



SKA SWG Update

Robert Braun, SKAO Science Director

21 September 2021



SKA Science Update

- Proposal allocation modelling
- SDC2: and the winner is ...
- AOB



SKA Proposal Allocation Modelling: The Challenge

- Simultaneously satisfy conflicting access constraints:
 1. Scientific Merit
 2. SKAO Members "share of the project"
 3. Fixed (and relatively small) Open Time fraction



SKA Proposal Allocation Modelling : Measuring “Access”

- SKAO “access” is more than just telescope time; also includes SDP processing, ODP transmission to SRC network, ...
- Needs full accounting, including suitable consideration of commensal access
- Models for measuring SKAO Member (and RoW) access:
 1. PI-only
 2. **All proposal roles (PI, MT and co-I) with weights proportional to role rarity**
 - Most fair and least prone to “gaming”
 3. Only PI and MT roles (for KSP proposals)



SKA Proposal Allocation Modelling : Adjusting “Access”

- In the **likely** event that Member access targets are not achieved based solely on scientific merit, then some adjustment will be needed
- Iteratively modify initial (science) proposal rankings based on current misfit to targets

- SKAO Members:

Compete *individually*
for (eg.) OT/4

$$W_{CJ} = \begin{cases} \left(\frac{F'_{CJ}}{S_C} \right)^\alpha & \text{if } F'_{CJ} < S_C, \\ 1 & \text{if } S_C < F'_{CJ} < (S_C + O/4), \\ \left(\frac{F'_{CJ}}{S_C + O/4} \right)^\alpha & \text{if } F'_{CJ} > (S_C + O/4). \end{cases}$$

- Non-members:

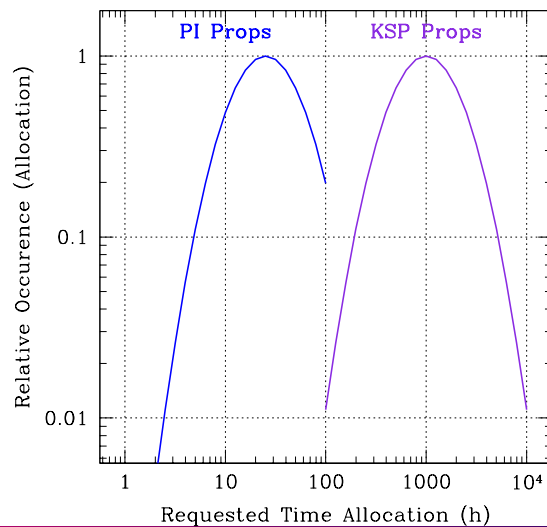
Compete *jointly*
for (eg.) OT/2

$$W_{RoWJ} = \begin{cases} 1 & \text{if } F'_{RoWJ} < O/2, \\ \left(\frac{F'_{RoWJ}}{O/2} \right)^\alpha & \text{if } F'_{RoWJ} > O/2. \end{cases}$$



SKA Proposal Allocation Modelling : Assumed Distributions

- Use SWG membership for science topic and country of affiliation distributions
- Randomly draw science topics and proposal teams
- Log-normal for durations



	CON	CSP	COS	COL	EOR	GRW	HIG	MAG	NSL	GAL	PUL	SHI	TRA	VLB	Total	CtyRep
AU	13	3	8	5	9	2	19	11	2	7	8	2	10	11	110	9.9%
CA	3	0	8	5	1	1	5	6	2	7	4	0	3	3	48	4.3%
CN	4	3	6	2	2	7	3	2	0	2	7	5	5	7	55	4.9%
FR	6	4	10	7	10	4	8	6	5	3	6	2	4	2	77	6.9%
DE	5	7	9	6	2	2	2	10	4	6	11	8	1	5	78	7.0%
IN	9	0	6	1	7	1	0	4	0	1	4	6	2	2	43	3.9%
IT	14	0	16	9	9	10	6	9	3	8	4	1	5	7	101	9.1%
NL	9	5	3	3	5	10	8	3	5	5	5	2	11	14	88	7.9%
PT	6	0	2	1	2	0	1	0	0	1	0	1	1	1	16	1.4%
ZA	6	0	17	1	5	2	4	2	0	0	3	0	2	4	46	4.1%
ES	5	1	7	8	3	3	4	6	0	2	2	0	4	4	49	4.4%
SE	3	0	4	8	5	1	2	4	6	8	0	0	3	5	49	4.4%
CH	0	0	5	1	0	3	0	1	2	1	0	1	0	0	14	1.3%
UK	21	4	33	10	8	8	7	8	12	9	9	21	10	9	169	15.2%
JP	2	0	5	1	6	1	3	3	1	3	3	1	2	5	36	3.2%
KR	0	0	1	0	2	1	0	1	0	0	0	0	0	1	6	0.5%
NZ	2	0	0	1	0	0	1	0	0	0	1	0	0	1	6	0.5%
US	5	4	7	10	7	6	3	1	0	2	10	10	8	1	74	6.6%
HR	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.1%
IR	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0.2%
PL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.1%
TW	1	0	1	0	1	0	2	0	0	0	0	0	0	0	5	0.4%
BE	0	3	0	0	0	0	0	0	0	0	0	1	0	0	4	0.4%
KE	0	1	0	0	0	0	0	0	0	0	0	1	0	0	2	0.2%
RU	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2	0.2%
FI	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2	0.2%
CL	0	0	0	1	0	0	0	0	0	0	1	1	0	0	3	0.3%
IL	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3	0.3%
MX	0	0	0	0	0	0	0	1	0	1	0	0	0	1	3	0.3%
PL	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.1%
AE	0	0	0	0	0	0	0	0	2	0	1	0	0	0	3	0.3%
AT	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0.1%
GR	0	0	0	0	0	0	0	0	0	1	0	2	0	0	3	0.3%
TH	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.1%
BR	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.1%
BG	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.1%
CZ	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	0.3%
IQ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.1%
IE	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	0.3%
MZ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.1%
NG	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.1%
NO	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.1%
HU	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.1%
RoW	11	9	9	13	9	7	6	3	3	5	14	27	9	5	130	11.7%
Total	117	36	149	81	85	63	78	79	45	68	80	77	72	85	1115	100.0%
SciRep	10.5%	3.2%	13.4%	7.3%	7.6%	5.7%	7.0%	7.1%	4.0%	6.1%	7.2%	6.9%	6.5%	7.6%	100.0%	



SKA Proposal Allocation Modelling: Early Indications

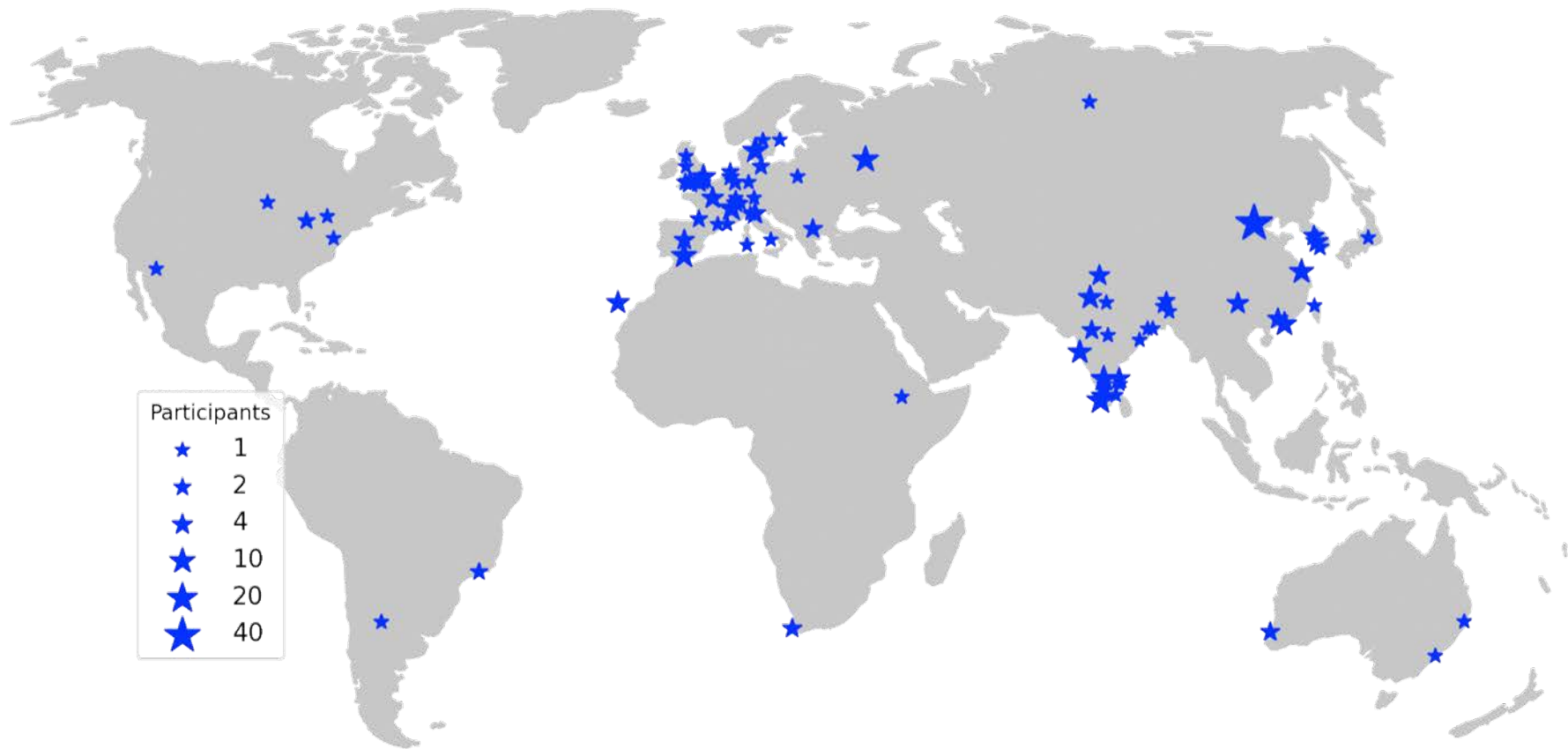
- Access measurement that includes all proposal roles (and not just the PI) is viable and is preferred
- Reasonable prospects to simultaneously satisfy access constraints with some precision (despite highly coupled problem) and very large number of constraints
- KSP access balance more challenging (than PI case) due to smaller number of allocated proposals
- Greater precision in yielding targets comes at the expense of reduced weighting of scientific merit
- KSP and PI access should be considered jointly rather than individually



Science Data Challenge 2

Duration Feb 1st – July 31st

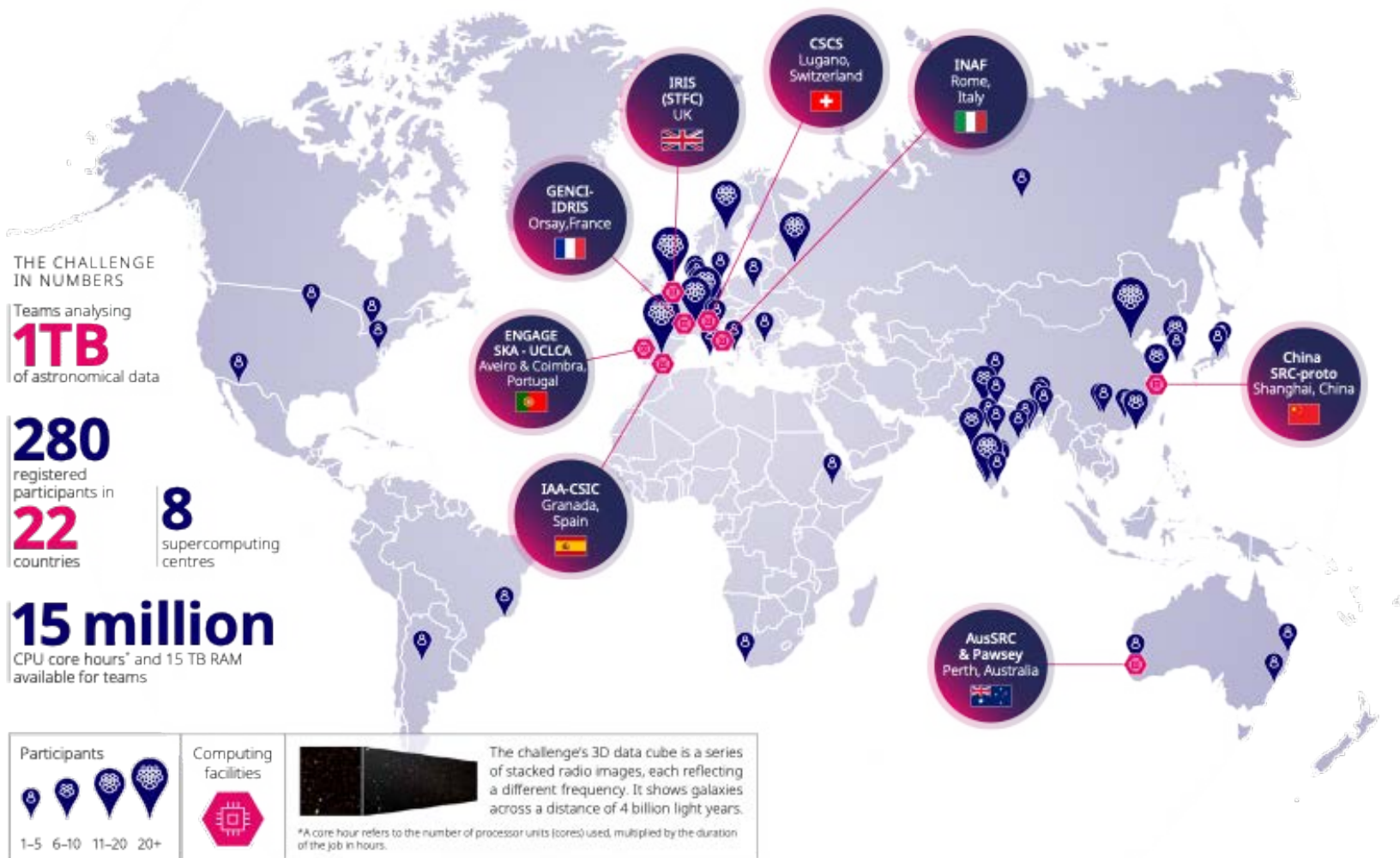
- 40 Teams with total of 276 participants, from 80 institutes in 23 countries



SKAO Science Data Challenge 2



MAP OF WORLDWIDE PARTICIPATION



SDC2, and the winner is ...

- About half of the 40 registered teams undertook significant analysis of the SDC2 data products
- 12 teams made final submissions
- Large range of scores
 - Score measures both goodness of fit for true detections and absence of false positives
- Top two teams used non-traditional methods, including internal cross-correlation of sub-teams for winner

Property	Error term	Threshold value
Sky position (RA, Dec) x, y	$D_{pos} = \frac{\sqrt{(x-x')^2 + (y-y')^2}}{S'}$	0.3
HI size S	$D_{HI\ size} = \frac{ S-S' }{S'}$	0.3
Integrated line flux F	$D_{flux} = \frac{ f-f' }{f'}$	0.1
Central frequency ν	$D_{freq} = \frac{ \nu-\nu' }{\nu'}$	0.3
Position angle θ	$D_{PA} = \theta - \theta' $	10.0
Inclination angle i	$D_{incl} = i - i' $	10.0
Line width w_{20}	$D_{line\ width} = \frac{ w_{20}-w_{20}' }{w_{20}'}$	0.3

$$\text{final score} = \sum_{i=1}^{N_{match}} w_i - N_{false}$$



SDC2, and the winner is ...

- About half of the 40 registered teams undertook significant analysis of the SDC2 data products
- 12 teams made final submissions
- Large range of scores
 - Score measures both goodness of fit for true detections and absence of false positives
- Top two teams used non-traditional methods, including internal cross-correlation of sub-teams for winner

LEADERBOARD

The live SDC2 leaderboard below reports the highest score for the full challenge dataset submitted by each team

Position	User	Group	Score	Date
1	minerva	MINERVA	23254.16	2021-07-31T22:08:25.716098
2	forska	FORSKA-Sweden	22489.43	2021-07-14T05:29:44.394263
3	sofia	SoFiA	16822.24	2021-07-27T02:35:21.234327
4	naoc-tianlai	NAOC-Tianlai	14416.02	2021-07-28T12:59:39.209828
5	hi-friends	HI-FRIENDS	13902.62	2021-07-31T20:39:01.416127
6	epfl	EPFL	8515.16	2021-07-31T20:30:40.569408
7	spardha	Spardha	5614.59	2021-07-30T13:54:14.229580
8	starmech	Starmech	2095.65	2021-07-31T15:42:40.105279
9	jlrat	JLRAT	1079.73	2021-07-31T18:13:38.347097
10	coin	Coin	-1.76	2021-07-31T22:48:57.226716
11	hiraxers	HIRAXers	-2.00	2021-07-15T10:55:52.222569
12	shao	SHAO	-471.00	2021-07-31T16:14:46.451245



SDC2 Next Steps

- Collecting full submission packages (fully documented, portable and customisable SDC2 analysis pipelines) for reproducibility award assessment
- Preparing announcement of winners (mid-October, to coordinate with institutes)
- Writing up SDC2 paper
 - Methods employed
 - Relative performance
 - Lessons learned for analysis pipelines
- Post SDC2 analysis and follow-up of HPC Data Centre participation
 - Securing of HPC support for future SDCs
 - Lessons learned for SRC network
- Investigate causes of team attrition



Any Other Business

- Upcoming meetings
- ...???



Thank you

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



www.skao.int