SWG Science Update

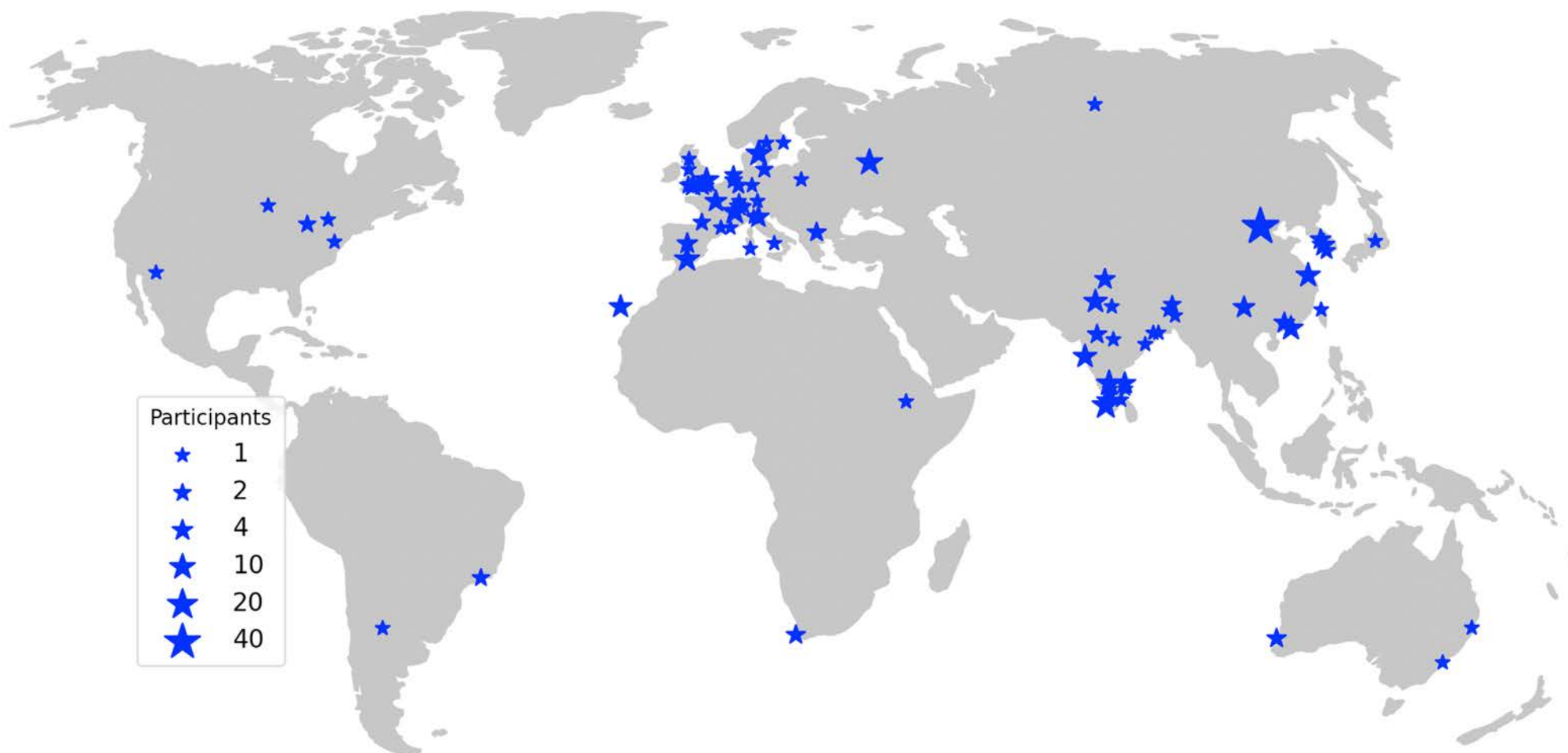**Robert Braun, Science Director**

18 May 2021

Science Activity Updates

- Science Data Challenges (Philippa, Anna)

- AOB

Science Data Challenge 2
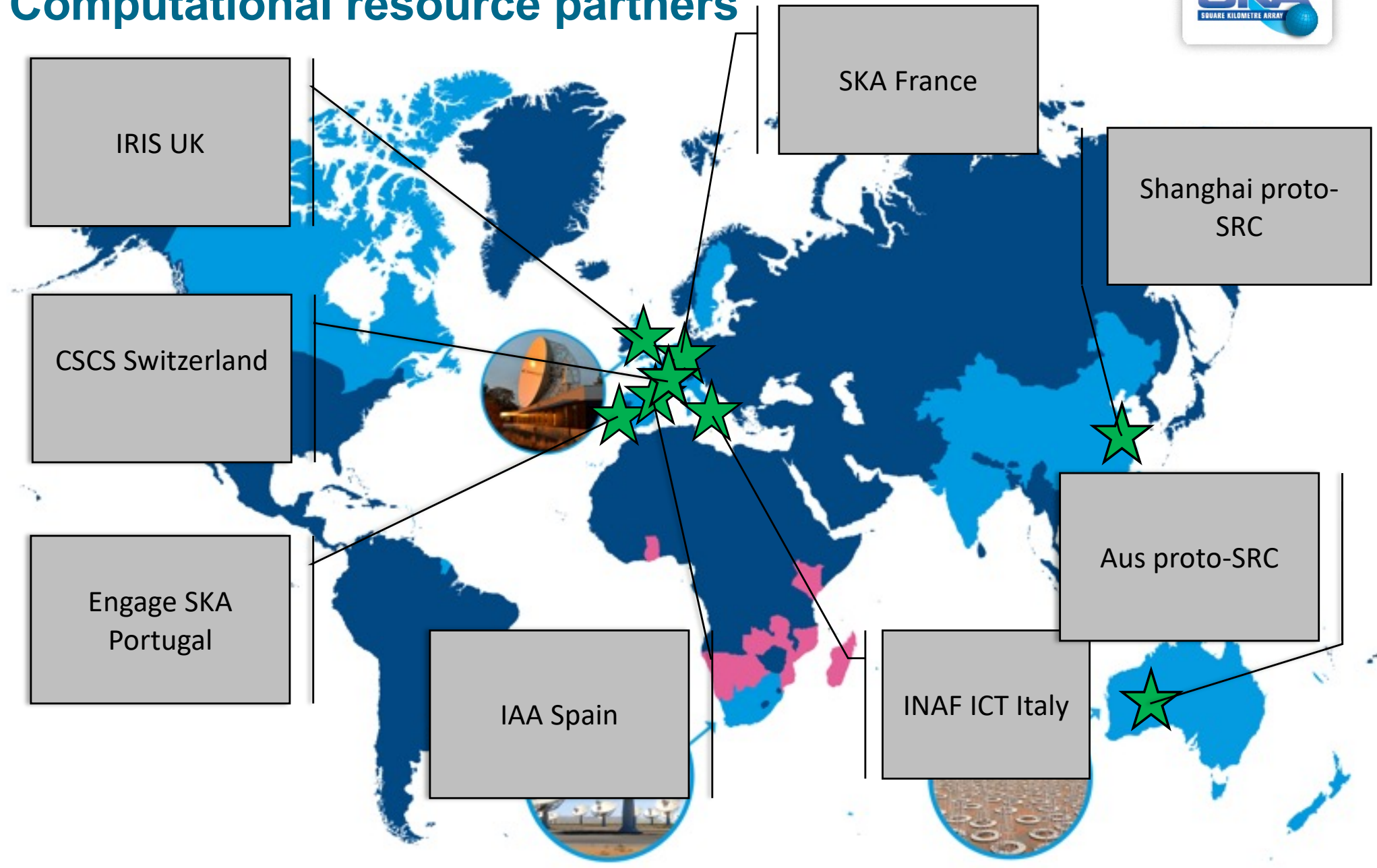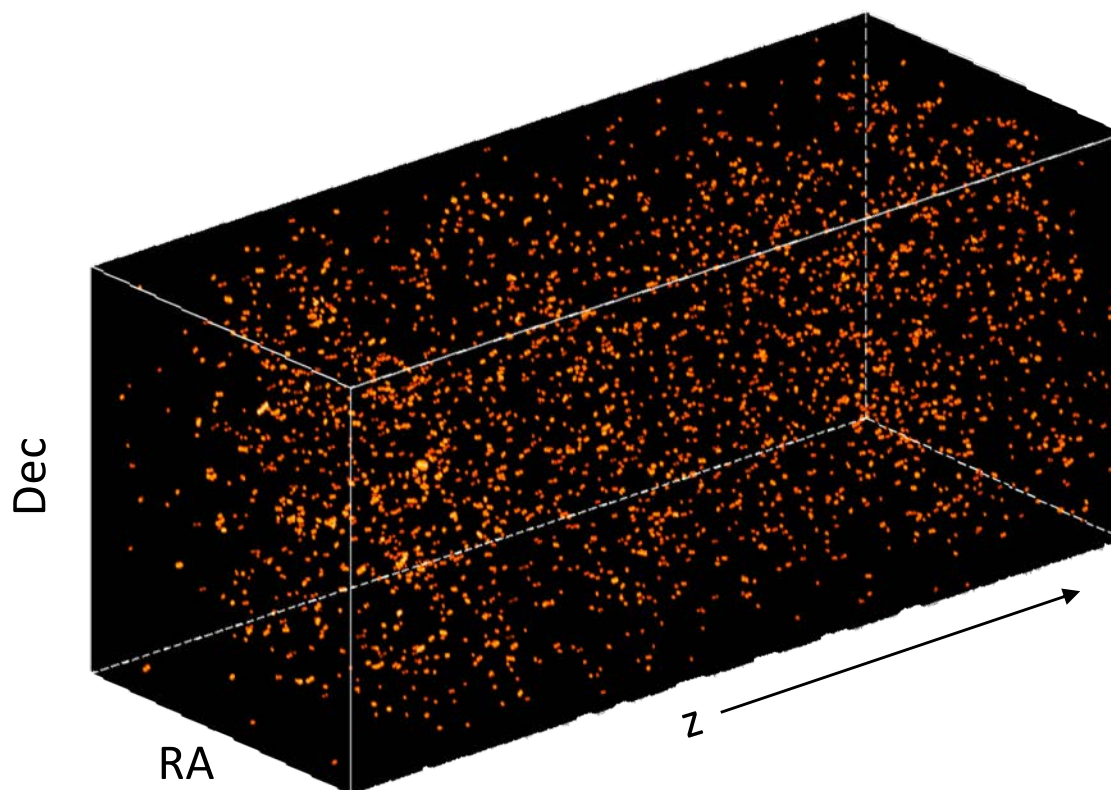## Duration Feb 1st – July 31st

Computational resource partners

IRIS UK

CSCS Switzerland

Engage SKA Portugal

SKA France

Shanghai proto-SRC

Aus proto-SRC

IAA Spain

INAF ICT Italy

• (Fairly) big data

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

Dec

RA

z

- Integration time = 2000h
- Spatial/Frequency resolution = 7 arcsec / 30 kHz
- RMS per channel 13 – 18 μJy
- 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 **6th-20th 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
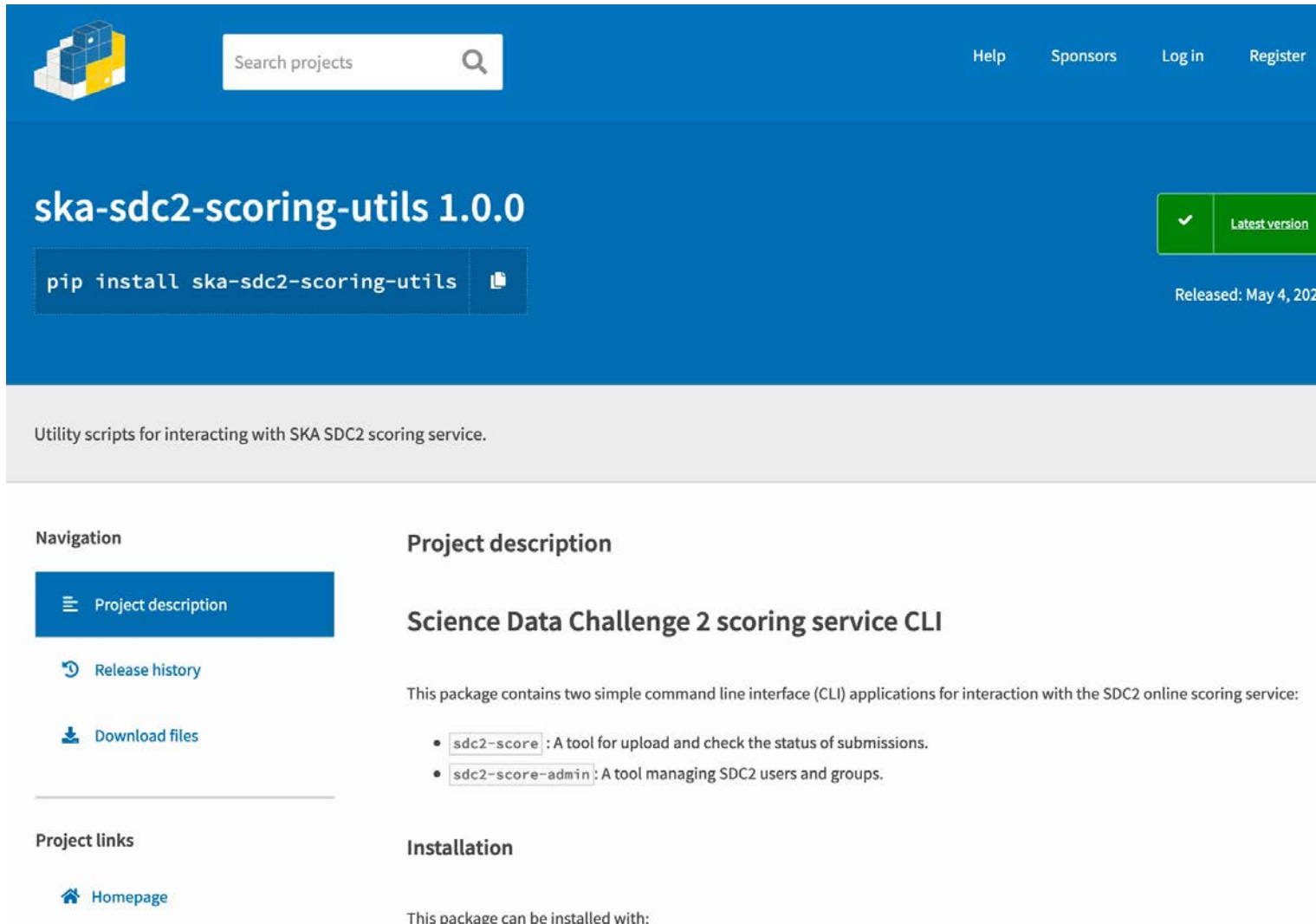
SDC2 team telecons

- 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/

SKAO SDC2 team

Anna Bonaldi

Robert Braun

David Bastien

Philippa Hartley

Science

Lara Alegre

Rohini Joshi

James Collinson

Ben Mort

Alex Clarke

Rob Barnsley

Piers Harding

Operations

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 31st)

- 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
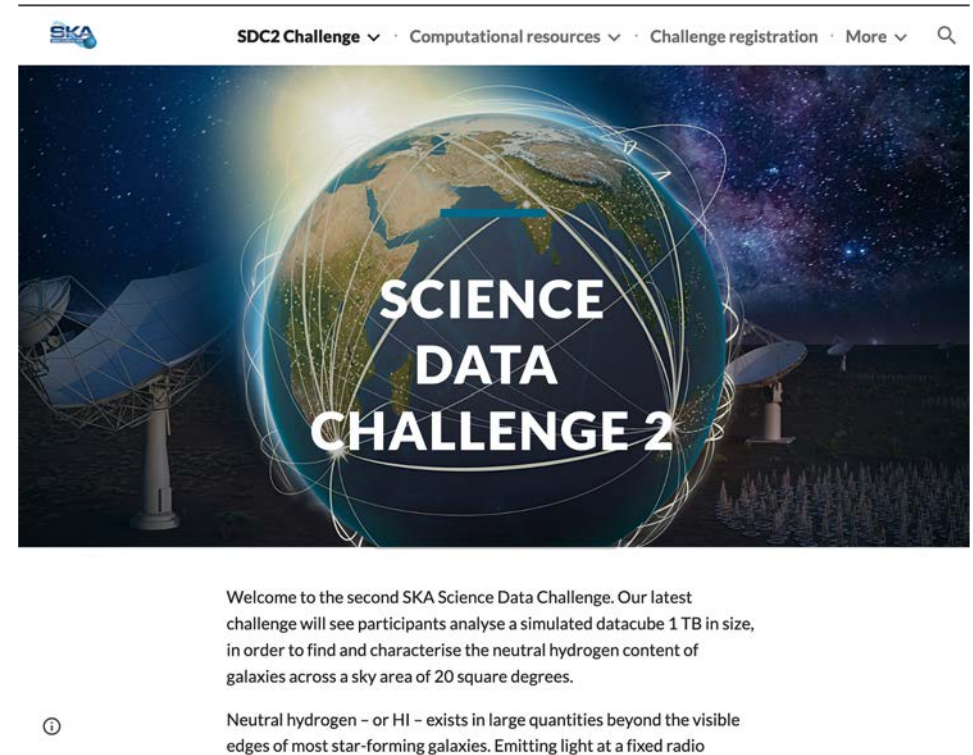 - SDP keen to work with community to inform SDP software development

Website
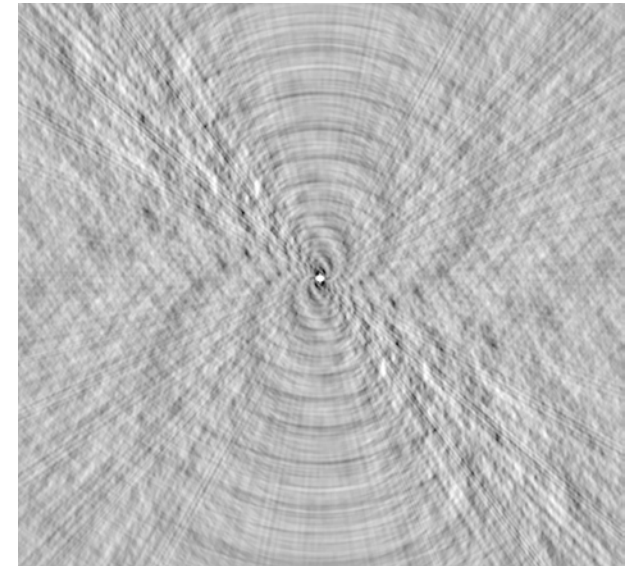
https://sdc2.astronomers.skatelescope.org/sdc2-challenge

- 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

SDC2 Challenge ∨ · Computational resources ∨ · Challenge registration · More ∨

SCIENCE DATA CHALLENGE 2

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

SDC Papers

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,[1]* A. Bonaldi[1]
[1]SKA Organisation, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK

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 footnote. Science Data

SDC Recruitment

AOB

- ???

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope