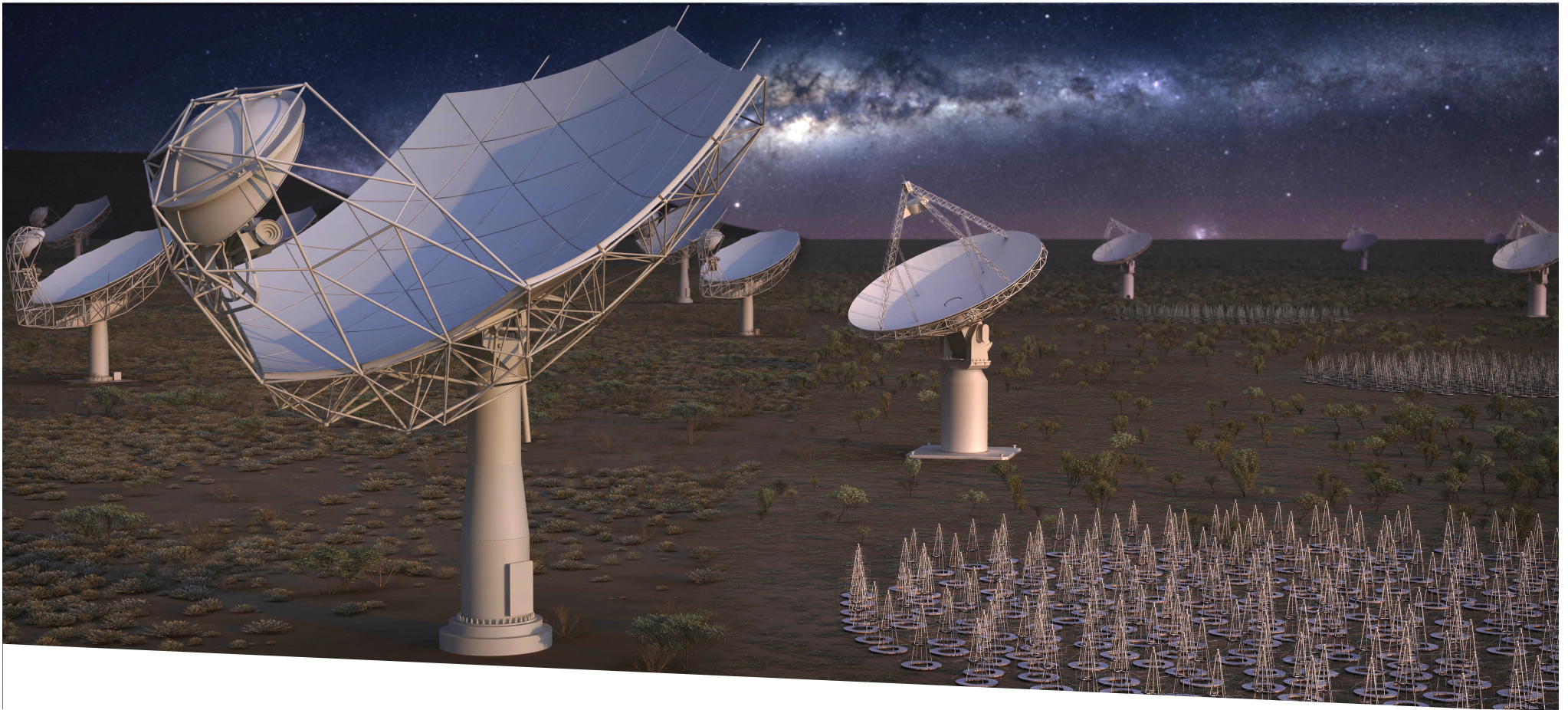# SKA Science data challenges

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope
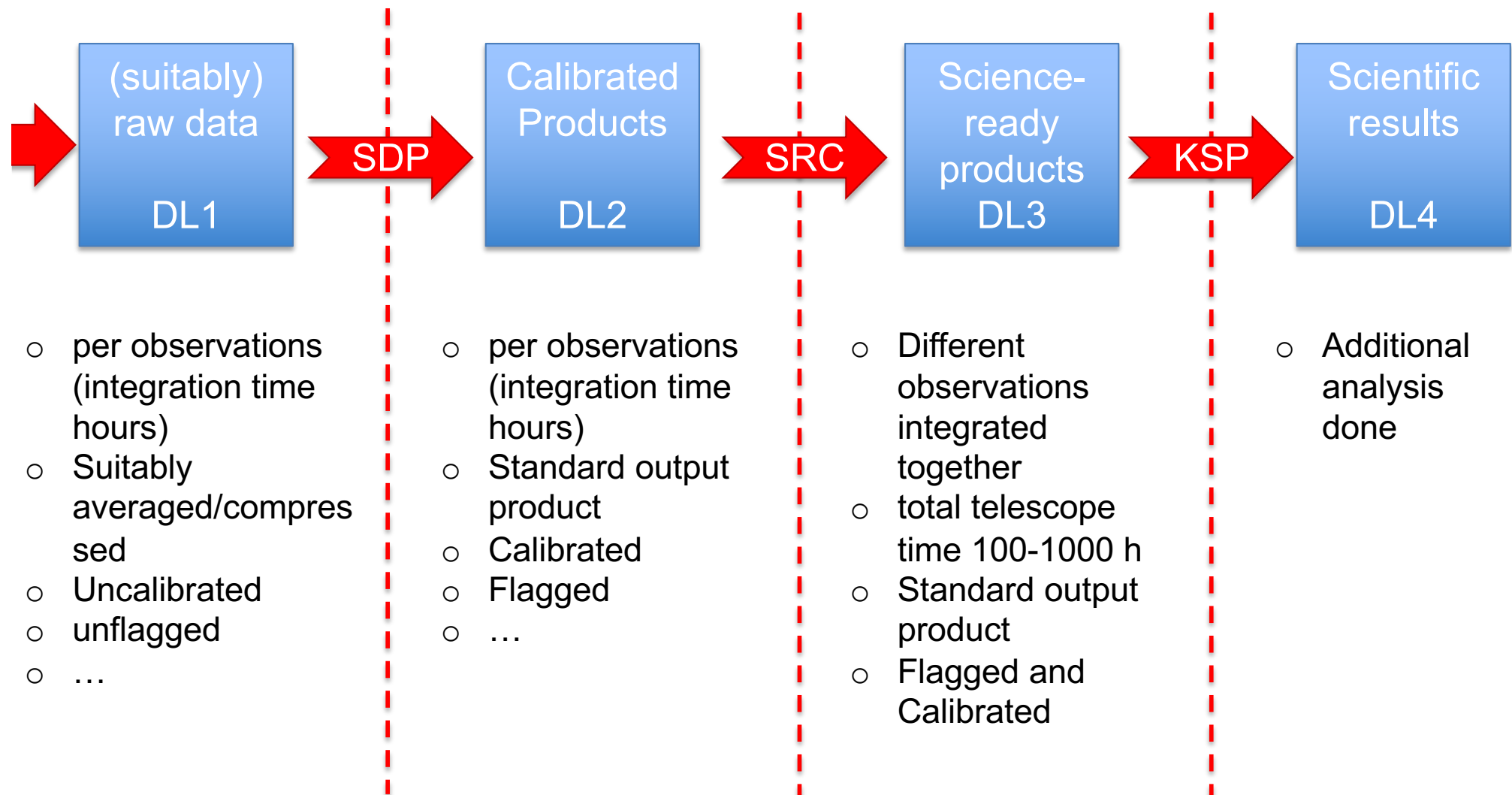
A. Bonaldi

Project scientist
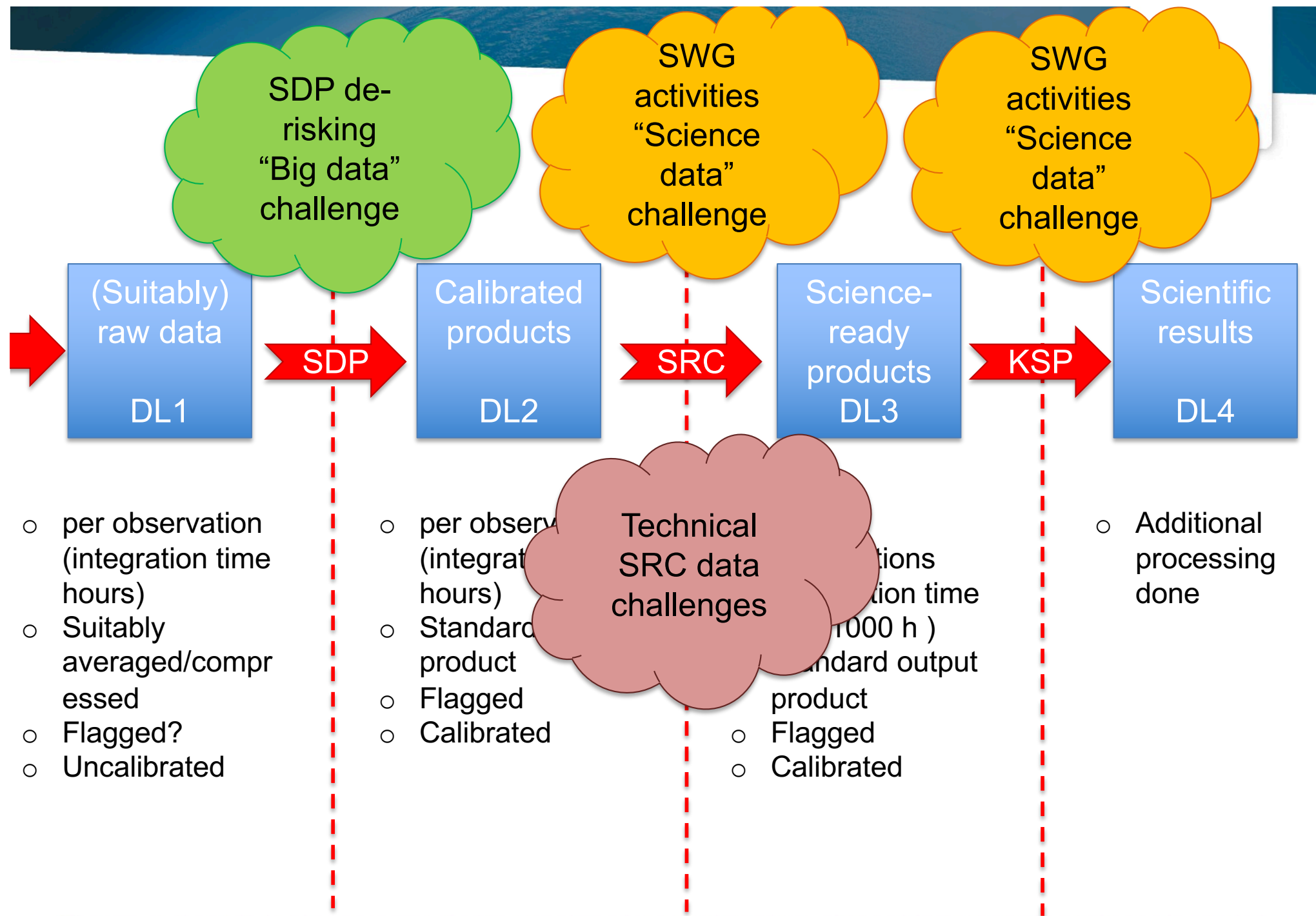
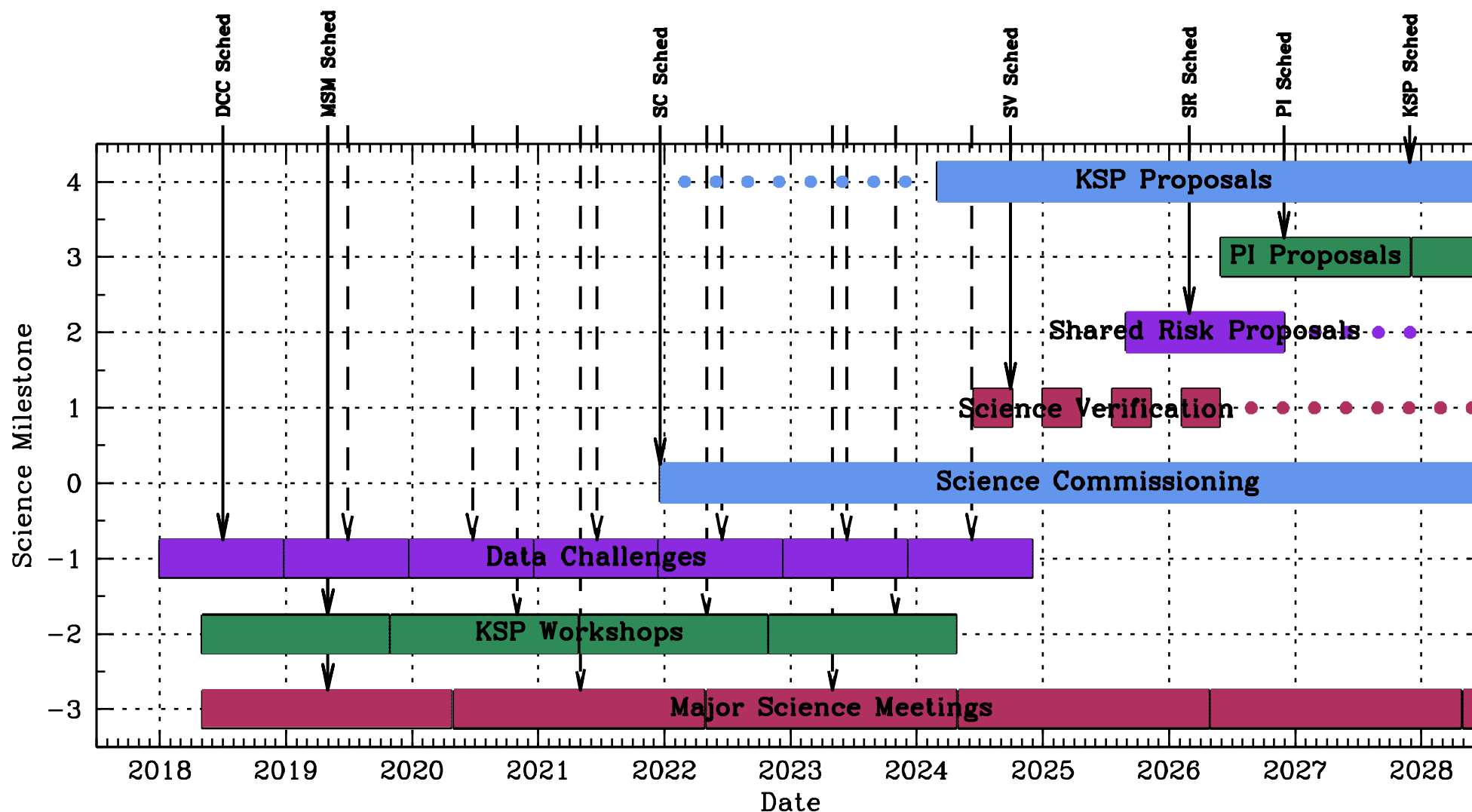AENEAS all-hands meeting 6 /3/19

SKA data challenges: Data Layers (DL)

DL0=raw data out of CSP

| (suitably) raw data DL1	→ SDP →	Calibrated Products DL2	→ SRC →	Science-ready products DL3	→ KSP →	Scientific results DL4

- o per observations (integration time hours)
- o Suitably averaged/compressed
- o Uncalibrated
- o unflagged
- o …

- o per observations (integration time hours)
- o Standard output product
- o Calibrated
- o Flagged
- o …

- o Different observations integrated together
- o total telescope time 100-1000 h
- o Standard output product
- o Flagged and Calibrated

- o Additional analysis done

SDP de-risking "Big data" challenge

SWG activities "Science data" challenge

SWG activities "Science data" challenge

(Suitably) raw data
DL1

SDP →

Calibrated products
DL2

SRC →

Science-ready products
DL3

KSP →

Scientific results
DL4

Technical SRC data challenges

- per observation (integration time hours)
- Suitably averaged/compressed
- Flagged?
- Uncalibrated

- per observ (integrat hours)
- Standard product
- Flagged
- Calibrated

- tions tion time 1000 h) ndard output product
- Flagged
- Calibrated

- Additional processing done

Exploring the Universe with the world's largest radio telescope

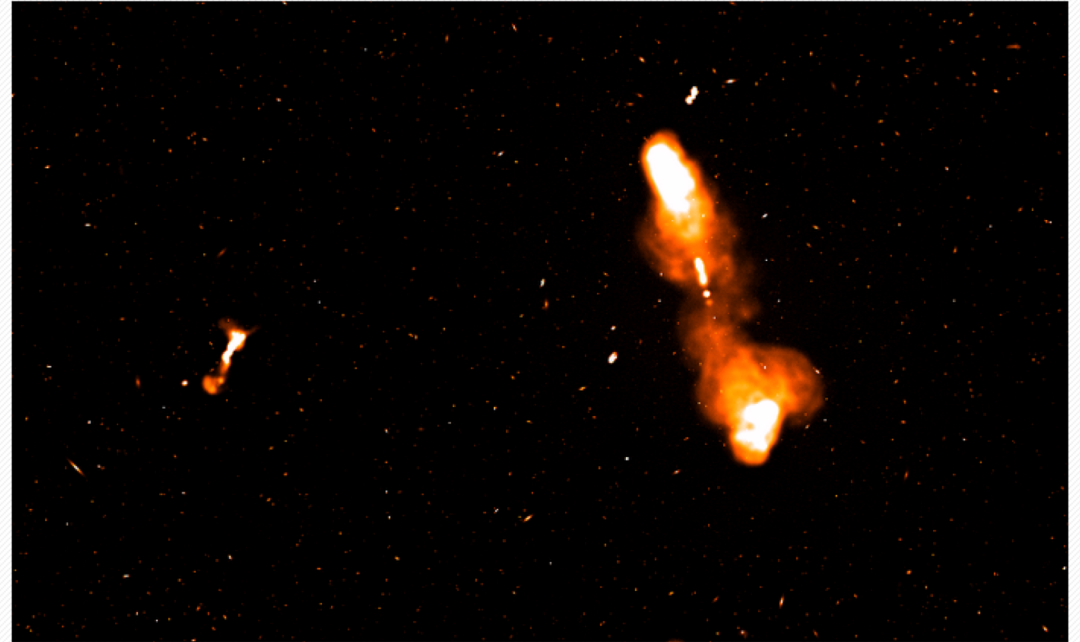# SKA science Timeline

SKA data challenges: simulations
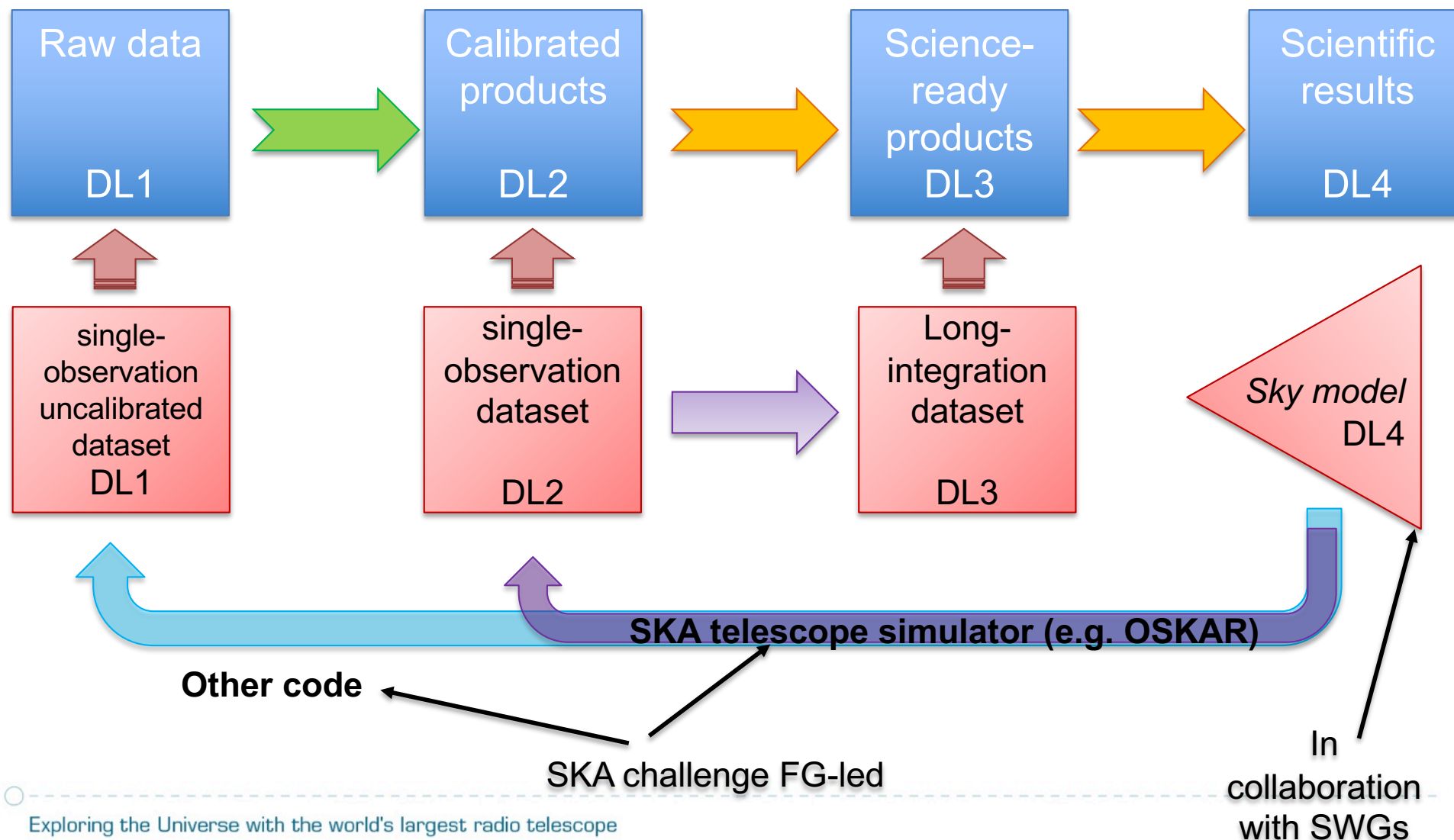
Science data challenge 1 (SDC1)

- Science-ready (SRC) imaging product

- Radio continuum, SKA Mid

- Not too challenging data sizes

- 1 pointing, 3 freqs, 3 depths

- Source finding

- Source identification, classification & characterization

➕ Print this page

SKA Launches First Science Data Challenge For Astronomy Community

A snapshot from the SKA Science Data Challenge image, showing a large Active Galactic Nucleus (AGN) as if observed by SKA-mid at 1.4 GHz. (Credit: SKA Organisation)

SKA Global Headquarters, 26 November 2018 – The Square Kilometre Array Organisation (SKAO) is today releasing its first ever Science Data Challenge, giving astronomers a taste of the highly detailed images the SKA will produce.

Developed by the SKAO's Project Science team, the challenge requires the analysis of a series of high resolution images created through data simulations. Researchers are invited to download the images and use their own software to find, identify and classify the sources.

The key aim of the series of Data Challenges is to prepare the science community for the kind of data products they will receive from SKA observations, and to gather valuable feedback which will inform the development of data reduction procedures.

Why SDC1?

❖ SKA unique features map into the data products:

- ✓ In the image plane, not visibilities
- ✓ "Benign" dirty beam
- ✓ Deconvolved down to 8h exposures
- ✓ Very deep -> confusion limited
- ✓ Very big number of sources to detect and classify

❖ SDC1 goals:

- ✓ Get the community familiar with this data product
- ✓ Develop efficient methods for source finding and source characterization -> SWG and SRC applications
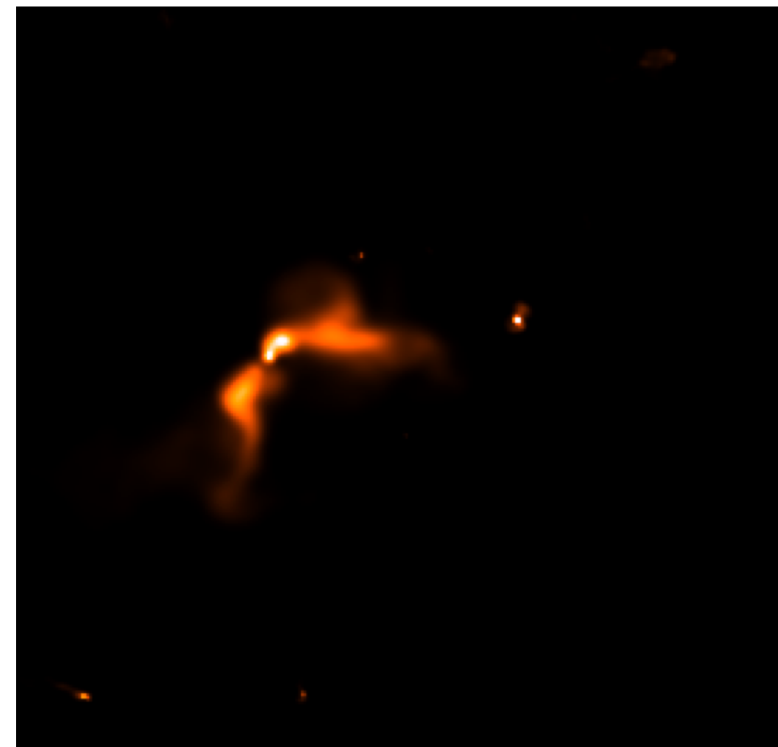
SKA data volumes

SKA UV coverage

SKA pipelines

SKA sensitivity

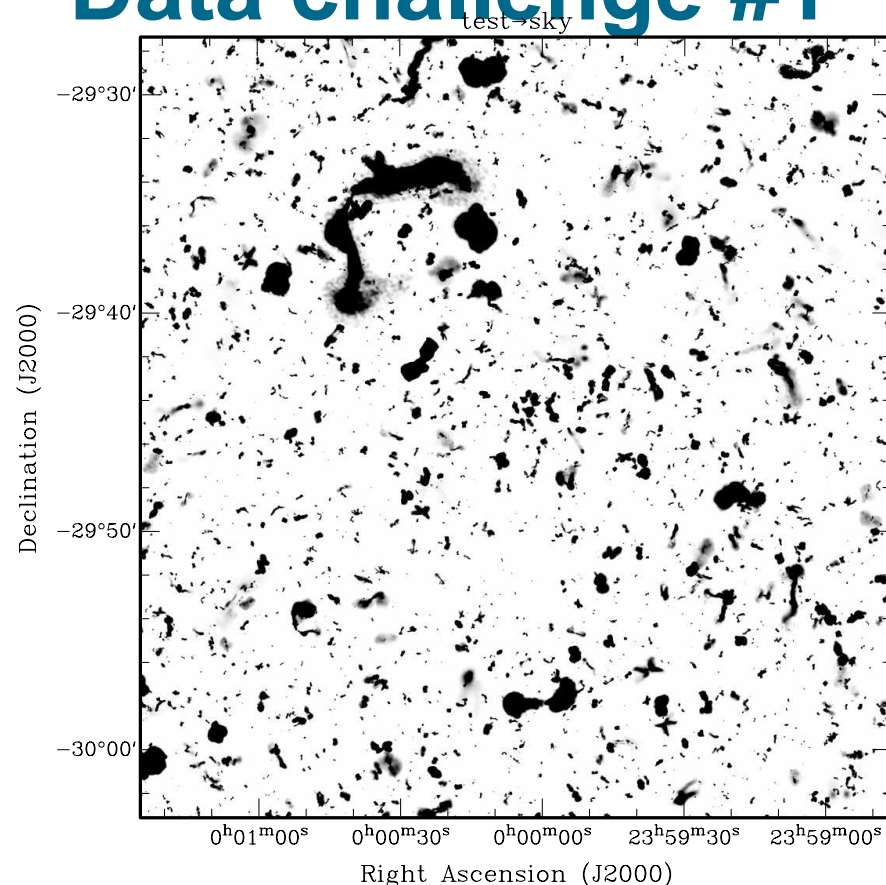# Data challenge #1

- ## Sky model:
  - ### List of sources: from T-RECS (Bonaldi et al. 2018)
    - AGNs and SFGs, spectra from 150 MHz-20 GHz
  - ### Image of the sky:
    - Galsim-based package by I. Harrison to draw SFGs, modelled as Sersic profiles
    - An atlas of real AGN images is used to add AGNs to the sky model
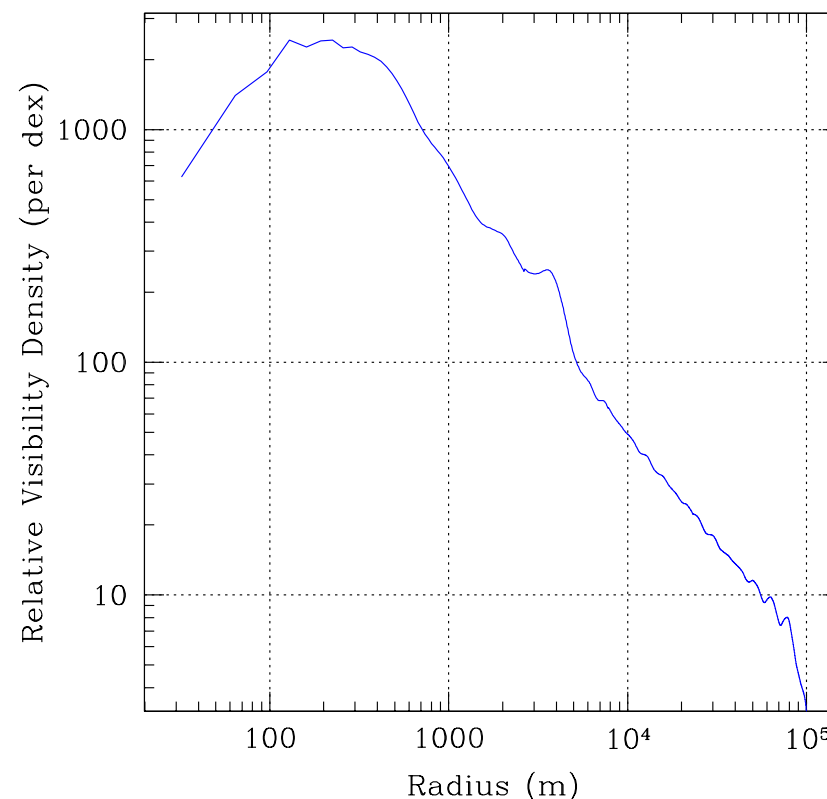

Zoom-in of AGN skymodel map

Data challenge #1

test→sky

SKA1–MID Visibility Distribution

- ## Telescope simulation:
  - – MIRIAD-based simulation code
  - – Using actual SKA dish voltage pattern and SKA1-Mid configuration
  - – Have developed new gridding algorithms to cope with data density contrast vs scale

SDC1 communication strategy

- Email to all SKA SWG participants
- SKA talks in various meetings/conferences
- Press release trough SKA public website
  - through SKA Communications Network (SKACON), posted on social media by SKA Canada, SKA Australia, ASTRON, CSIRO, STFC, SKA India and many other partners
- White paper arXiv:1811.10454, to reach astronomical community at large

Exploring the Universe with the world's largest radio telescope

Square Kilometre Array Science Data Challenge 1

ANNA BONALDI & ROBERT BRAUN, FOR THE SKAO SCIENCE TEAM *

SKA Organization, Jodrell Bank, Lower Withington, Macclesfield, Cheshire, SK11 9DL, United Kingdom
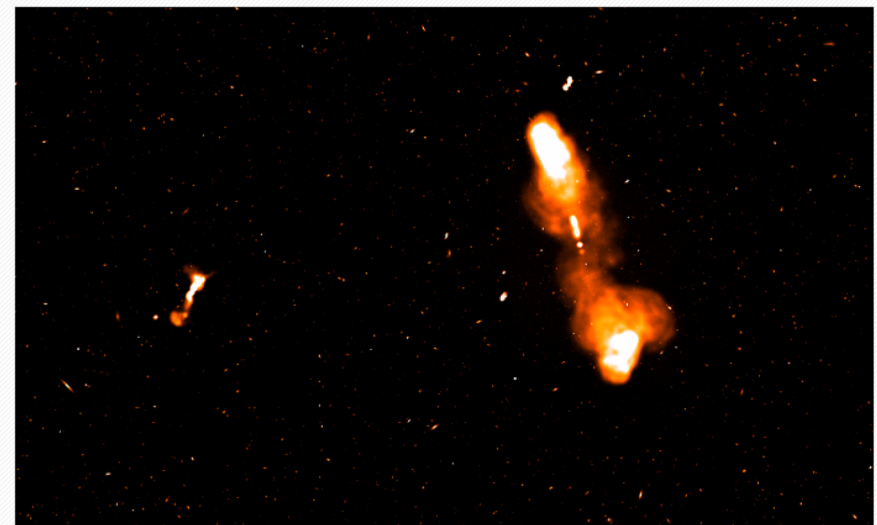
November 27, 2018

Abstract

The Square Kilometre Array (SKA, https://skatelescope.org) will be the world's largest radio telescope. SKA Science Data Challenges will be regularly issued to the community as part of the science preparatory activities. The purpose of these challenges is to inform the development of the data reduction workflows, to allow the science community to get familiar with the standard products the SKA will deliver, and optimise their analyses to extract science from them. These challenges may consist of real data from currently operating radio facilities or of simulated SKA data. The purpose of this document is to provide information on how the SKA Science data challenge #1 (SDC1) has been produced and to set the challenge for the community. For more information on how to take part in the challenge and to download the data see https://astronomers.skatelescope.org/ska-science-data-challenge-1/

A snapshot from the SKA Science Data Challenge image, showing a large Active Galactic Nucleus (AGN) as if observed by SKA-mid at 1.4 GHz. (Credit: SKA Organisation)
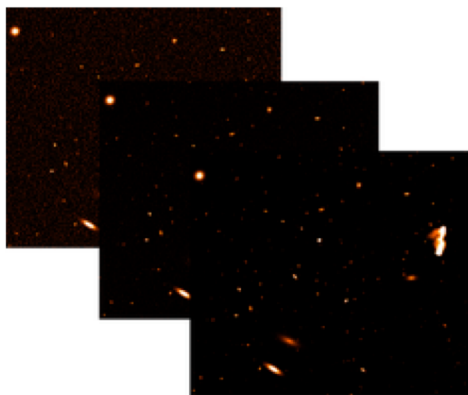
SKA Global Headquarters, 26 November 2018 – The Square Kilometre Array Organisation (SKAO) is today releasing its first ever Science Data Challenge, giving astronomers a taste of the highly detailed images the SKA will produce.

Developed by the SKAO's Project Science team, the challenge requires the analysis of a series of high resolution images created through data simulations. Researchers are invited to download the images and use their own software to find, identify and classify the sources.

The key aim of the series of Data Challenges is to prepare the science community for the kind of data products they will receive from SKA observations, and to gather valuable feedback which will inform the development of data reduction procedures.

1 [astro-ph.IM] 26 Nov 20

SKA Science Data Challenge #1

Zoom-in of the 1.4 GHz maps, showing the same region of the sky with different telescope integration: 8, 100, 1000 h left to right.

The SKA Science Data Challenge #1 (SDC1) release consists of 9 files, with the format of FITS images. Each file is a simulated SKA continuum image in total intensity of the same field at 3 frequencies (560 MHz, representative of SKA Mid Band 1, 1.4 GHz, representative of SKA Mid Band 2 and 9.2 GHz, representative of SKA Mid Band 5) and 3 telescope integrations (8, 100, 1000 h as representative of a single, medium-depth and deep integration, respectively).

Ancillary data consist of primary beams and synthesized beams for each frequency. An explanatory supplement describes the data and the challenge that is set for the community. A training set is also released, which consists in truth catalogues listing the objects in the simulated 1000 h data and their properties for a 5% of the field-of-view.

https://astronomers.skatelescope.org/

Challenge Description

The challenge set for the community is to undertake:

- source finding (RA, Dec) to locate the centroids and/or core positions,
- source property characterization (integrated flux density, possible core fraction, major and minor axis size, major axis position angle)
- source population identification (one of SFG, AGN-steep, AGN-flat)

The full descritpion of the data and of the challenge set is here:
SKA Data Challenge #1 description   DOWNLOAD

Take up the challenge!

Submission of results are accepted from either individuals or teams. If you would like to participate, please let us know by registering your interest, as detailed on Sec. 8 of the Challenge Description file, so that we can you keep you in the loop of challenge progress updates and other communications. Details on how to format and submit your results are also given in the document. The challenge starting day is 26/11/2018 and the deadline for submitting results is 15/3/2019, after which results will be graded.

Have a taste of what SKA data will be!

You are welcome to use the data for your own research, and to perform analyses and tests beyond the set challenge. Please acknowledge the use of these data as "SKAO data challenges, science data challenge #1". Details on how the dataset has been made and further references, including scientific papers, can be found in the Challenge Description file.

Data

560 MHz, 8 hours	4 Gb	DOWNLOAD
560 MHz, 100 hours	4 Gb	DOWNLOAD
560 MHz, 1000 hours	4 Gb	DOWNLOAD

Data

560 MHz, 8 hours	4 Gb	DOWNLOAD	
560 MHz, 100 hours	4 Gb	DOWNLOAD	B1
560 MHz, 1000 hours	4 Gb	DOWNLOAD	
1400 MHz, 8 hours	4 Gb	DOWNLOAD	
1400 MHz, 100 hours	4 Gb	DOWNLOAD	B2
1400 MHz, 1000 hours	4 Gb	DOWNLOAD	
9200 MHz, 8 hours	4 Gb	DOWNLOAD	
9200 MHz, 100 hours	4 Gb	DOWNLOAD	B5
9200 MHz, 1000 hours	4 Gb	DOWNLOAD	

- Short
- Medium
- Long

Ancillary data

560 MHz, primary beam	300 Kb	DOWNLOAD
560 MHz, synthesized	4 Gb	DOWNLOAD
1400 MHz, primary beam	300 Kb	DOWNLOAD
1400 MHz, synthesized	4 Gb	DOWNLOAD
9200 MHz, primary beam	300 Kb	DOWNLOAD
9200 MHz, synthesized	4 Gb	DOWNLOAD

Training set

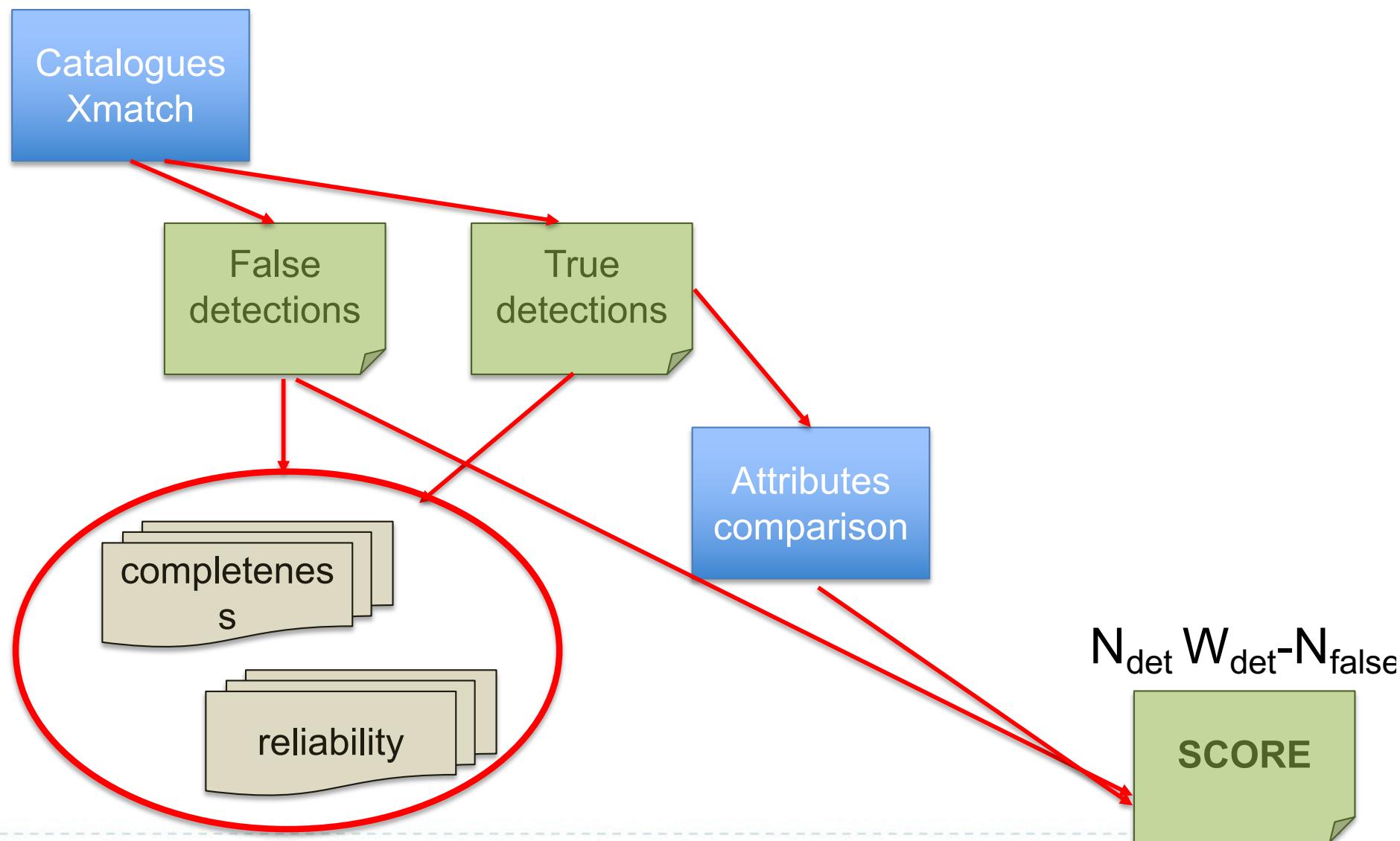|			
560 MHz, truth catalogue	54 Mb	DOWNLOAD	
1400 MHz, truth catalogue	14 Mb	DOWNLOAD	Truth table for a 5% sky area: training set
9200 MHz, truth catalogue	340 Kb	DOWNLOAD	

Explori

The SDC1 teams at work!

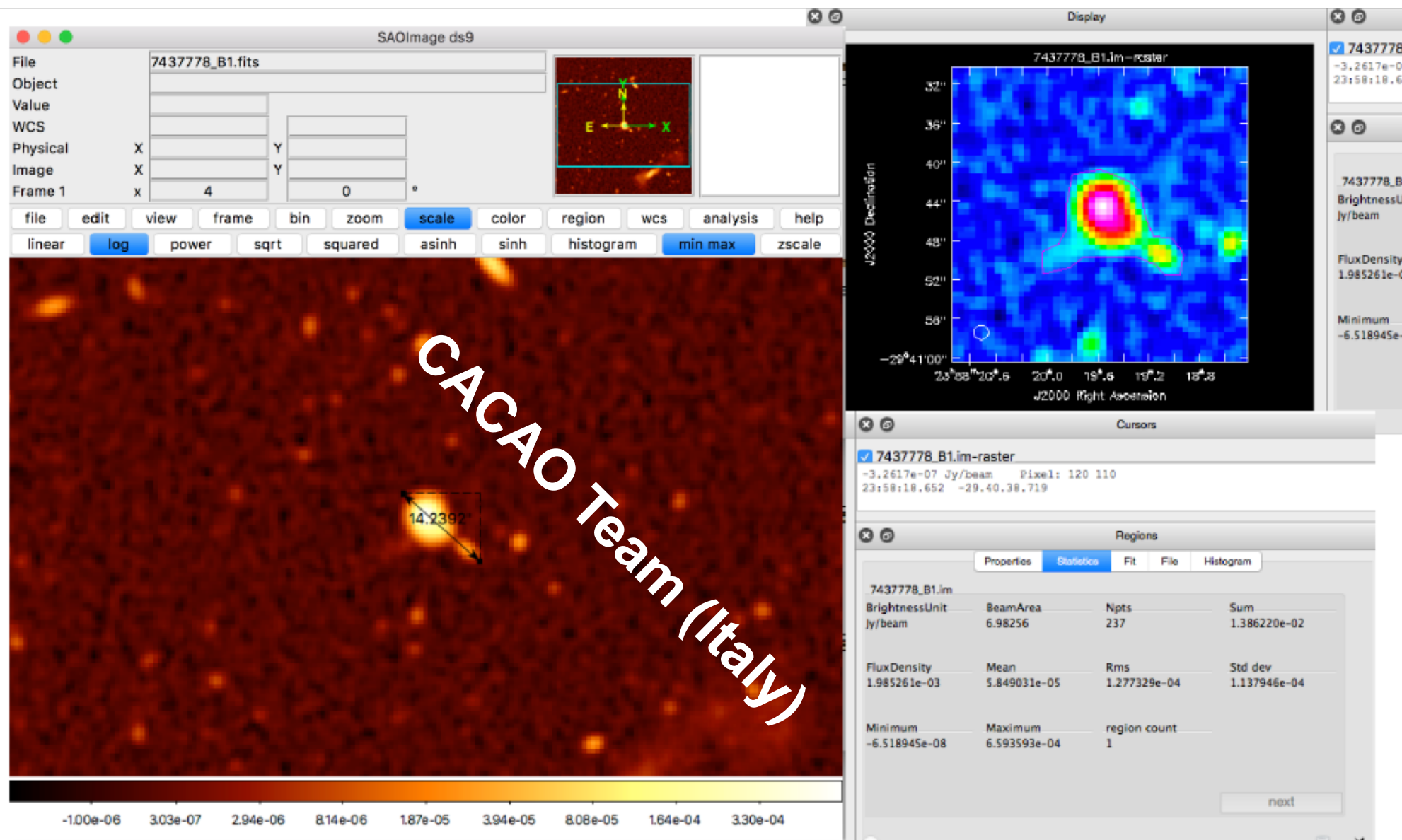13 teams, both inside and outside the scientific SKA community
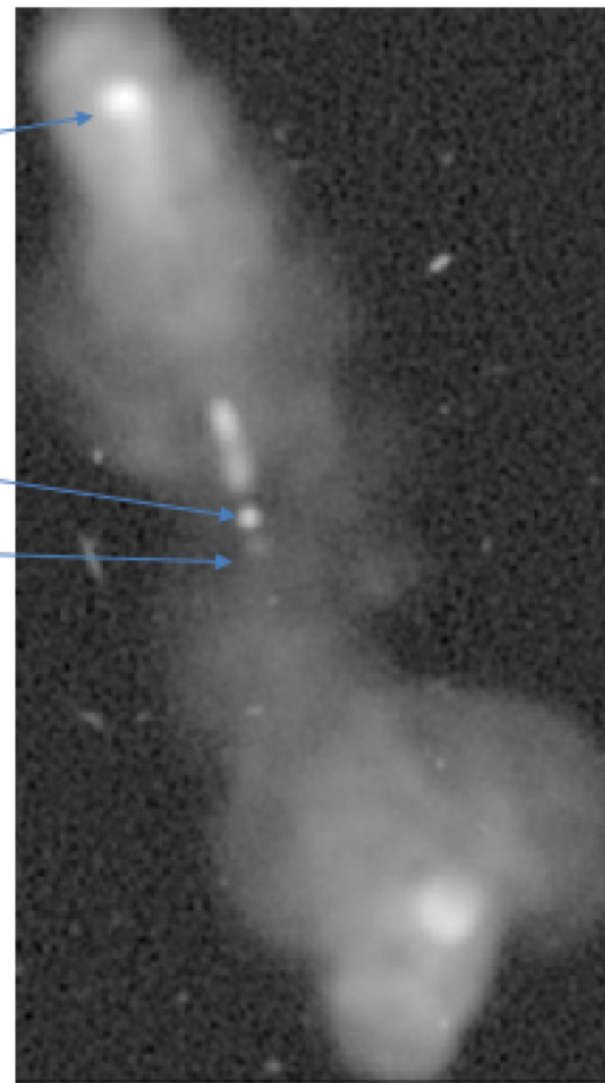
SDC1 scoring method

Catalogues Xmatch

False detections

True detections

Attributes comparison

completeness

reliability

$N_{det} W_{det} - N_{false}$

SCORE

WIP examples

CACAO Team (Italy)

WIP examples
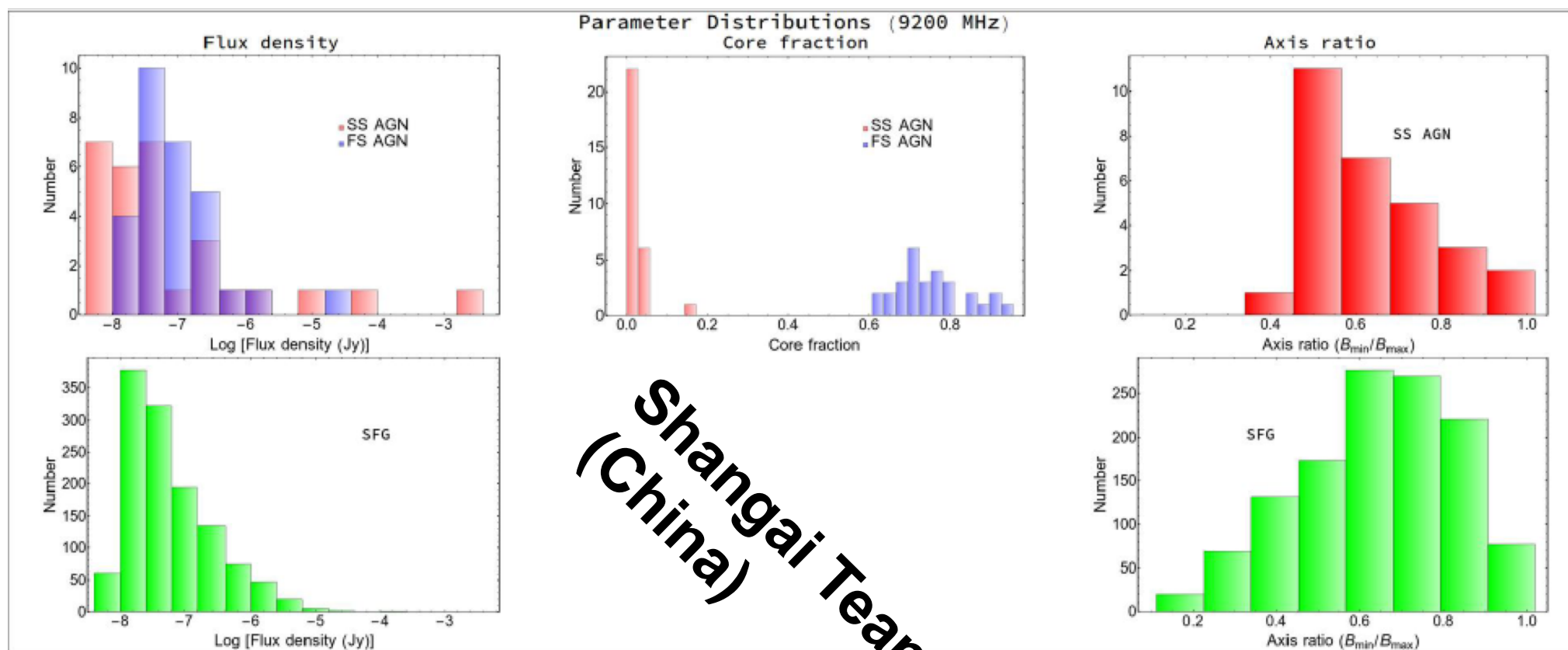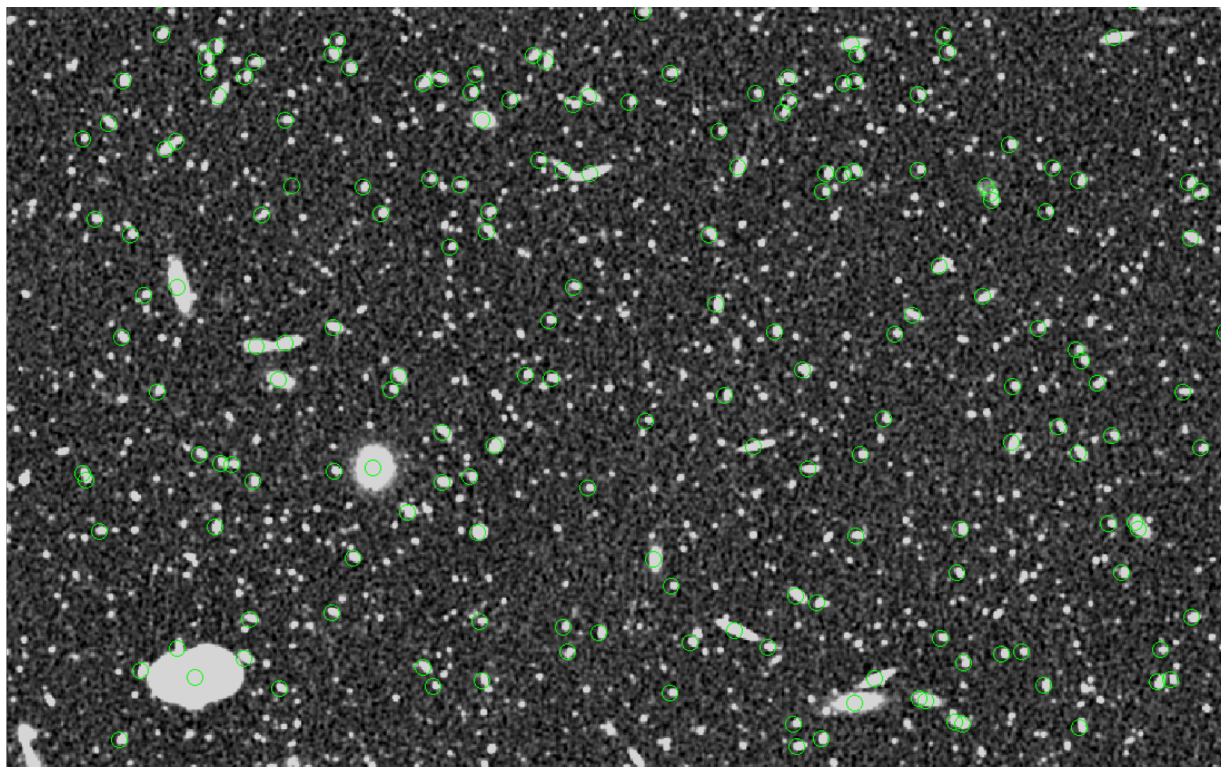
WIP examples

WIP examples

I. Toledo (Chile)
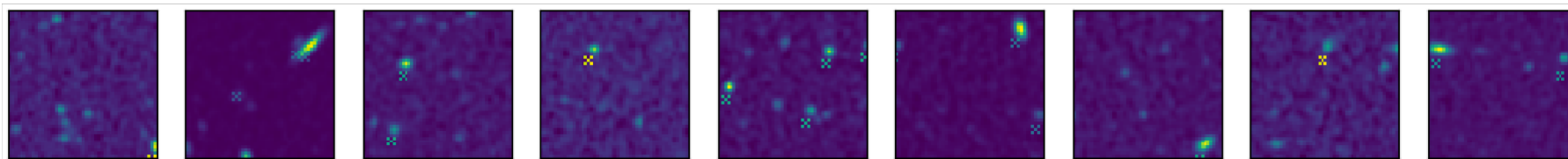
V. Lukic (Germany)

SDC1 timeline and progress

Report/
publication
on results

Challenge
winner

Final
results
submission

Update at
SKA
conference

Preliminary
results
submission

Telecon with
participants

Updates to
the dataset

Data released

Nov 2018   Dec 2018   Jan 2019   Feb 2019   March 2019   Apr 2019   May 2019
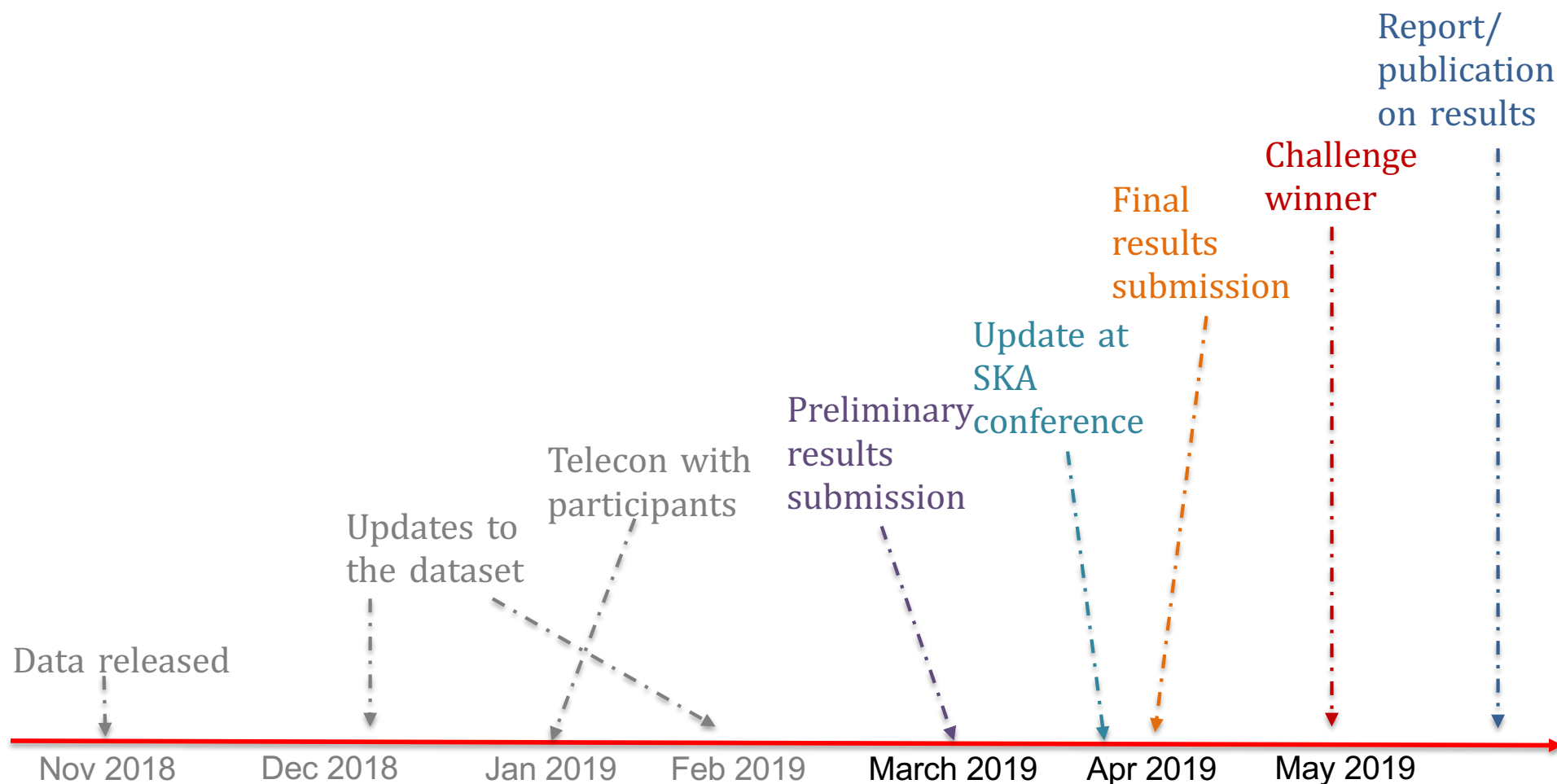
Ideas for the next SDCs

- IM simulation **Tianyue Chen, SFTC industry placement @ SKAO**

    - Wide area, including diffuse Galactic emission

    - Total power, "single dish" measurement

    - Investigate scan strategy

    - Investigate foreground removal methods

    Philippa Hartley, postdoc @ SKAO

- Represent the variety of SKA observing modes and science cases

- Increasing realism:
    - Time variability
    - Polarization
    - Instrumental systematics

- **Long-term goal: combine with the SDP and SRC challenges to "end to end" simulations**

Exploring the Universe with the world's largest radio telescope

…Stay tuned!