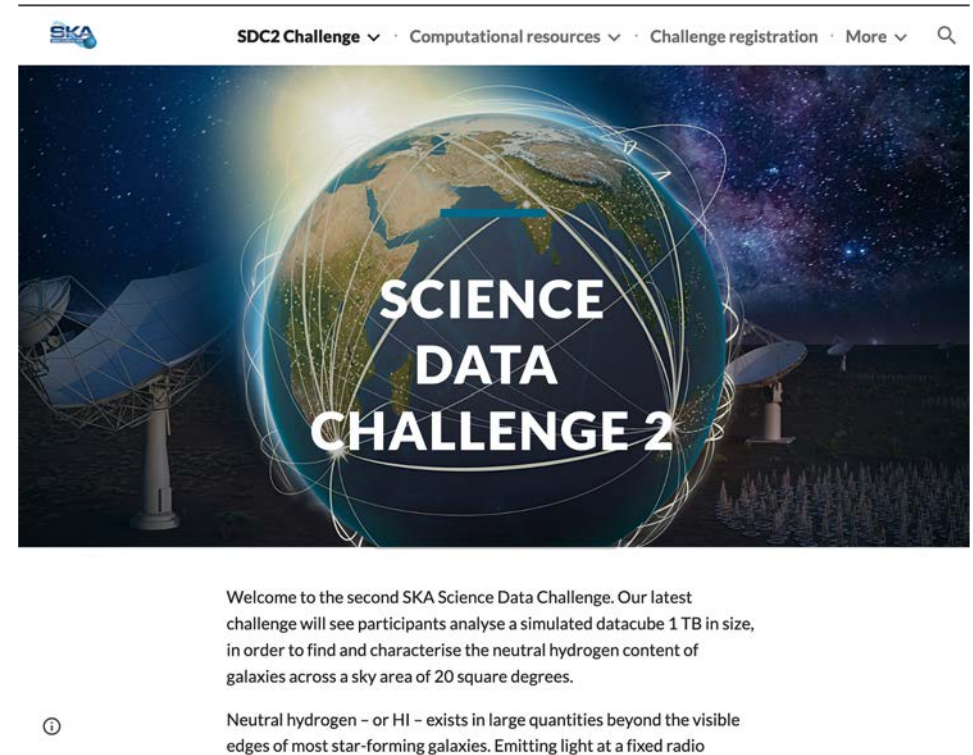


# Website



<https://sdc2.astronomers.skatelescope.org/sdc2-challenge>

- Test datasets and catalogues
- Full simulation description
- Details of the HPC facilities
- Challenge description and rules
- Reproducibility award criteria
- Discussion forum
- Scoring service and scoring algorithm description



## The Tiered Radio Extragalactic Continuum (T-RECS) simulation II: HI emission and continuum-HI cross-correlation.

Anna Bonaldi, Philippa Hartley

17 May 2021

### ABSTRACT

In this paper we extend the Tiered Radio Extragalactic Continuum Simulation (T-RECS) to include HI emission.

**Key words:** radio continuum; galaxies; galaxies: luminosity function, mass function, large-scale structure of Universe

### 1 INTRODUCTION

In section 2, measured or expected correlations between ra-

chrotron, free-free and thermal dust emission as a function of SFR.

## SKA Science Data Challenge 2: analysis and results

P. Hartley,<sup>1</sup>\* A. Bonaldi<sup>1</sup>

<sup>1</sup>SKA Organisation, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK

Accepted XXX. Received YYY; in original form ZZZ

### ABSTRACT

**Key words:**

### 1 INTRODUCTION

The Square Kilometre Array (SKA) project was born from an ambition to create a telescope sensitive enough to trace the formation of the first galaxies. Observing this era via the very weak emission from neutral hydrogen atoms will be possible only by using a collecting area of 1 square kilometre: large enough not only to provide a window onto *Cosmic Dawn* but – thanks to an order of magnitude increase in sensitivity over current instruments – also to explore new frontiers in galaxy evolution and cosmology, cosmic magnetism, the laws of gravity, extraterrestrial life and – in the strong tradition of radio astronomy (Wilkinson et al. 2004) – the unknown (see the SKA Science Book, Braun et al. (2015) for a comprehensive description of the full SKA science case).

in a very large number of overlapping sources that will require detection and classification.

In order to support the community to prepare for such rich datasets, the SKAO has established a series of Science Data Challenges (SDCs). Each challenge involves some combination of real or simulated datasets designed as closely as possible to represent future SKA data. The purpose of each challenge is then to exercise analysis methods needed to extract science from the data, with the goal of fostering new ideas and methods via wide participation and engagement. The challenges also aim to familiarise the community with the standard products that the SKA will deliver, providing the opportunity to test the validity of scientific proposals and to optimise survey. For these reasons, all SDC data products are made available publicly for the long-term footprint. Science Data

# SDC Recruitment



**SKAO** SKA Observatory Recruitment Portal

Search the SKAO website

JobsWho we areWhat we OfferOur LocationsHow to ApplyCandidate Menu

## Postdoctoral Researcher In Radio Astronomy Simulations (10125)

At the SKAO we are coordinating a global effort to deliver one of the largest science facilities on the planet. The SKAO telescopes will be next-generation instruments that will help to answer key questions in astrophysics, drive technological innovation and support human capital development.

More than a dozen partner countries and over a thousand scientists and engineers around the world are already on board, making the SKA project an international collaboration like no other, and one of the most ambitious science and engineering endeavours of the 21st century.

Headquartered in the UK, with telescope sites in Australia and South Africa, the SKAO will be one observatory operating two telescopes, an endeavour spanning three continents. Multinational by design, in early 2021 the SKAO became an intergovernmental organisation, one of only a select handful of science infrastructures to do so worldwide.

Our Global HQ is located near the buzzing city of Manchester at the historic Jodrell Bank Observatory, a UNESCO World Heritage site located in the Cheshire countryside. It is home to a rapidly expanding international team of more than a hundred, working in a highly collaborative and inclusive way, in coordination with SKAO teams in Australia and South Africa as part of a global astronomy and engineering community.

Being part of Team SKA means contributing to something extraordinary: a uniquely exciting project to advance human knowledge that will last for decades. In return, we offer a diverse multicultural working environment, 30 days annual leave plus public holidays, a generous pension scheme, regular social events, a family-friendly environment and much more. Come and join us!

A unique opportunity for a Postdoctoral research position in Radio Astronomy Simulations.

### The Role

The successful candidate will play a major part in developing solutions for the SKA Science Data Challenge in collaboration with existing Science and Operations team members. This initiative aims to release with a regular cadence increasingly realistic simulated SKA datasets and data products to the international scientific user community, together with a specific data analysis exercise for them to solve (the "challenge") and a set of tools and services to support the completion of the exercise and the evaluation of the results.

The post holder will report to the SKA Project Scientist and be responsible for the coordination of the SKA Science Data Challenges.

The role is a 3-year fixed term contract. Full-time hours will be 35 hours per week and flexible working may be considered compatible with reviews, milestones and office activities.

**Apply For This Job**

**Send To A Friend**

**Job Location**  
UK (Head Office)

**Area / Department**  
Science

**Contract Type**  
Fixed Term Contract

**Salary**  
Competitive salary with generous pension and leave benefits

**Closing Date**  
31/08/2021

**Ref No**  
10125

<https://recruitment.skatelescope.org/vacancy/postdoctoral-researcher-in-radio-astronomy-simulations-440429.html>



# AOB



- ???





# SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

