Real-time data processing for the BlackGEM Array of telescopes

Radboud Universiteit

RADBOUD RADIO LAB

NOVA

Steven Bloemen
Project Manager

2018

MeerLICHT

1 prototype telescope

SAAO Sutherland, S-Africa

2023

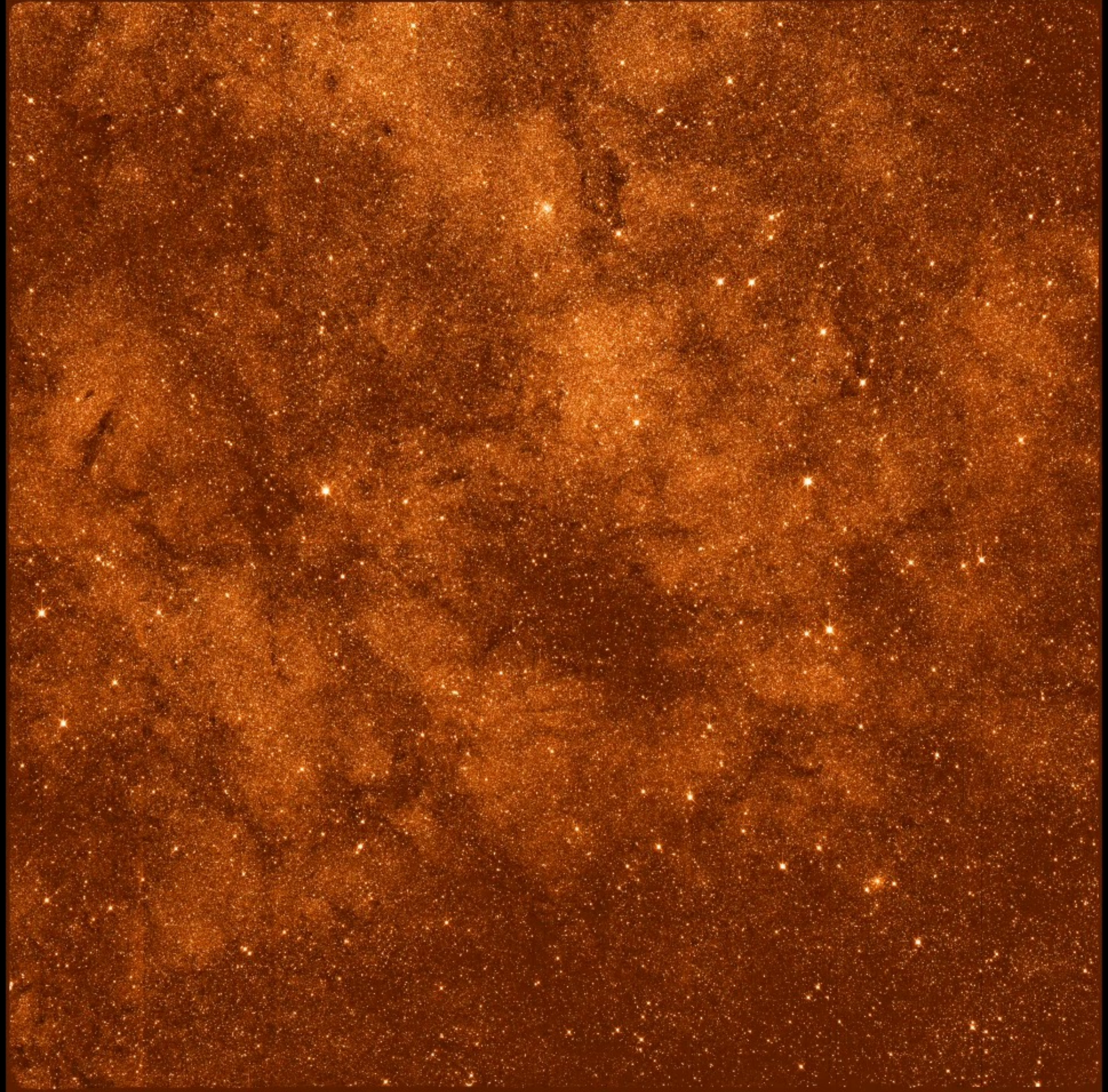
BlackGEM

3 telescopes

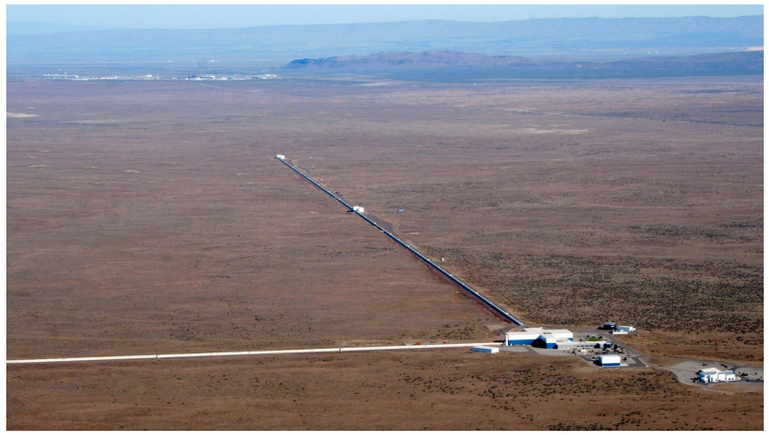
ESO La Silla, Chili

100 Mpix camera
1 image / minute
Sky area 10x full moon per image

BlackGEM inauguration by minister Dijkgraaf
9 January 2024

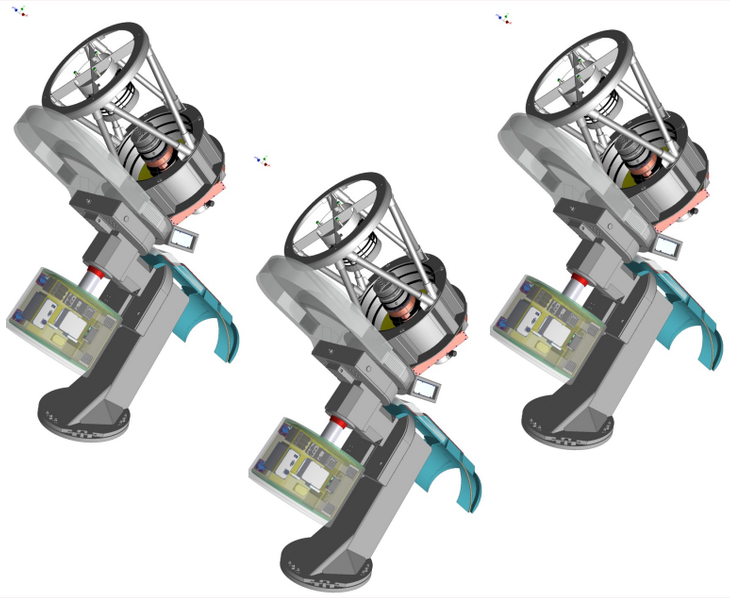


Main goal: optical counterparts of gravitational wave sources



Detection LIGO/Virgo

Rough sky position

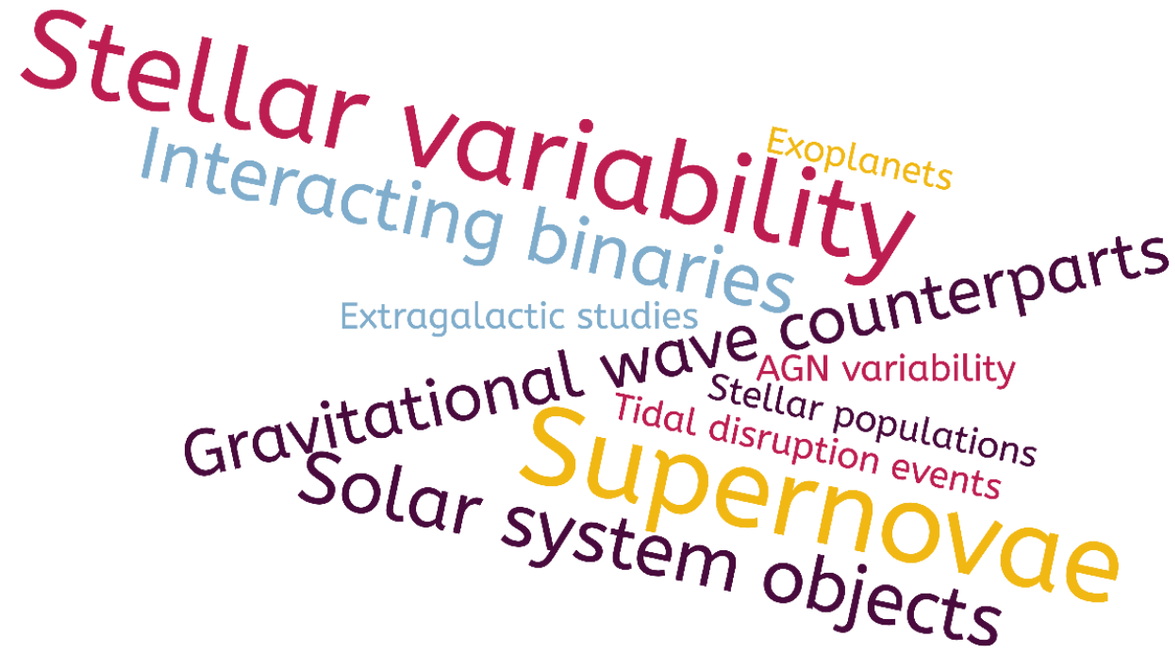


Find (faint!) source using wide-field telescopes

Follow-up with large facilities

Consortium & data rights

- All consortium members get access to all data products
- Science working group leaders are appointed in the consortium



WARWICK
THE UNIVERSITY OF WARWICK

UNIVERSITAT DE BARCELONA

LCO
Las Cumbres Observatory

NOVA

Radboud University

KU LEUVEN

ARMAGH OBSERVATORY & PLANETARIUM

MANCHESTER 1824
The University of Manchester

UCDAVIS
UNIVERSITY OF CALIFORNIA

TEL AVIV UNIVERSITY
אוניברסיטת תל אביב

TEXAS TECH UNIVERSITY

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

Universität Potsdam

DTU

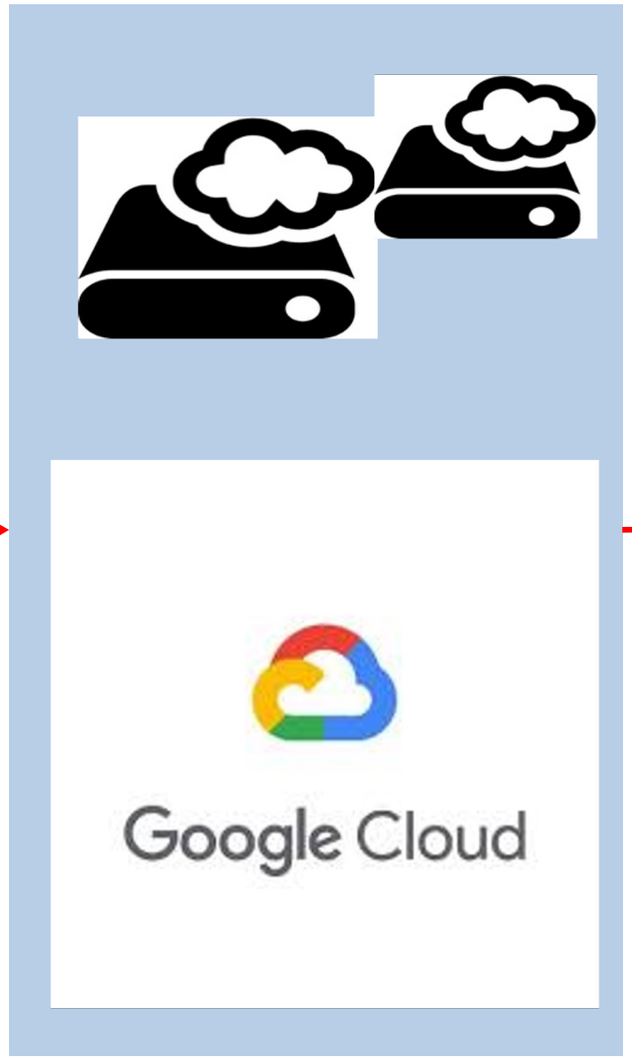
Data flow MeerLICHT

Raw data (600 images, 50 GB per night) transmitted to Cape Town via fiber

Processing, storage & databases at IDIA/Ilifu (**South African research cloud**)

From telescope to fully processed data in < 15 minutes

Data flow BlackGEM



Raw data (2000 images, 150 GB per night) transmitted via microwave link and Google fiber

Processing, storage & databases in **Google Cloud**

Image processing

Runs on slurm cluster in Ilifu (MeerLICHT) and Google Cloud (BlackGEM)

BlackBOX (Paul Vreeswijk) – https://github.com/pmvreeswijk/BlackBOX

- Standard astro-image processing steps (bias, flatfield, gain,…)
- Astrometric calibration based on Gaia DR3 (incl. proper motion)
- Photometric calibration based on Gaia DR3 (low-resolution spectra)

VARIABILITY PIPELINE

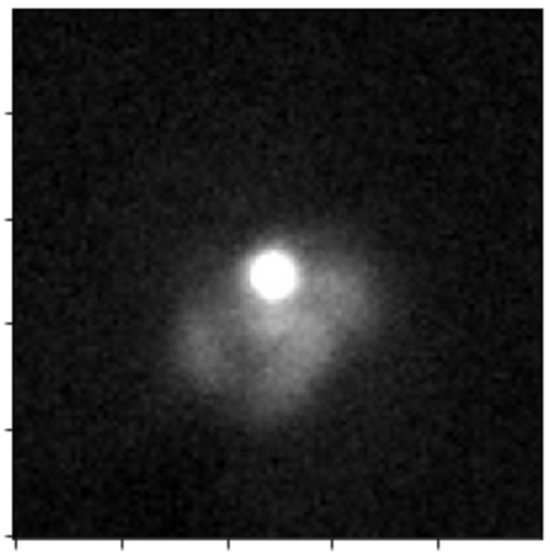
- Forced photometry for all Gaia DR3 sources
- ~10^8 sources per night

TRANSIENT PIPELINE

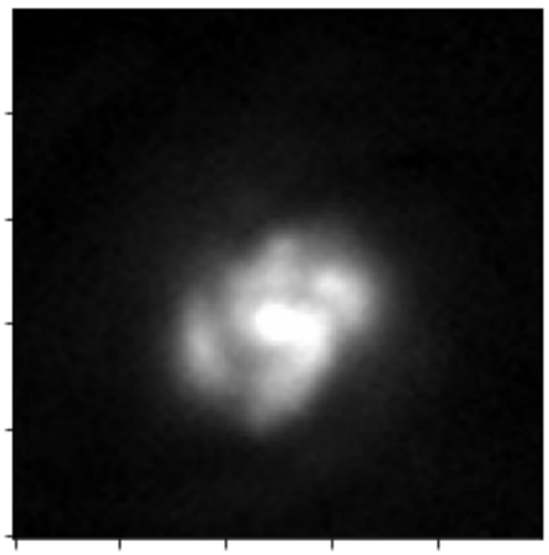
- Difference imaging (Zogy)
- SourceExtractor search for transients
- ~10 000 candidates per night
- Deep learning real-bogus classification

Example: real transient (supernova)

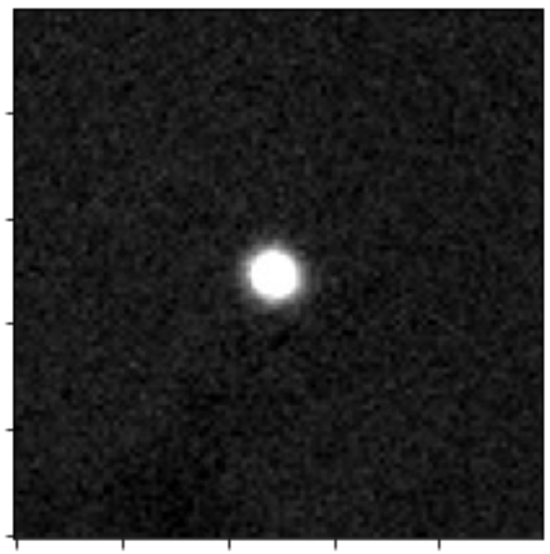
NEW IMAGE



'OLD' REFERENCE IMAGE

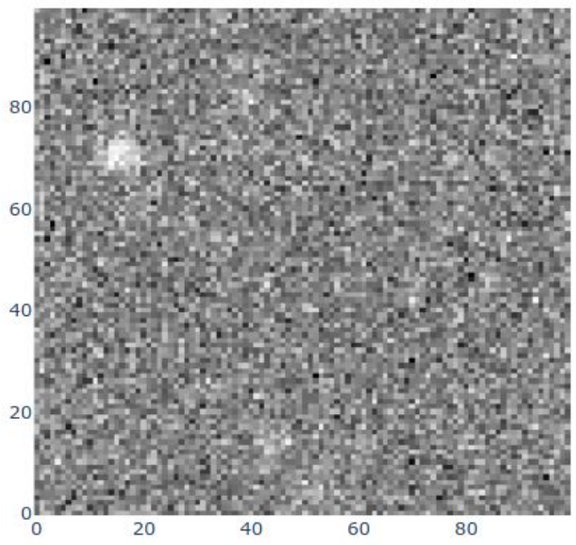


DIFFERENCE IMAGE

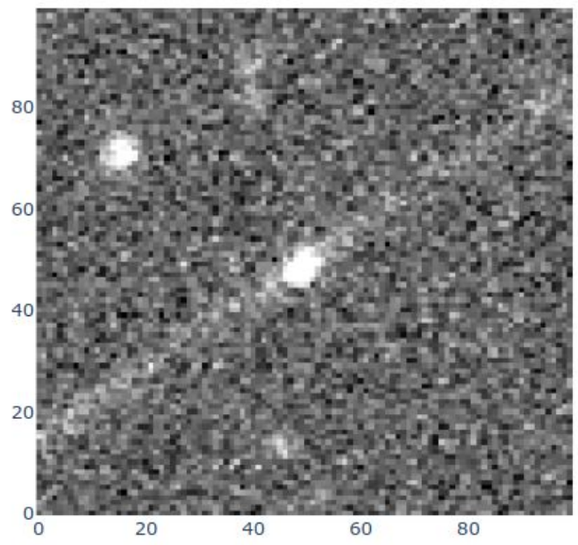


Example: artefact

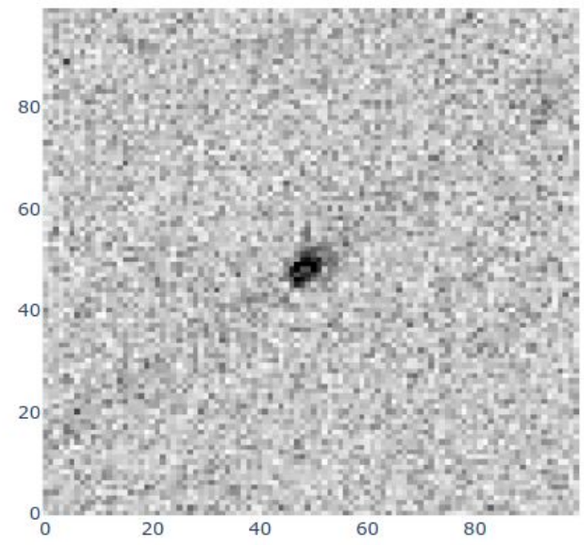
NEW IMAGE



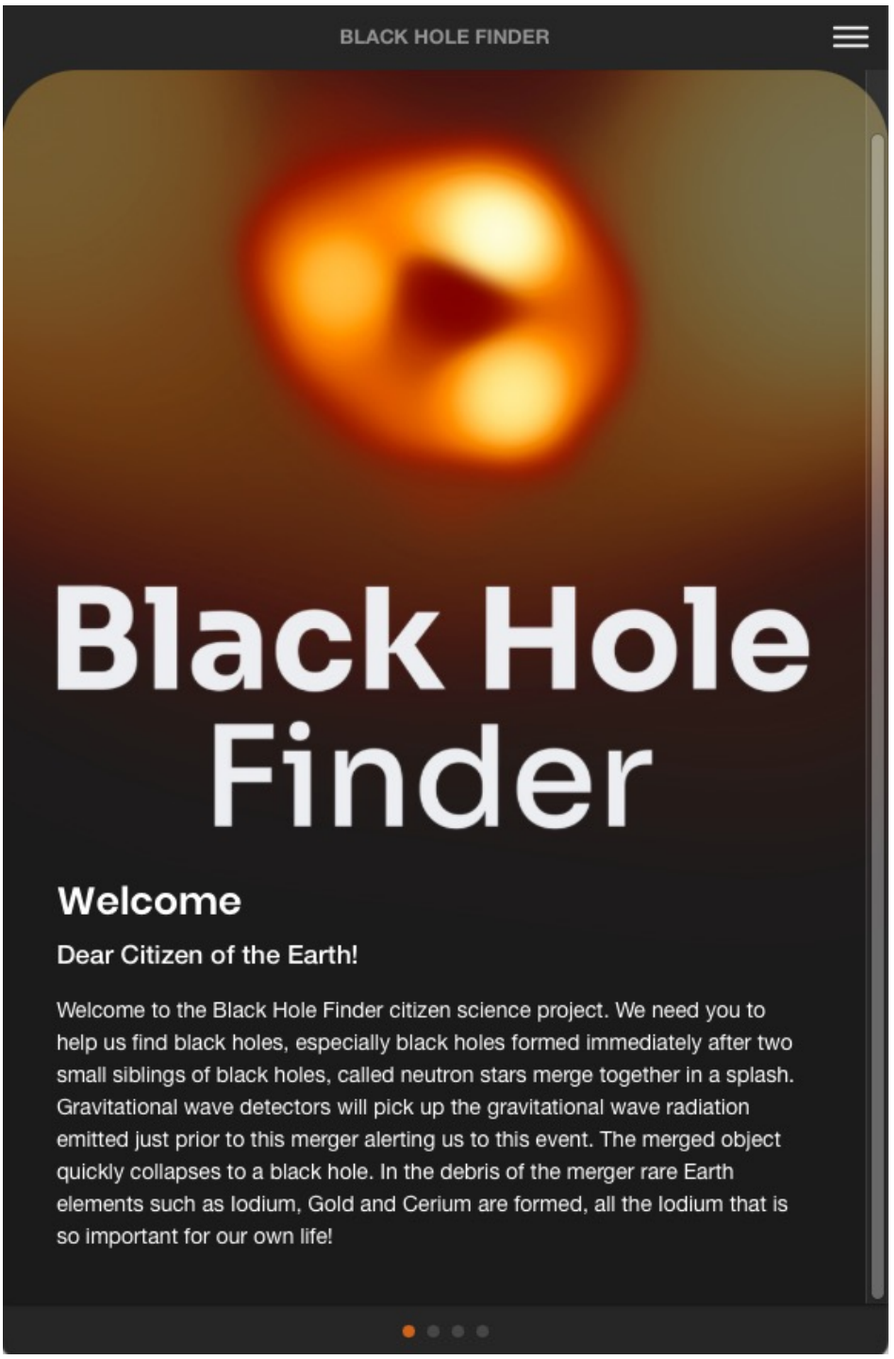
'OLD' REFERENCE IMAGE

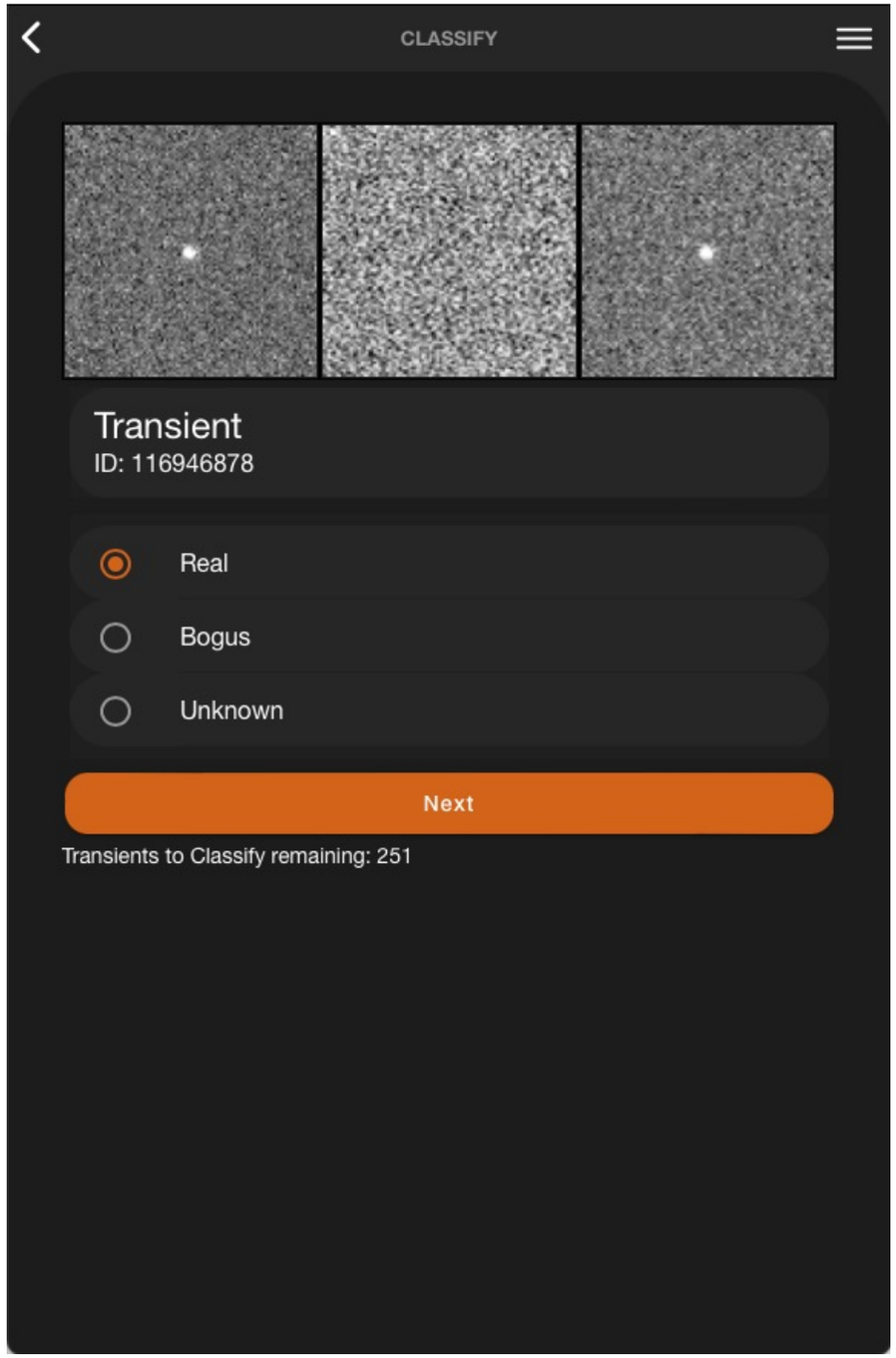


DIFFERENCE IMAGE

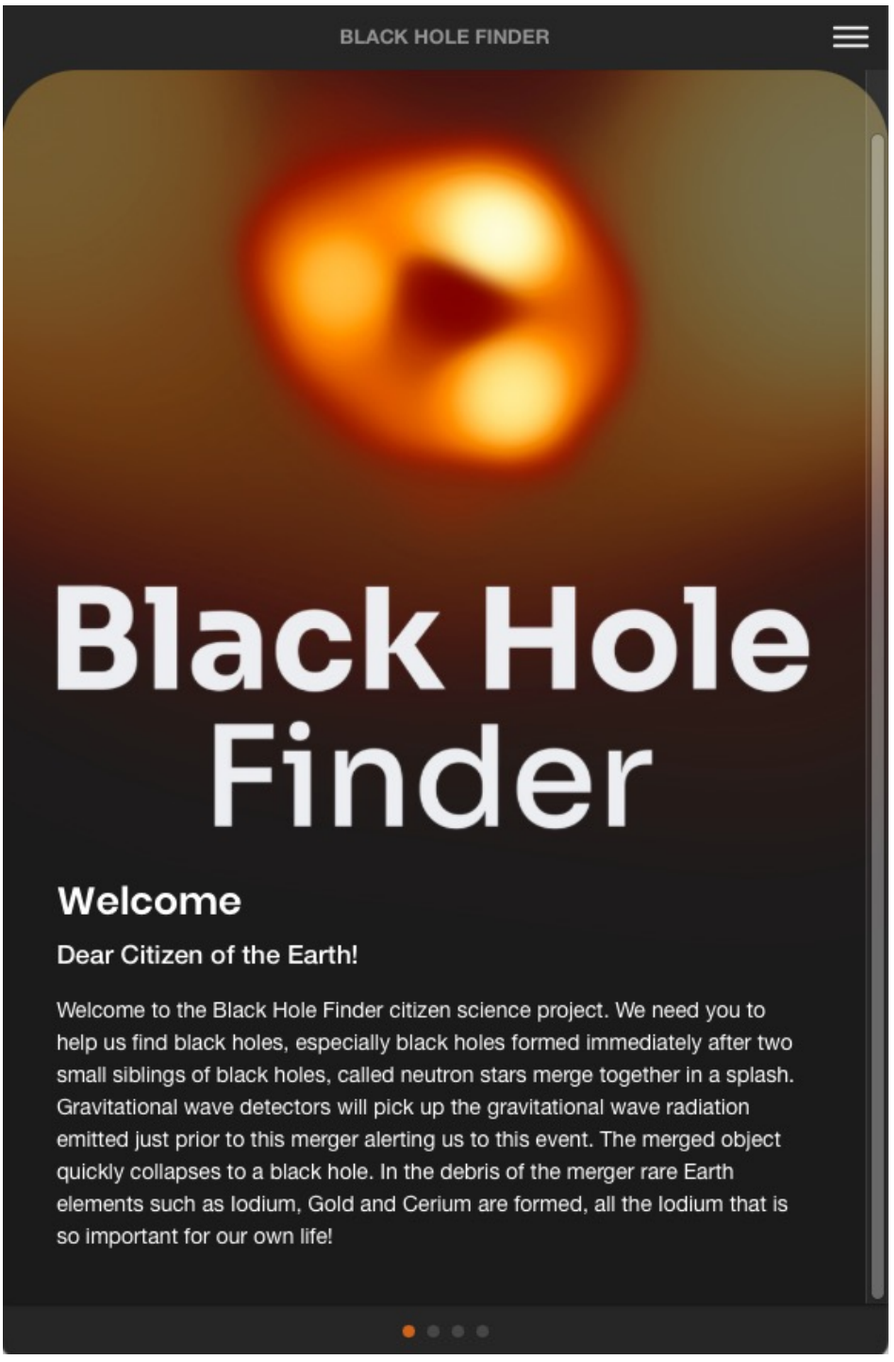


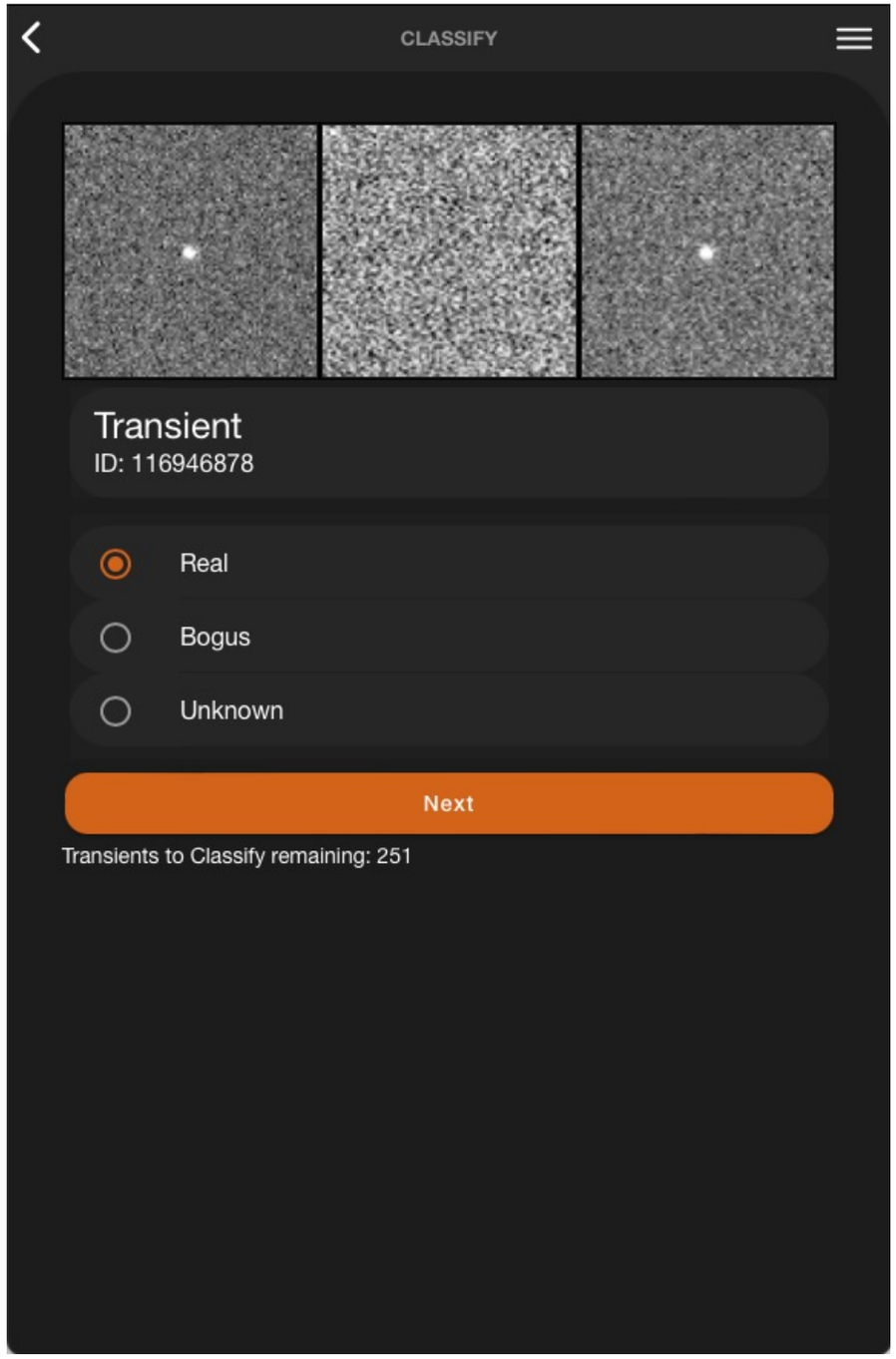
Citizen Science app

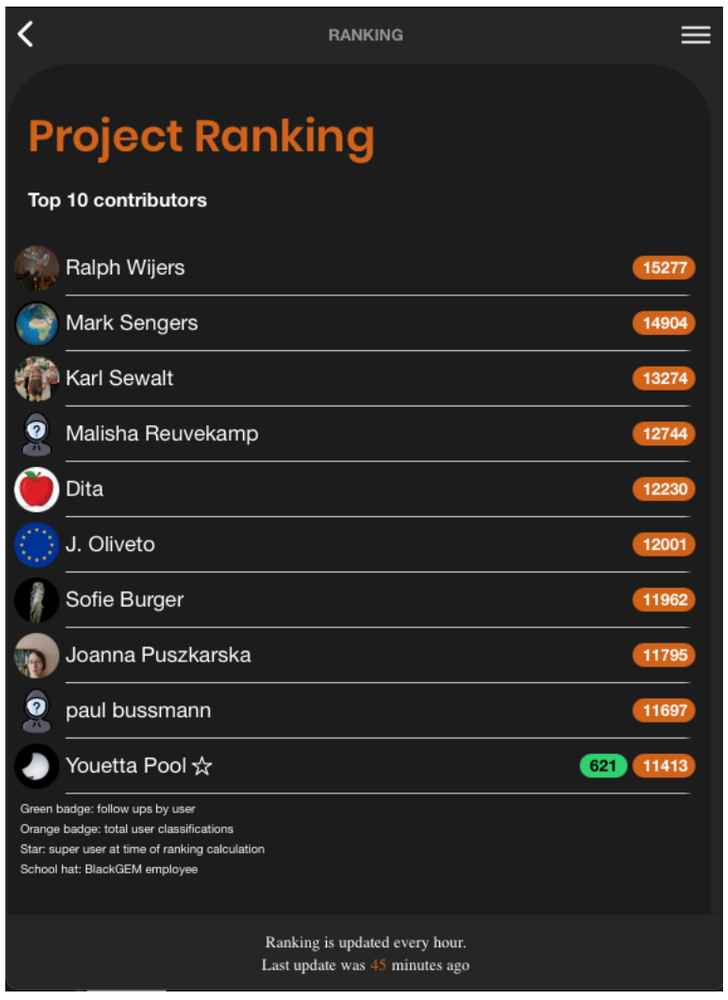




Citizen Science app

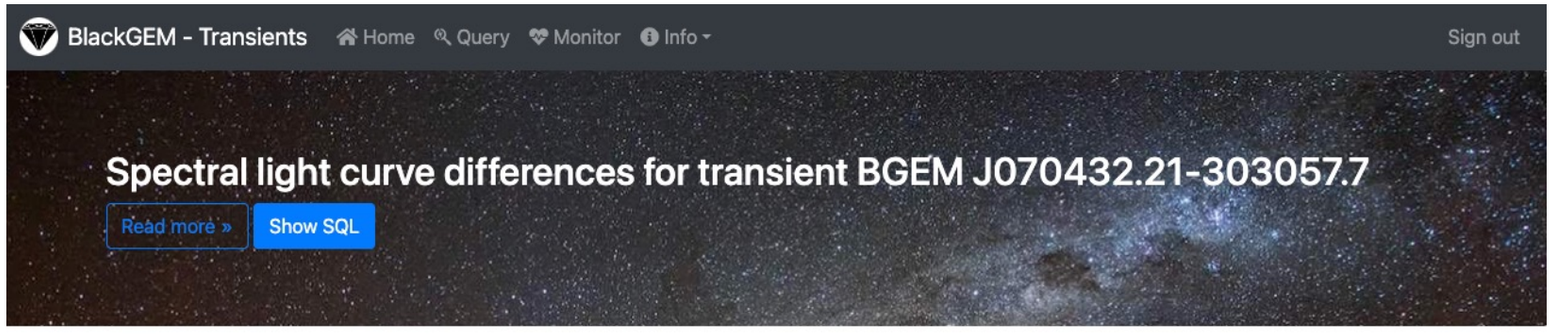


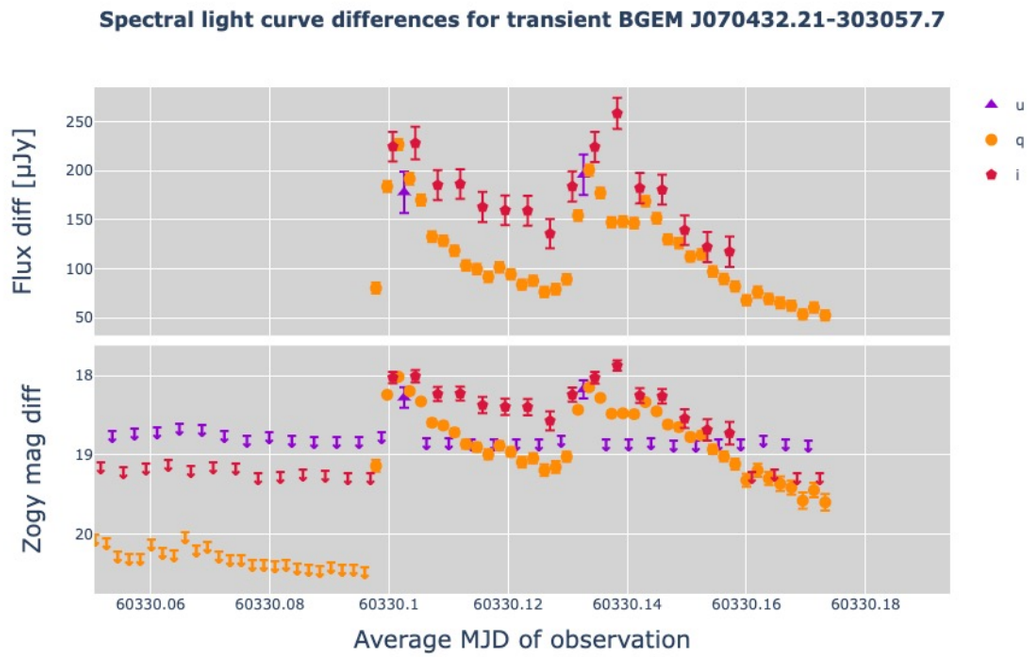




Dataspex database for transients

- Database developed by Dataspex in MonetDB
- Source matching at based on coordinates, on-the-fly updates of statistics
- Web and Jupyter notebook access





BigQuery database:

Forced photometry for all Gaia sources

Individual detections

(Variability) statistics for each source + filter combination

Image metadata

Source list = Gaia DR3 catalogue

Detection

Detection_ID	STRING
Image_ID	STRING
Source_ID	INTEGER
X_POS	FLOAT
Y_POS	FLOAT
FLAGS_MASK	INTEGER
BACKGROUND	FLOAT
MAG_APER_R0P66xFWHM	FLOAT
MAG_APER_R1P5xFWHM	FLOAT
MAG_APER_R5xFWHM	FLOAT
MAGERR_APER_R0P66xFWHM	FLOAT
MAGERR_APER_R1P5xFWHM	FLOAT
MAGERR_APER_R5xFWHM	FLOAT
MAG_OPT	FLOAT
MAGERR_OPT	FLOAT
SNR_OPT	FLOAT
LIMMAG_OPT	FLOAT
FNU_OPT	FLOAT
FNUERR_OPT	FLOAT

Image

Image ID	STRING
FILENAME	STRING
ORIGFILE	STRING
ALTITUDE	FLOAT
EPOCH	FLOAT
RA	FLOAT
RA_CNTR	FLOAT
DEC	FLOAT
DEC_CNTR	FLOAT
OBJECT	INTEGER
IMAGETYP	STRING
FILTER	STRING
EXPTIME	FLOAT
DATE_OBS	TIMESTAMP
MJD_OBS	FLOAT

Source_Stats

Source ID	INTEGER
FILTER	STRING
N	INTEGER
F_AVE	FLOAT
F_VAR	FLOAT
F_MED	FLOAT
F_SKEW	FLOAT
F_IQR	FLOAT
CHI2_RED_W	FLOAT
F_KUR	INTEGER
F_KUR_W	FLOAT
A	INTEGER
A_W	FLOAT

Source

Source_ID	INTEGER
RANDOM_INDEX	INTEGER
REF_EPOCH	FLOAT
RA	FLOAT
RA_ERROR	FLOAT
DEC	FLOAT
DEC_ERROR	FLOAT
PARALLAX	FLOAT
PM	FLOAT
PHOT_G_MEAN_MAG	FLOAT

BigQuery database
Forced photometry for all Gaia sources

- Optimized for retrieval of light curves (flux timeseries)
- Allows for large scale analytics

  e.g. statistics of light curves

  average and median flux, skewness, kurtosis,…

  4 billion datapoints
  250 million light curves (source+filter combinations)
  300 GB of data

  → 72 seconds (25h of CPU time)

- Cost model: 7 EUR/TB processed
- Fast rebuild (e.g. after reprocessing): 1 year of data in <1 day
- Currently being tested before roll-out to consortium

Using a commercial cloud environment

PROs

- Flexibility – scale up massively for short periods of time
- Fast international networks out-of-the-box
  e.g. 4 Gb/s Ilifu @ Cape Town → NL
- Well built and maintained ecosystem
  Managed services, (web) interfaces, (Python) access libraries, user management tools,…
- Cost effective for our 'lots of data, not a lot of users' scenario
- Very reliable
  High uptime, fast fault resolution

CONs

- Steep learning curve
- Need external consultants to make the most out of it
- Mistakes can be costly – need to monitor spending
- Needs a change of mindset compared to own hardware
  Hard to exhaust resources, easy to exhaust your bank account
- Certainly not suitable for every use case

www.blackgem.org @BlackGEM_Array