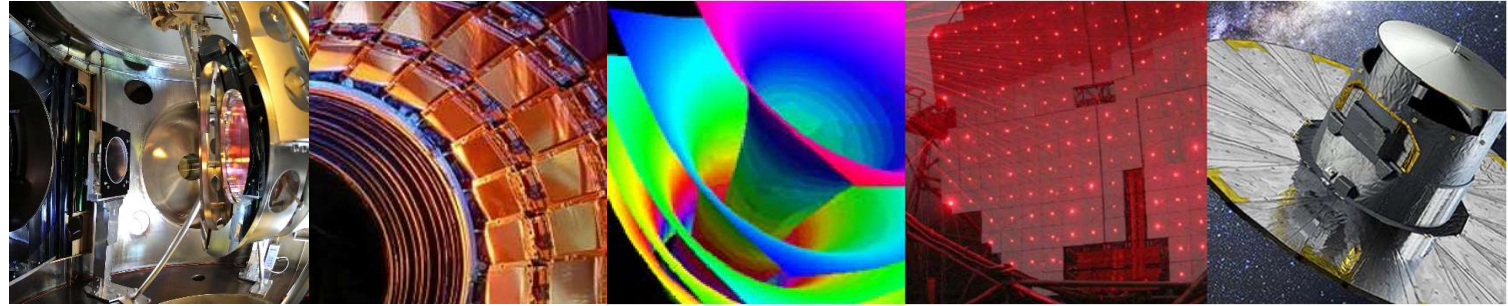




EXCELENCIA
MARÍA
DE MAEZTU
2020-2023

Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA



Developments at the ICCUB for the preparation of **Gaia** DR4 and its exploitation

Institute of Cosmos Sciences
Universitat de Barcelona

Xavier Luri Carrascoso
ICCUB director
Jordi Portell i de Mora
ICCUB deputy technical director

AstroHEP-PPCC24

Zaragoza
6 June 2024

IEEC
Institut d'Estudis
Espacials de Catalunya

Funded by:  **Unió Europea**
Fons Europeu
Next Generation



 **Plan de Recuperación,**
Transformación
y Resiliencia

 **Next Generation**
Catalunya

 **Generalitat**
de Catalunya

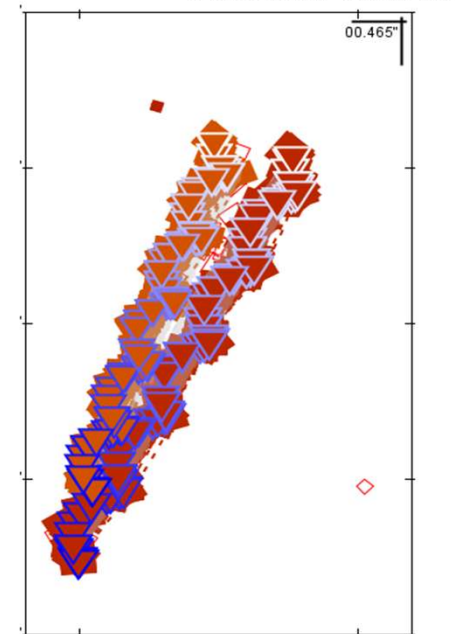
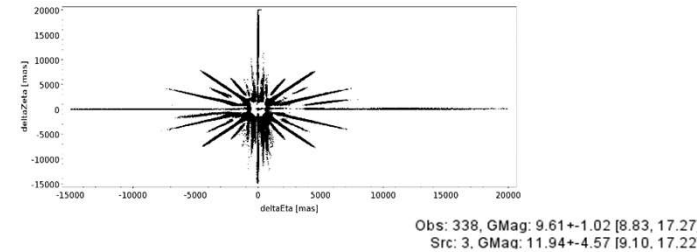
 **UNIVERSITAT DE**
BARCELONA

Data processing, validation and visualization



Some recent activities within the frame of the **Gaia DPAC (Data Processing and Analysis Consortium)**:

- CU3 (**Core Processing**), IDU (**Intermediate Data Updating**), Cross-Matching
 - Development and integration of algorithms: calibrations, image parameters, spurious detections...
 - Improvement of very bright stars astrometry
 - On-ground detection and resolution of **close star pairs**
 - Identification and modelling of **resolved binary stars**
--> improve catalogue resolution and completeness (**DR4-DR5**): clusters, binaries, dense areas...
- DPCB (**Data Processing Centre of Barcelona**)
 - **Operational runs** at BSC (**MareNostrum**): **5.5 years** of mission data (**DR4**), up to **154E9 observations** processed, more than **280 TB** generated...
 - Now processing **~9.5 years** of mission data (already working towards **DR5!**)
 - **Official backup** of the full MainDB and raw TM Archive
 - Migration to MareNostrum 5



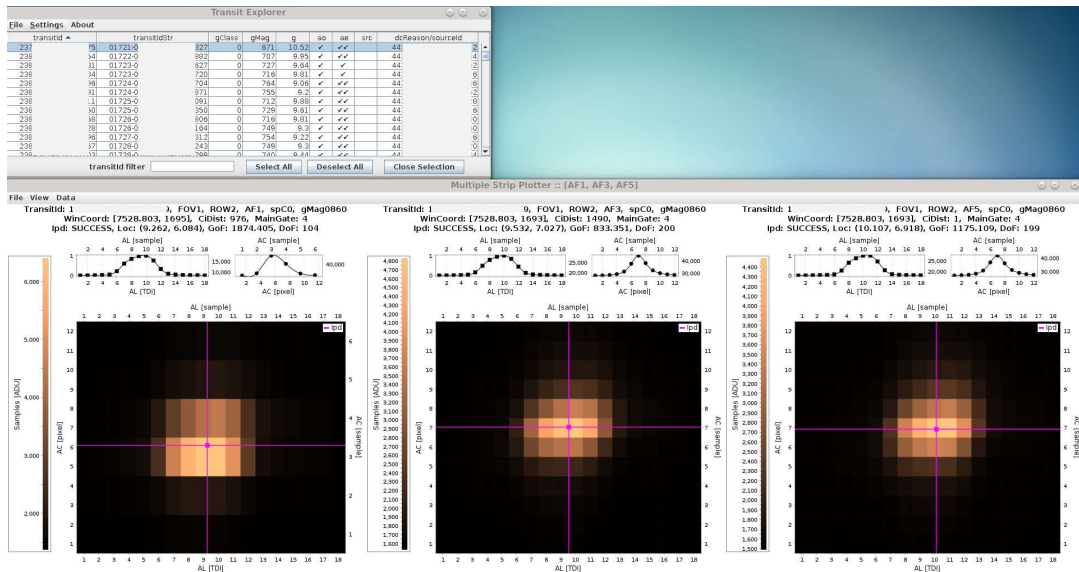
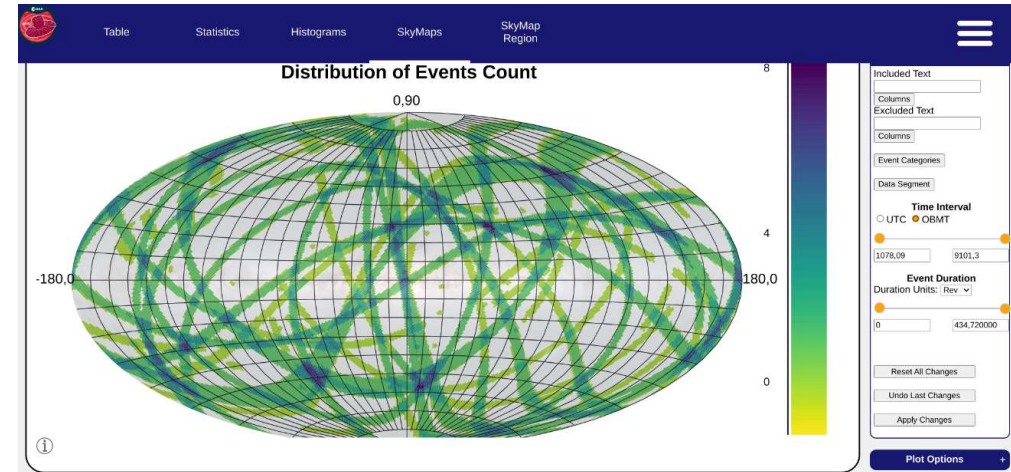
Data processing, validation and visualization



Some recent activities within the frame of the **Gaia DPAC (Data Processing and Analysis Consortium)**:

- Still in DPCB, **Data visualization tools**
 - **Catalogue Explorer**, to visualize the “scene” (observations and their match to sources) and run cross-matching tests
 - **Transit Explorer**, to visualize the observations
 - **Event Explorer**, to examine the spacecraft and mission events

| Event | System | Wavelength Start | Wavelength End | OBMT Start/End | OBMT Start/End | Description | Event length (min) | Event length (max) | Revs. status |
|---|------------|------------------|------------------|----------------|----------------|---|--------------------|--------------------|--------------|
| 7 ACDC MAINOP coverage | Spacecraft | 06/08/2014 09:02 | 06/08/2014 09:13 | 1125 140 | 1125 170 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 39.75 |
| 8 ACDC MAINOP coverage | Spacecraft | 12/08/2014 09:00 | 12/08/2014 09:11 | 1149 630 | 1149 660 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 24.46 |
| 9 ACDC MAINOP coverage | Spacecraft | 13/08/2014 04:18 | 13/08/2014 04:29 | 1150 360 | 1150 390 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 6.60 |
| 10 APR 11 switch-off (PSM) | SWM | 17/08/2014 18:38 | 17/08/2014 18:38 | 1171 800 | 1171 840 | SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 16.83 |
| 11 Solar Flaring Measurement | Spacecraft | 21/08/2014 04:10 | 21/08/2014 04:19 | 1185 180 | 1185 240 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 6.07 |
| 12 ACDC MAINOP coverage | Spacecraft | 21/08/2014 04:11 | 21/08/2014 04:24 | 1187 660 | 1187 700 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 2.31 |
| 13 SCM 20.4.4 MPS Offset C | Spacecraft | 22/08/2014 09:10 | 22/08/2014 09:19 | 1189 180 | 1189 200 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 1.46 |
| 14 ACDC MAINOP coverage | Spacecraft | 22/08/2014 09:27 | 22/08/2014 09:37 | 1190 960 | 1190 990 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 0.89 |
| 15 12.4 of antenna A1 EPSR | Spacecraft | 22/08/2014 11:38 | 22/08/2014 11:38 | 1190 630 | 1190 630 | From Gaia/ST/Thruster SC | 2:00:00.00 | 2:00:00.00 | 1.51 |
| 16 Close out of Epoch | CAR | 22/08/2014 12:00 | 22/08/2014 12:00 | 1192 120 | 1192 120 | | 0:00:00.00 | 0:00:00.00 | 0.66 |
| 17 - Transition from LPSM to PSM | Spacecraft | 22/08/2014 22:58 | 22/08/2014 22:58 | 1192 630 | 1192 630 | From Gaia/ST/Thruster PO | 0:00:00.00 | 0:00:00.00 | 0.49 |
| 18 ACDC MAINOP coverage | Spacecraft | 22/08/2014 18:11 | 22/08/2014 18:11 | 1195 750 | 1195 770 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 1.22 |
| 19 VPLM autonomous observation | VPLM | 22/08/2014 22:20 | 22/08/2014 00:39 | 1220 520 | 1220 740 | VPLM segment in Science, HI | 0:20:00.00 | 0:20:00.00 | 24.75 |
| 20 ACDC MAINOP coverage | Spacecraft | 30/08/2014 11:02 | 30/08/2014 11:02 | 1222 920 | 1222 940 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 2.17 |
| 21 RINGOUT target | VPLM | 30/08/2014 11:00 | 30/08/2014 11:00 | 1224 130 | 1224 140 | VPLM observation in Science | 0:00:00.00 | 0:00:00.00 | 1.38 |
| 22 Calibration problem: Bad fit: RPTI DR: RPTI FL | Spacecraft | 09/09/2014 09:36 | 09/09/2014 09:36 | 1233 730 | 1233 760 | RPTI observation in Science | 42:20:00.00 | 42:20:00.00 | 9.69 |
| 23 RINGOUT target | DR | 09/09/2014 09:35 | 09/09/2014 09:35 | 1245 840 | 1245 840 | DR observation in Science | 0:00:00.00 | 0:00:00.00 | 6.61 |
| 24 RINGOUT target: Bad fit: RPTI DR: RPTI FL | Spacecraft | 10/09/2014 09:31 | 10/09/2014 09:31 | 1245 830 | 1245 830 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 1.00 |
| 25 Saturn Transit (opening) Scienc | Science | 09/09/2014 18:28 | 09/09/2014 18:28 | 1245 960 | 1245 960 | From FOC, RINGOUT, 220.17N | 0:20:00.00 | 0:20:00.00 | 34.4 |
| 26 Mars Transit (opening) Science | Science | 09/09/2014 07:14 | 09/09/2014 07:14 | 1245 940 | 1245 940 | From FOC, RINGOUT, 230.37E | 0:00:00.00 | 0:00:00.00 | 35.29 |
| 27 WFTS 2 observation: Bad fit: SWM | Spacecraft | 09/09/2014 09:00 | 09/09/2014 09:00 | 1261 120 | 1261 240 | | 0:10:00.00 | 0:10:00.00 | 0.12 |
| 28 Transition MA to TM: SWM | Spacecraft | 09/09/2014 04:22 | 09/09/2014 06:24 | 1261 360 | 1261 390 | From VPLM, VPLM to ZODIAC/EAR | 0:20:00.00 | 0:20:00.00 | 0.12 |
| 29 Solar Flare and Coronal Mass | Spacecraft | 11/09/2014 01:01 | 11/09/2014 01:01 | 1268 80 | 1270 130 | Sci MA-A-05-1 in Solar HI | 7:30:00.00 | 7:30:00.00 | 7.27 |
| 30 VPLM MA Tracking (Solar) VPLM | Spacecraft | 11/09/2014 01:01 | 11/09/2014 01:01 | 1268 800 | 1268 800 | From Gaia/ST/Thruster SC | 0:00:00.00 | 0:00:00.00 | 1.64 |
| 31 WING CDB | DOC DT | 12/09/2014 09:00 | 12/09/2014 17:00 | 1284 600 | 1284 600 | WING CDB, opening: 240.0 | 0:00:00.00 | 0:00:00.00 | 14.63 |
| 32 Detector threshold update: VPLM | Spacecraft | 13/09/2014 09:00 | 13/09/2014 09:00 | 1284 600 | 1284 740 | From 200 sec. observation threshold and | 0:00:00.00 | 0:00:00.00 | 0.02 |
| 33 4RT confirmation: SWM | VPLM | 15/09/2014 09:00 | 15/09/2014 09:00 | 1284 740 | 1284 740 | From 200 sec. observation | 0:00:00.00 | 0:00:00.00 | 0.11 |
| 34 VPLM autonomous observation | VPLM | 15/09/2014 09:00 | 15/09/2014 09:00 | 1284 600 | 1284 600 | Short SERVICE modes for V | 0:00:00.00 | 0:00:00.00 | 0.75 |
| 35 SWM test start | SWM | 16/09/2014 18:24 | 16/09/2014 18:24 | 1282 300 | 1282 300 | From J05. Sci. Probably not | 0:00:00.00 | 0:00:00.00 | 2.66 |
| 36 WFTS sweep | VPLM VPLM | 16/09/2014 09:00 | 16/09/2014 09:00 | 1282 600 | 1282 640 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 1.53 |
| 37 WFTS sweep | VPLM VPLM | 17/09/2014 09:00 | 17/09/2014 09:00 | 1282 600 | 1282 640 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 0.83 |
| 38 WFTS sweep | VPLM VPLM | 17/09/2014 18:01 | 17/09/2014 18:01 | 1282 300 | 1282 300 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 2.04 |
| 39 WFTS sweep | VPLM VPLM | 18/09/2014 09:00 | 18/09/2014 09:00 | 1282 600 | 1282 640 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 1.53 |
| 40 WFTS sweep | VPLM VPLM | 18/09/2014 00:01 | 18/09/2014 00:01 | 1282 600 | 1282 640 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 0.10 |
| 41 DRPTI DR: RPTI FL | Spacecraft | 19/09/2014 18:24 | 19/09/2014 18:24 | 1282 300 | 1282 300 | From J05. Sci. Probably not | 0:00:00.00 | 0:00:00.00 | 1.75 |
| 42 DRPTI DR: RPTI FL | Spacecraft | 19/09/2014 18:22 | 19/09/2014 18:22 | 1282 300 | 1282 300 | From J05. Sci. Probably not | 0:00:00.00 | 0:00:00.00 | 0.14 |
| 43 WFTS sweep | VPLM VPLM | 19/09/2014 09:00 | 19/09/2014 09:00 | 1282 300 | 1282 300 | Short SERVICE modes for V | 0:00:00.00 | 0:00:00.00 | 0.89 |
| 44 WFTS sweep | VPLM VPLM | 20/09/2014 09:00 | 20/09/2014 09:00 | 1282 600 | 1282 640 | Short SERVICE modes for V | 0:00:00.00 | 0:00:00.00 | 0.39 |
| 45 WFTS sweep | VPLM VPLM | 20/09/2014 04:36 | 20/09/2014 04:36 | 1282 300 | 1282 340 | Short SERVICE modes for V | 0:00:00.00 | 0:00:00.00 | 0.75 |
| 46 WFTS sweep | VPLM VPLM | 21/09/2014 09:00 | 21/09/2014 09:00 | 1282 600 | 1282 640 | Short SERVICE modes for V | 0:00:00.00 | 0:00:00.00 | 0.34 |
| 47 SP-1 calibration: new VPLM | Spacecraft | 22/09/2014 00:22 | 22/09/2014 00:22 | 1282 600 | 1282 600 | Admission: new fields: k | 0:00:00.00 | 0:00:00.00 | 3.88 |
| 48 Time correlation (not done): MOI DOC | Spacecraft | 22/09/2014 07:00 | 22/09/2014 08:22 | 1282 600 | 1282 600 | From SC EAR, not in Thrust | 0:10:00.00 | 0:10:00.00 | 1.21 |
| 49 ACDC MAINOP coverage | Spacecraft | 22/09/2014 12:00 | 22/09/2014 12:00 | 1312 400 | 1312 400 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 2.47 |
| 50 Close out of Epoch | CAR | 23/09/2014 01:13 | 23/09/2014 01:13 | 1317 000 | 1317 000 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 0.81 |
| 51 ACDC MAINOP coverage | Spacecraft | 23/09/2014 01:13 | 23/09/2014 01:13 | 1317 000 | 1317 000 | From SC EAR, not in Thrust | 0:00:00.00 | 0:00:00.00 | 0.81 |



Data processing, validation and visualization



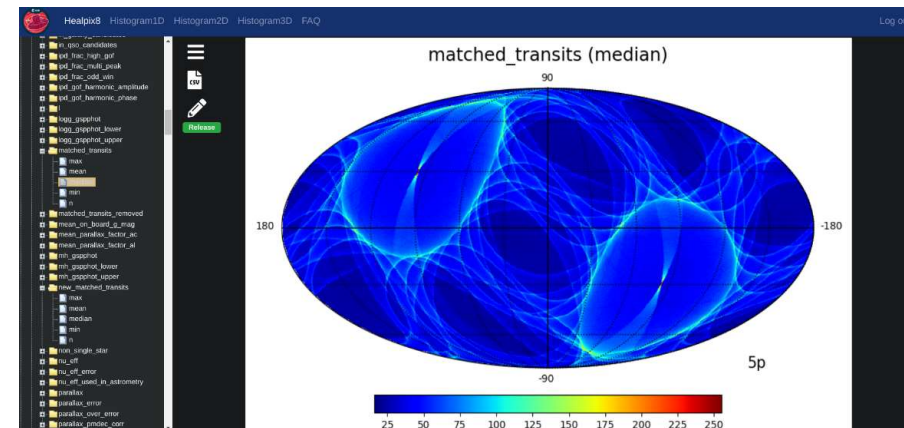
Some recent activities within the frame of the **Gaia DPAC (Data Processing and Analysis Consortium)**:

- CU9 (**Catalogue Preparation**)
 - Catalogue validation for **DR3** and the **FPR**: many new data types, tables, parameters...
 - Development of software tools for **statistics and validation**, e.g. the **Gaia Analysis Tool (GAT)**
 - **Now working hard on the many DR4 products**



- **Project Office**
 - **Technical interfaces** between Units and Centres; **technical support** to other Units
 - Estimation of database and transfer sizes
 - Curation of Operational Event Logs, support to visualization tools
 - Support to additional (often cross-unit) investigations

- CU3 / IDT (**Initial Data Treatment**)
 - Support to daily operations, monitoring and resolution of onboard/onground issues



Recent achievements and outlook



- **Data Release 3 (DR3):**

- Released **13 June 2022**
- Lots of new data products

- **Focused Product Release (FPR):**

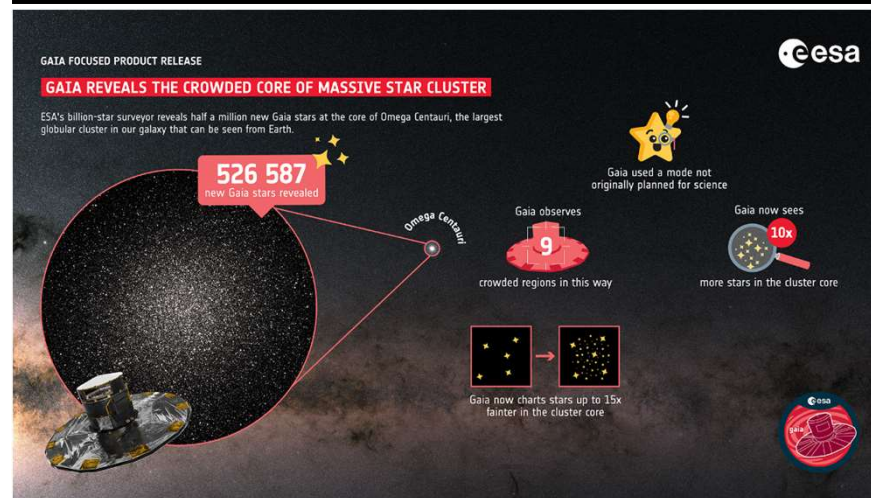
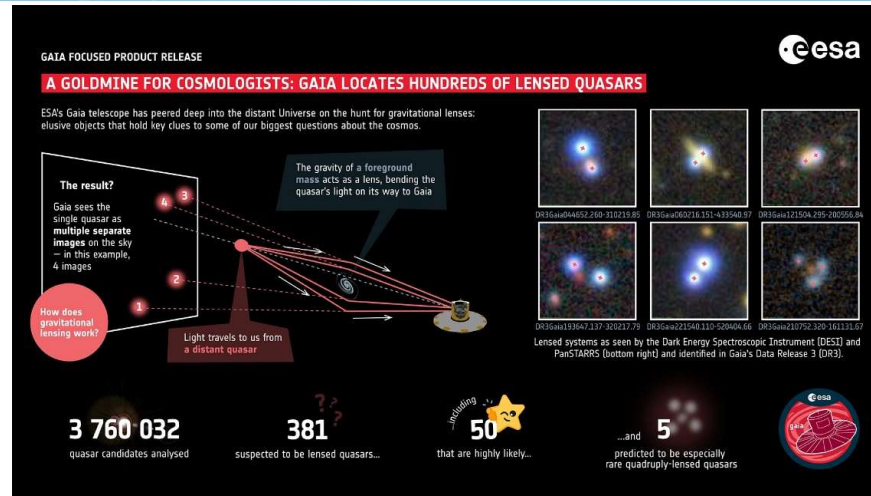
- Released **10 October 2023**
- Additional OmegaCen sources, Gravitational Lenses, improved SSO astrometry, LPVs, DIBs

- **Data Release 4 (DR4):**

- Full nominal mission (66 months)
- During **2026**
- **Epoch data** for all data products and sources (incl. astrometry, spectra, etc.)

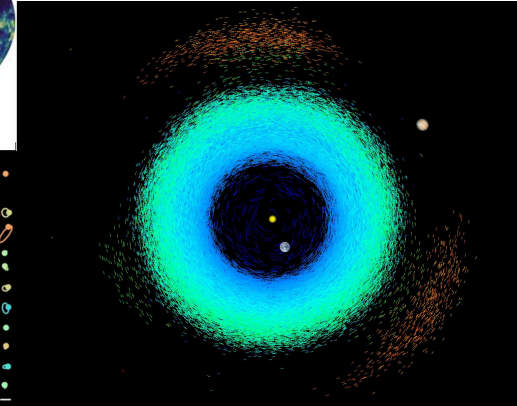
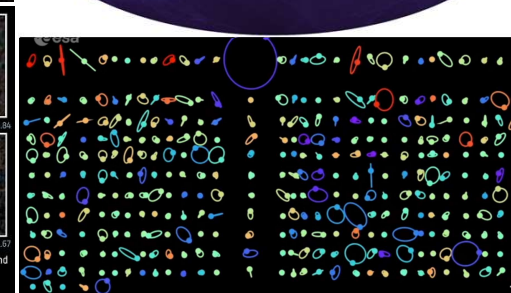
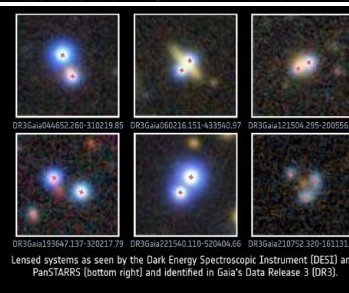
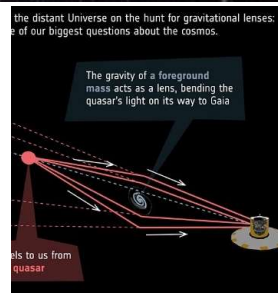
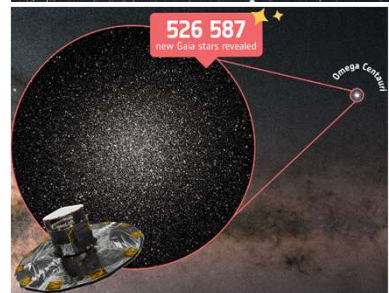
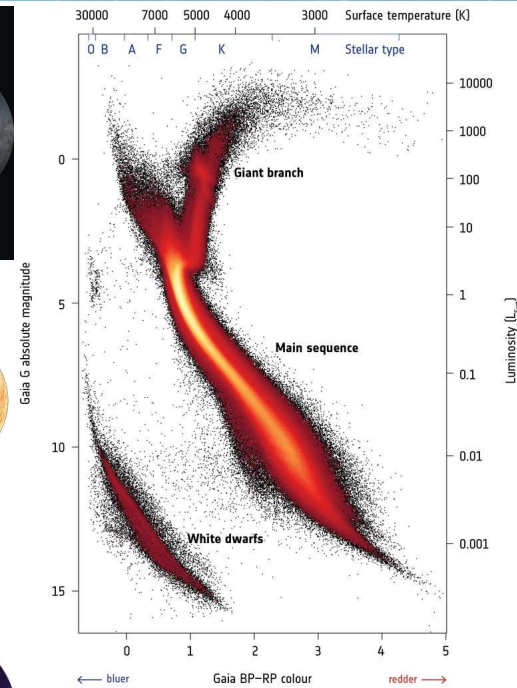
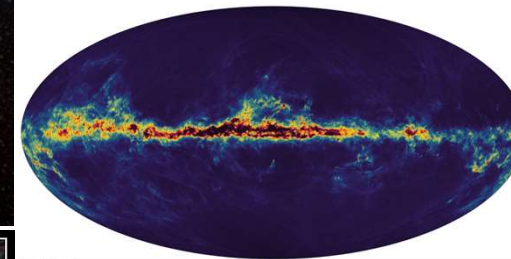
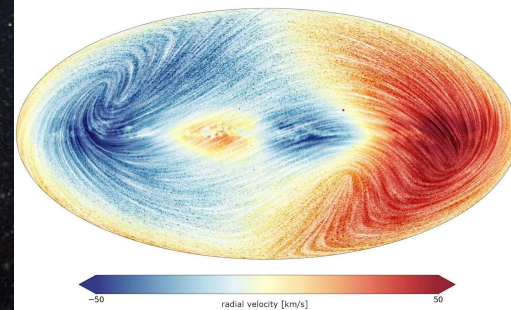
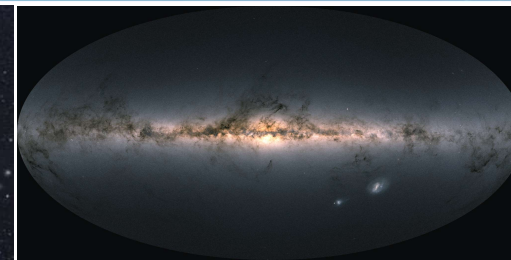
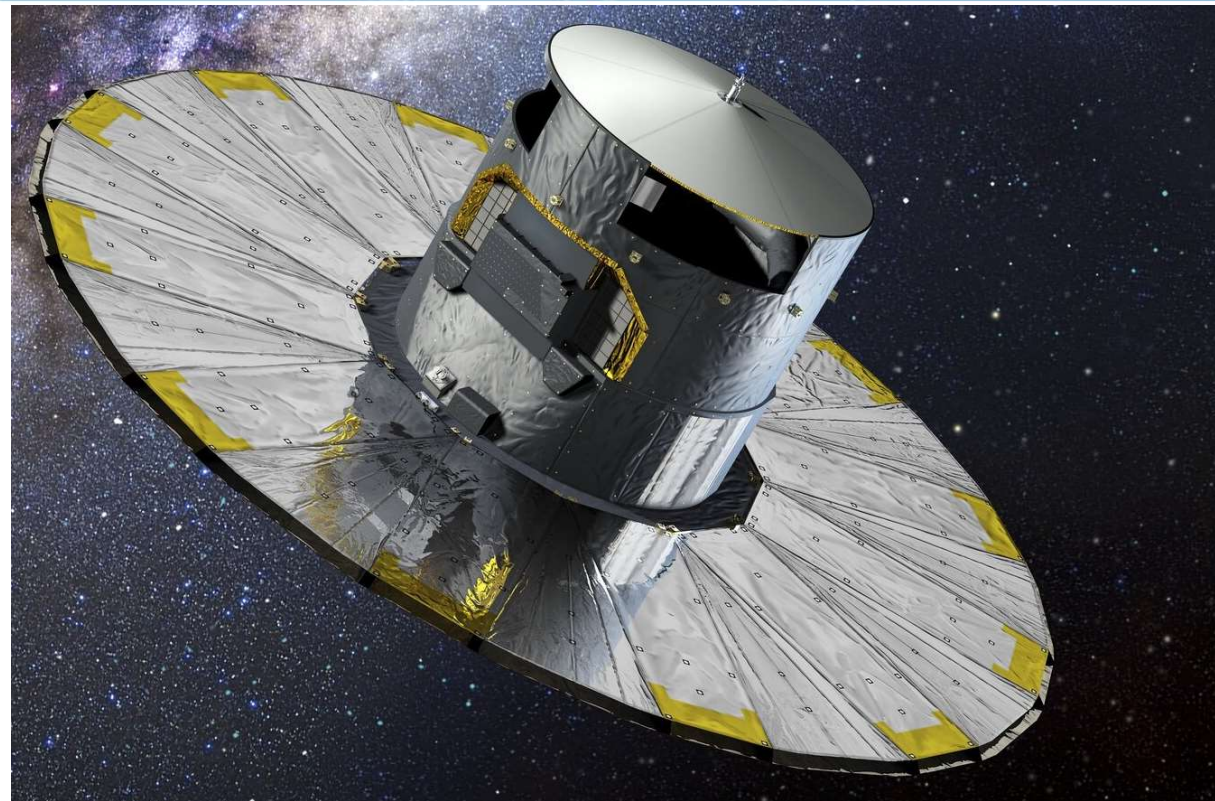
- **Data Release 5 (DR5):**

- *Extended mission (~10 years), date TBD (around 2031)*
- *Already working on it!*



| | # sources in Gaia DR3 |
|--|-----------------------------------|
| Total number of sources | 1,811,709,771 |
| | Gaia Early Data Release 3 |
| Number of sources with full astrometry | 1,467,744,818 |
| Number of 5-parameter sources | 585,416,709 |
| Number of 6-parameter sources | 882,328,109 |
| Number of 2-parameter sources | 343,964,953 |
| Gaia-CRF sources | 1,614,173 |
| Sources with mean G magnitude | 1,806,254,432 |
| Sources with mean G_{BP} -band photometry | 1,542,033,472 |
| Sources with mean G_{RP} -band photometry | 1,554,997,939 |
| | New in Gaia Data Release 3 |
| Sources with radial velocities | 33,812,183 |
| Sources with mean G_{RPVS} -band magnitudes | 32,232,187 |
| Sources with rotational velocities | 3,524,677 |
| Mean BP/RP spectra | 219,197,643 |
| Mean RVs spectra | 999,645 |
| Variable-source analysis | 10,509,536 |
| Variability types (supervised machine learning) | 24 |
| Supervised machine-learning classification for variables | 9,976,881 |
| Specific Object Studies – Cepheids | 15,021 |
| Specific Object Studies – Compact companions | 6,306 |
| Specific Object Studies – Eclipsing binaries | 2,184,477 |
| Specific Object Studies – Long-period variables | 1,720,588 |
| Specific Object Studies – Microlensing events | 363 |
| Specific Object Studies – Planetary transits | 214 |
| Specific Object Studies – RR Lyrae stars | 271,779 |
| Specific Object Studies – Short-timescale variables | 471,679 |
| Specific Object Studies – Solar-like rotational modulation variables | 474,026 |
| Specific Object Studies – Upper-main-sequence oscillators | 54,476 |
| Specific Object Studies – Active galactic nuclei | 872,228 |

Activities within the frame of PPCC



PPCC AstroHEP LIA4 (large surveys) at ICCUB



Overview of activities funded by PPCC:

- **Preparation of Gaia DR4:**
development of algorithms and software for the Gaia **data processing** and analysis, transforming the raw Gaia data into usable science data products
- **Cloud-based data mining of Gaia data:**
define methods and technologies to efficiently **exploit massive amounts of data**
- **Data fusion of Gaia with other catalogues:**
investigate **data fusion techniques** and apply them to Gaia and **J-PAS/J-PLUS** (proof-of-concept)
- **Gaia knowledge transfer and transversal support:**
apply lessons learned in Gaia to **other surveys and projects**

1 postdoc

1 engineer

Commercial Cloud Services

1 engineer

1 engineer

- Strengthen the **national leadership** in Gaia
- Define strategies for **massive data science**
- Find and exploit **synergies** of Gaia with other surveys and projects

PPCC AstroHEP LIA4 (large surveys) at ICCUB



- General progress:
 - So far, mostly **focused on the preparations for Gaia DR4** and the validation of its data products
 - Also good progress on **Cloud** technologies and knowledge transfer
 - More modest progress on catalogues fusion; defining proof-of-concept → discussions with OAJ/CEECA
- Preparation of Gaia DR4:
 - Definition of **multi-epoch descriptors**:
easy catalogue search for certain transients and peculiar objects without requiring massive epoch data analysis
 - Support to **non-single stars** processing unit:
filters and criteria to select the initial candidates of astrometric binary stars (resolved and non-resolved)
 - Support to **extended objects** unit:
gravitational lenses (incl. epoch photometry calibration), galaxies, quasars; understanding of instrumental effects
 - Support to **Solar System objects** unit:
pre-processing and inputs determination
 - Support to **spectroscopic processing** unit:
epoch and catalogue cross-checks; instrumental effects
 - **Core astrometric processing** unit:
analysis of cross-matching resolution and astrometric results; definition of useful indicators for DR4 users
 - Next: **progressive ingestions** of Gaia data into the Archive for **incremental validation** of all products and technical preparations

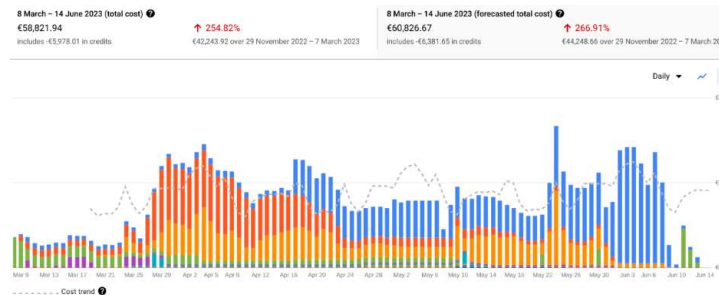
PPCC AstroHEP LIA4 (large surveys) at ICCUB



Expertise on Cloud-based data mining technologies:

- Extensive use of **Commercial Cloud Services** through European funds (OCRE)

- Large Linux Virtual Machines,
- Apache **Spark** cluster, Data Lake,
- **Machine Learning** services,
- Notebooks, **BigQuery**...
- Run large simulations, get richer statistics, find correlations, improve current models
- Outstanding performance achieved with BigQuery on the DR3 bulk catalogue: **1.8B sources ingested in 3 min, complex queries in <10 sec**



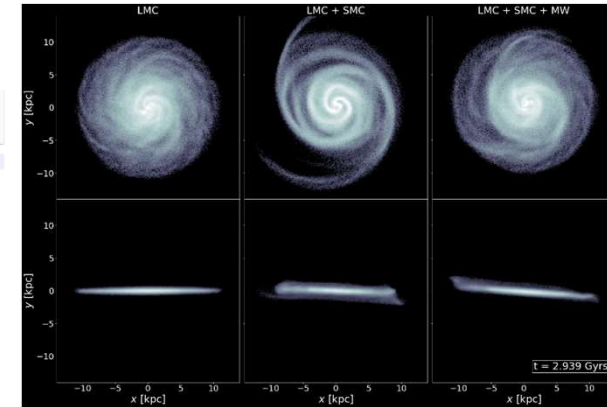
- Gaia Data Analysis Framework (**GDAF**):

- Hadoop + Spark + Parquet + libraries + interfaces, formerly deployed at CESSA/CSUC

- **SPACIOUS**, European project recently granted; massive data mining on Gaia and other missions

- PPCC-funded activities:

- GDAF revision and **deployment at BSC** → tests on DR3 data in their cloud platform
- Now studying tools for quick queries and tests on huge internal DPAC tables (at BSC, to ensure data governance and privacy)



OCRE Open Clouds for Research Environments

Google Cloud

Compute Engine

PPCC AstroHEP LIA4 (large surveys) at ICCUB



Data fusion of Gaia with other catalogues:

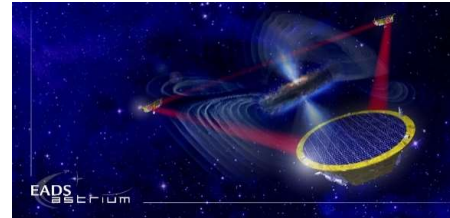
- Identify **limitations and complementary features** between Gaia and other catalogues
- Initial proof-of-concept: OAJ catalogues (**J-PAS, J-PLUS, J-VAR?**)
- Initial concepts and ideas:
 - Cross-match Gaia DR3 with J-PLUS DR3 (if not done already)
 - **Train classifiers** to determine **spectral types** from Gaia data (using astrophysical parameters and spectra), extend training using J-PLUS data → cross-check and extend to unmatched stars in both catalogues
 - **Refine Gaia photometry** by using the J-PLUS one on matching sources, then extending it to the rest of Gaia sources
 - Depending on the outcome, **publish** the combined (fused) catalogue
- On the longer term:
 - Identify additional rich products or data quality improvements achievable through Gaia + J-PLUS data fusion
 - Investigate the **application to other surveys**: WEAVE, Euclid, LSST, PhotSat
 - Develop a **Cloud-based service** to support data fusion of two catalogues, including cross-match and ML services

PPCC AstroHEP LIA4 (large surveys) at ICCUB

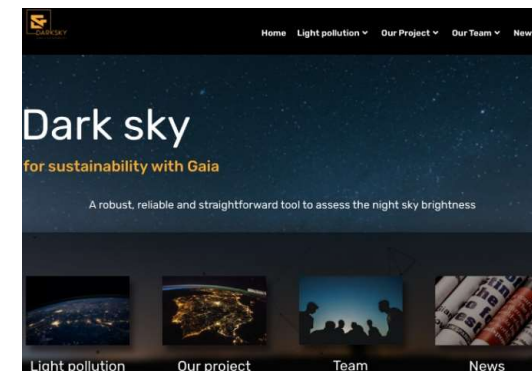


Knowledge transfer from Gaia to other projects:

- Virgo, **PLATO**, Jasmine, **LISA**, GaiaNIR...
- In general: massive data processing pipelines, data handling techniques, astrometric algorithms (attitude, cross-match, instrument model)
- PPCC engineer: focus on **PhotSat** ground segment definition and implementation



- Also proof-of-concept and pre-market projects:
 - **Gaia4Sustainability** → Gaia map of the brightness of natural sky
Evaluate and identify sources of light pollution
 - **B2CATS** → Cloud-based continuous authentication based on behavioral sensing
Apache Kafka, Docker/Kubernetes, optimized data streaming



Thank you

Xavier Luri (xavier.luri@ub.edu)
Jordi Portell (jportell@icc.ub.edu)

on behalf of the Gaia ICCUB-IEEC Team

Funded by:



Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA

