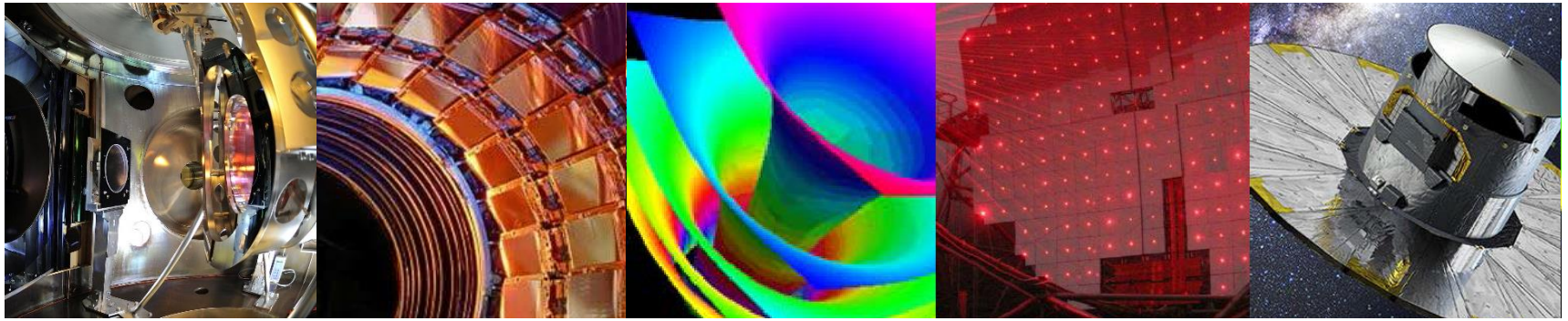
ICCUB

Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA

EXCELENCIA MARÍA DE MAEZTU 2020-2023



Current and future activities on **computing and software engineering** at the ICCUB Technological Unit

Jordi Portell i de Mora
Deputy technical director

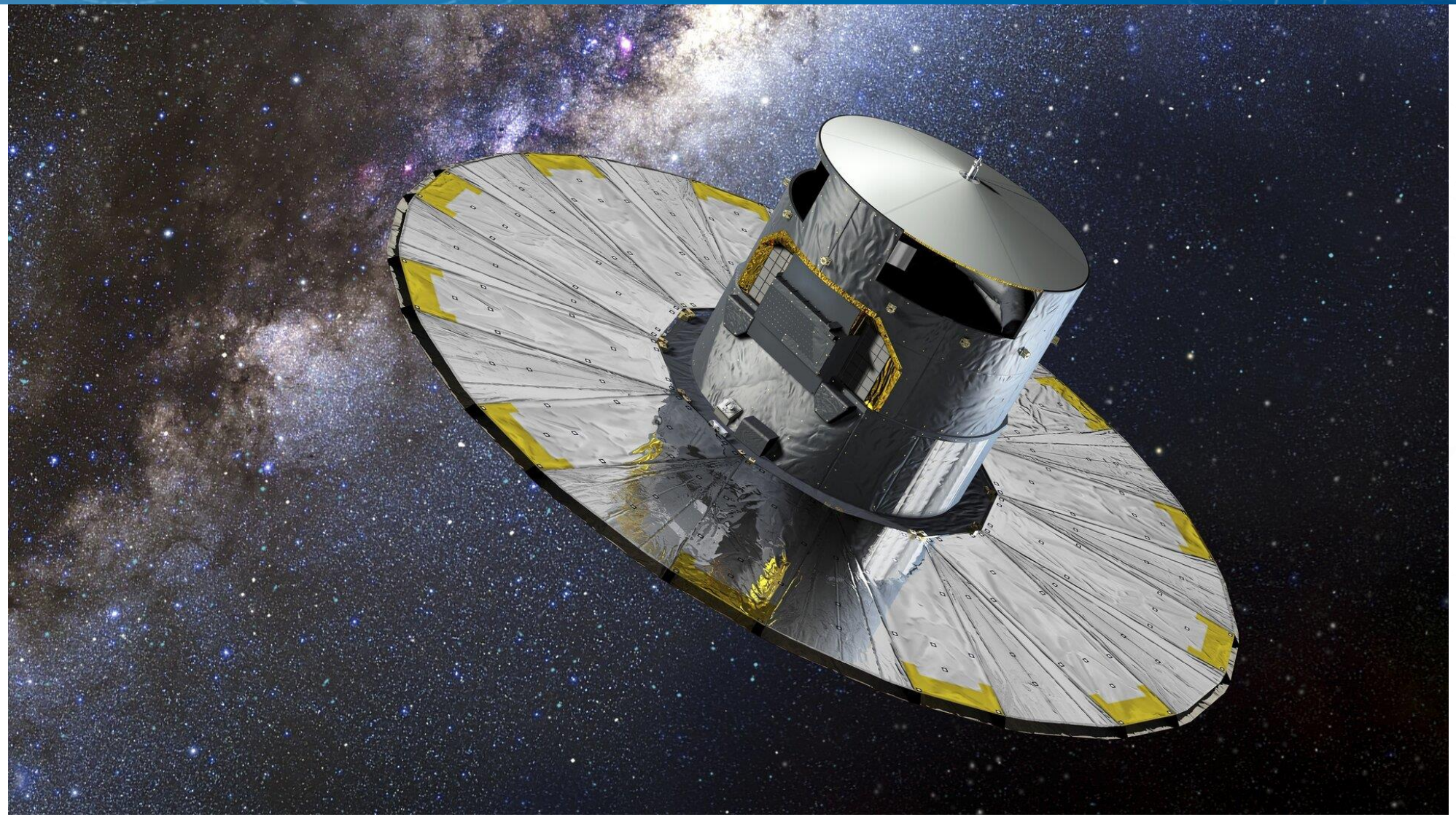
Institute of Cosmos Sciences
Universitat de Barcelona

ICCUB Winter Meeting

Barcelona
8 February 2022

UNIVERSITAT DE BARCELONA

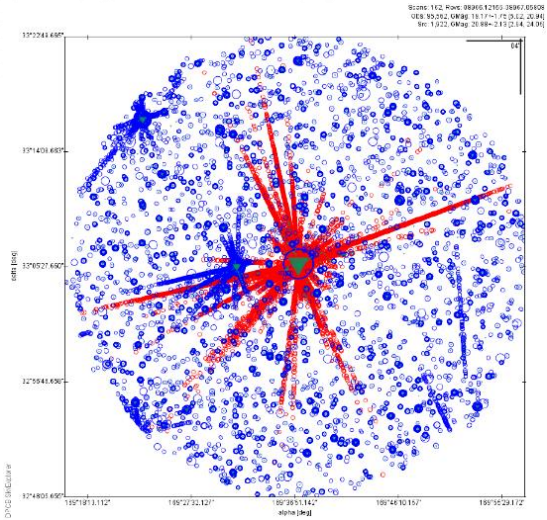
Gaia

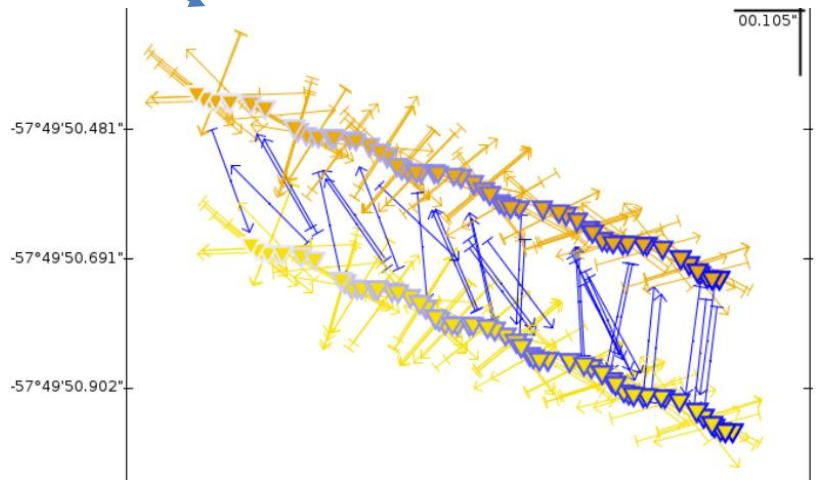


Gaia: Data processing, validation and visualization

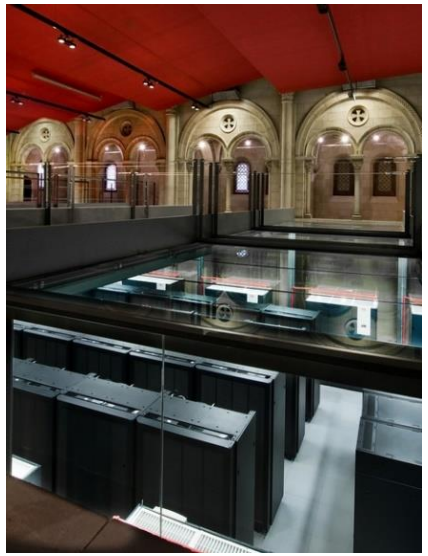
Most activities within the frame of the Gaia DPAC (Data Processing and Analysis Consortium):

- CU3 (Core Processing) unit, IDU (Intermediate Data Updating) system
  - Integration and test of algorithms: Instrumental calibrations, image parameters determination
  - **Development of new algorithms**: Attitude & Calibration bootstrap, spurious detections classification, cross-matching…
  - Recently: on-ground detection and resolution of **close star pairs**
    → improve catalogue resolution and completeness, specially in dense areas (e.g. clusters)





- DPCB (Data Processing Centre of Barcelona)
  - Operational runs at BSC (MareNostrum):
    5.5 years of mission, **65 TB** output, **5M CPU** hours, **142E9** observations processed…
  - Two runs like this (so far!) for this cycle (prep. DR4)
  - Official backup of the full MainDB and raw TM Archive
  - Data visualization tools



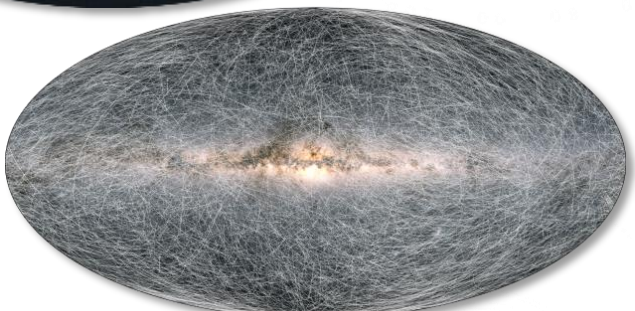
- CU3 / IDT (Initial Data Treatment)
  - Support to daily operations, monitoring and resolution of onboard/onground issues

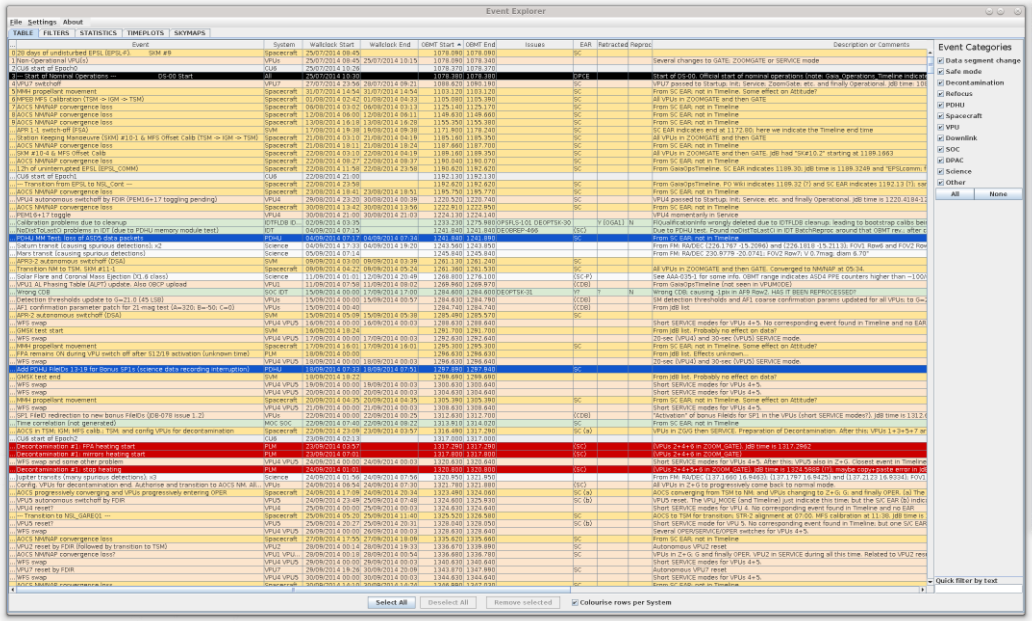
ICCUB Institut de Ciències del Cosmos UNIVERSITAT DE BARCELONA
EXCELENCIA MARÍA DE MAEZTU 2020-2023

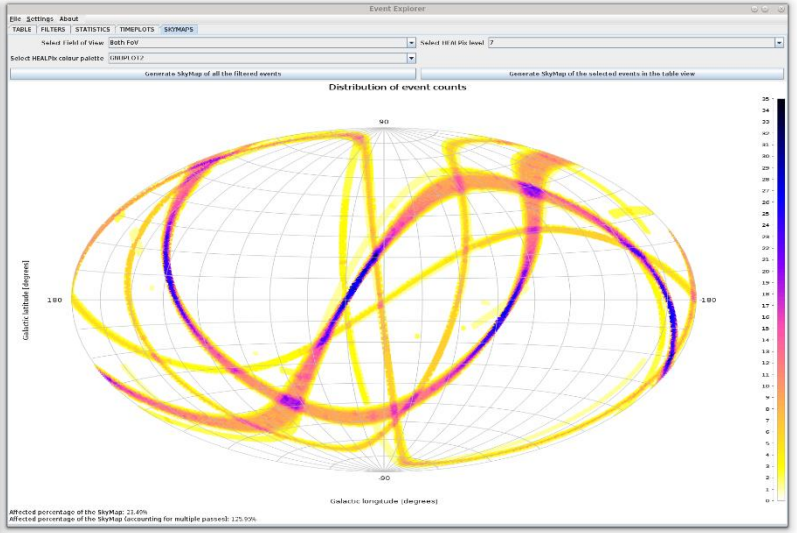
Gaia: Data processing, validation and visualization

Most activities within the frame of the Gaia DPAC (Data Processing and Analysis Consortium):

- CU9 (Catalogue Preparation)
  - Catalogue validation for EDR3 and **DR3**: many new data types, tables, parameters…
  - Development of software tools for statistics and validation
- Project Office
  - Technical interfaces between Units and Centres; technical support to other Units
  - Estimation of database and transfer sizes
  - Curation of Operational Event Logs, development of visualization tools
  - Support to additional (typ. cross-unit) investigations







ICCUB Institut de Ciències del Cosmos UNIVERSITAT DE BARCELONA

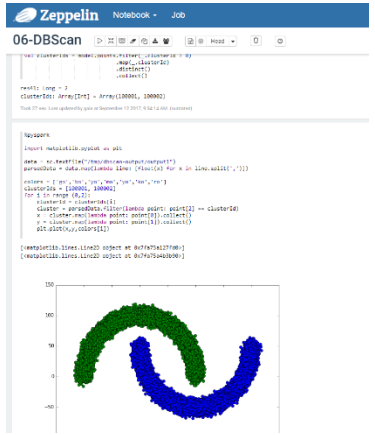
EXCELENCIA MARÍA DE MAEZTU 2020-2023

Gaia: Additional activities

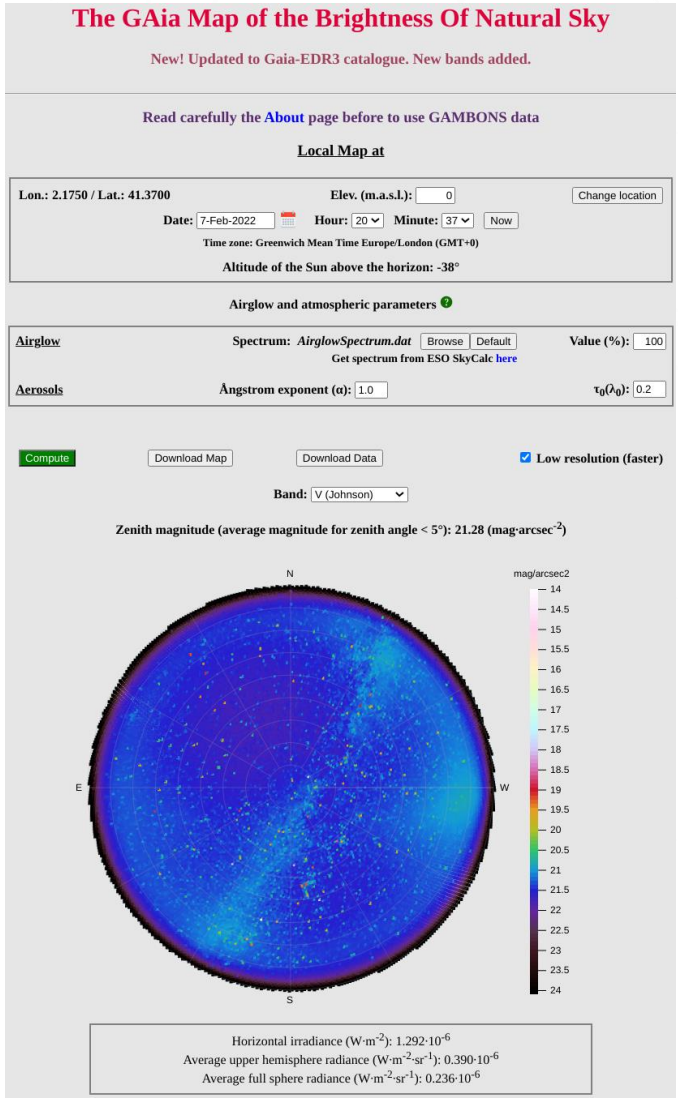
Beyond DPAC, some activities related to Gaia:

- OCRE / GalacticRainCloudS:
  - Galactic Research in Cloud Services
  - **Commercial Cloud Services** granted by OCRE (European Initiative) to do data mining and research on Gaia (E)DR3 data
  - Spark cluster + Data lake + Linux Virtual Machines + Machine Learning services + Notebooks
  - Run large simulations, get richer statistics, find correlations, improve current models

- GDAF:
  - Gaia Data Analysis Framework
  - Hadoop + Spark + Parquet + libraries + interfaces
  - Allow queries, plots and investigations on **Big Gaia Data**
  - Formerly deployed at CESCA/CSUC, now migrating to BSC



- Gaia4Sustainability (a.k.a. GAMBONS Plus):
  - Gaia map of the brightness of natural sky
  - Evaluate and identify sources of **light pollution**
  - Now improving it to offer a “proof-of-concept” service to users
  - Software development including modern web technologies and GPU programming



ICCUB
Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA
EXCELENCIA MARÍA DE MAEZTU 2020-2023

Gaia: Recent achievements and outlook

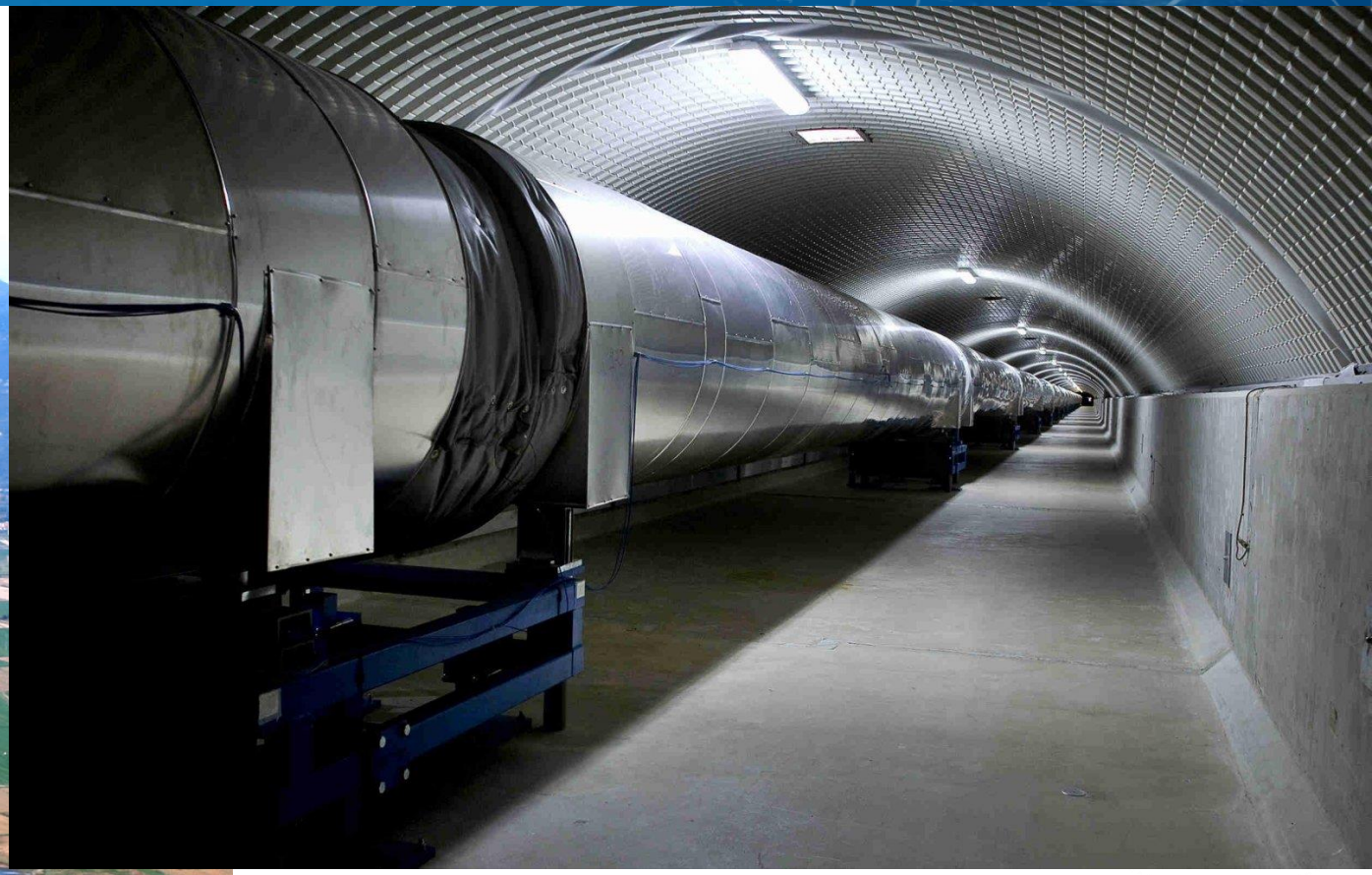
- Early Data Release 3 (EDR3):
  - Released 3 December 2020
  - Mainly Astrometry+Photometry
  - Based on 34 months of data
- **Data Release 3 (DR3):**
  - **Scheduled 13 June 2022**
  - **Lots of new data products**

- Data Release 4 (DR4):
  - Full nominal mission (66 months)
  - Envisaged ~2025
  - **Epoch data** for all data products and sources (incl. astrometry, spectra, etc.)
- *Data Release 5 (DR5):*
  - *Extended mission, date TBD*
  - *Already working on it!*

	# sources in Gaia DR3	# sources in Gaia DR2	# sources in Gaia DR1
Total number of sources	**1,811,709,771**	**1,692,919,135**	**1,142,679,769**
	Gaia Early Data Release 3		
Number of sources with minimally 5 astrometric parameters	1,467,744,818	1,331,909,727	2,057,050
Number of 5-parameter sources	585,416,709		
Number of 6-parameter sources	882,328,109		
Number of 2-parameter sources	343,964,953	361,009,408	1,140,622,719
Gaia-CRF sources	1,614,173	556,869	2,191
Sources with mean G magnitude	1,806,254,432	1,692,919,135	1,142,679,769
Sources with mean G_{BP}-band photometry	1,542,033,472	1,381,964,755	-
Sources with mean G_{RP}-band photometry	1,554,997,939	1,383,551,713	-
	New data in Gaia Data Release 3 (pending validation)		
Sources with radial velocities	≈ 33,000,000	7,224,631	-
BP/RP spectra	> 100,000,000	-	-
RVS spectra	≈ 1,000,000	-	-
Variable source classifications	≈ 13,000,000	550,737	3,194
Object classifications	≈ 1,000,000,000	-	-
Sources with astrophysical parameters	≈ 500,000,000	161,497,595	-
Non-single stars	≈ a few 100,000	-	-
QSO host and galaxy morphological characterisation	≈ a few 1,000,000	-	-
Solar system objects	≈ 150,000	14,099	-
Reflectance spectra for solar system objects	≈ 50,000	-	-
Average BP/RP reflectance spectra of asteroids	≈ 10,000	-	-
Gaia Andromeda Photometric Survey (GAPS)	≈ 1,000,000	-	-

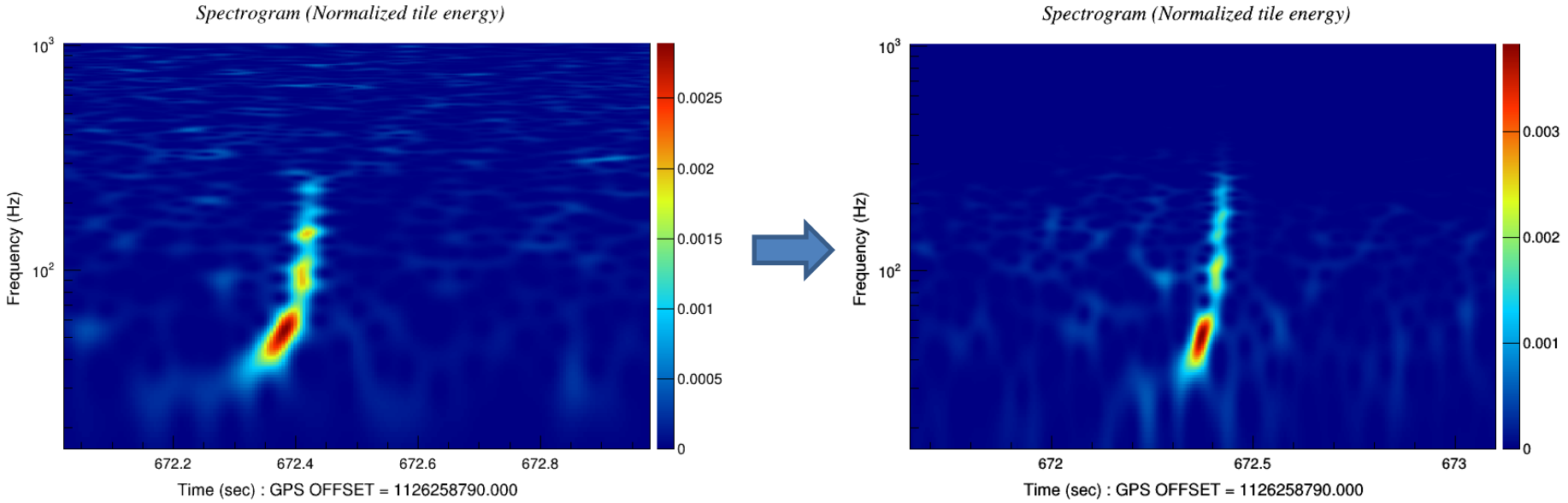
ICCUB
Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA
EXCELENCIA MARÍA DE MAEZTU 2020-2023

Virgo and GW



Virgo: Contributions to the Gravitational Waves Observatory

- ICCUB Virgo participation was triggered from the Technological Unit
  - Initially aiming at contributions on **Computing** and **Instrumentation**
  - Now also **outreach**, **data analysis** and **science modelling and exploitation** activities
  - Full member of the Virgo Collaboration since July 2019
  - ICCUB-Virgo group has grown a lot! Now 17 members and ~6 FTEs
  - Here we just focus on computing and data analysis
- Data analysis:
  - **Denoising** plugin (based on iterative rROF) for Bursts pipeline (unmodelled searches), up to ~2dB SNR improvement



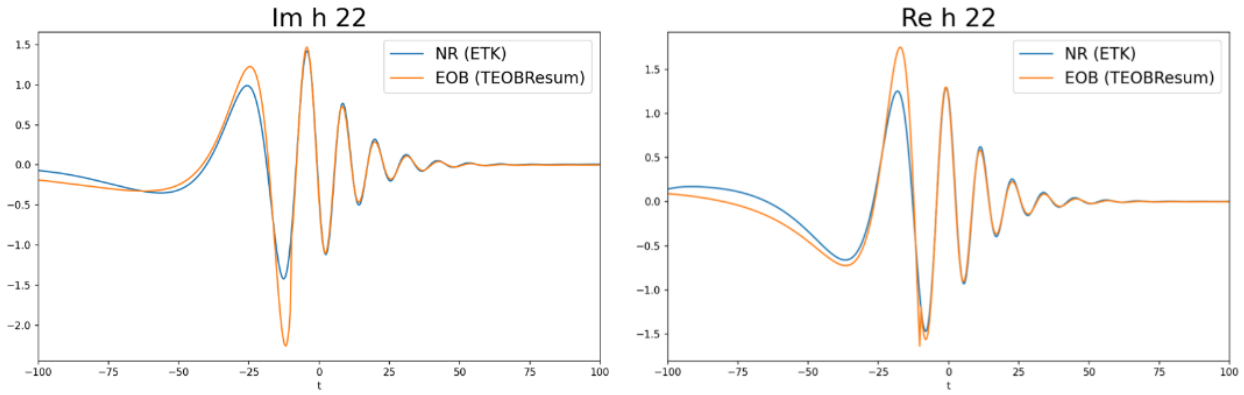
VIRGO ICCUB

ICCUB Institut de Ciències del Cosmos UNIVERSITAT DE BARCELONA

EXCELENCIA MARÍA DE MAEZTU 2020-2023

Virgo: Contributions to the Gravitational Waves Observatory

- Data analysis:
  - Working on new **GW templates** and models:
    High eccentricity, precession, gravitational lensing…
  - Also on new **pipelines** and improved template **interfaces**

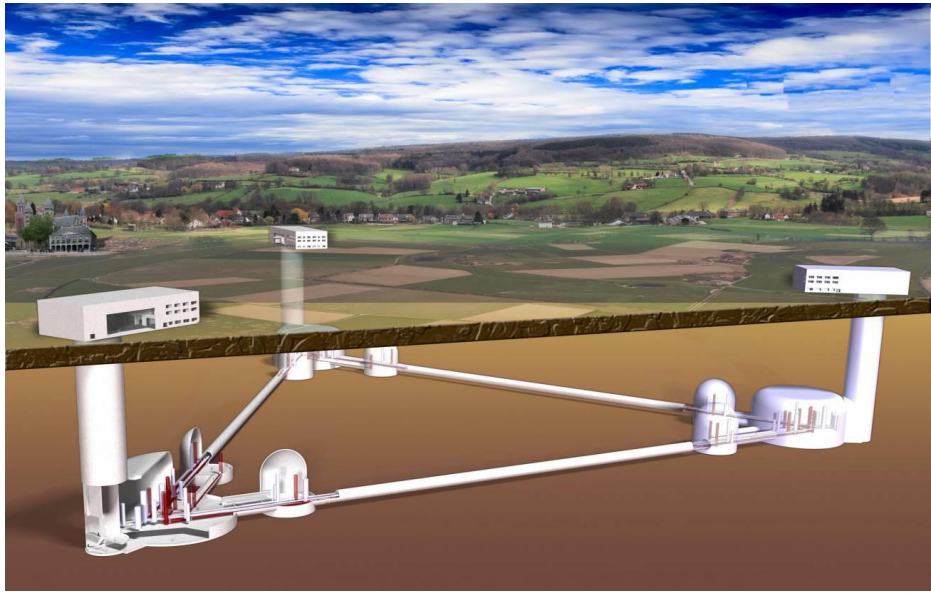


- Computing:
  - Now working on a federated **authentication service** for Virgo (in collab. with LIGO and KAGRA)
  - Soon:
    optimization of Continuous Wave pipeline
    support to scientists
    centralized monitoring of Rucio data handling and HTCondor jobs

Contributions to other **Gravitational Waves** projects

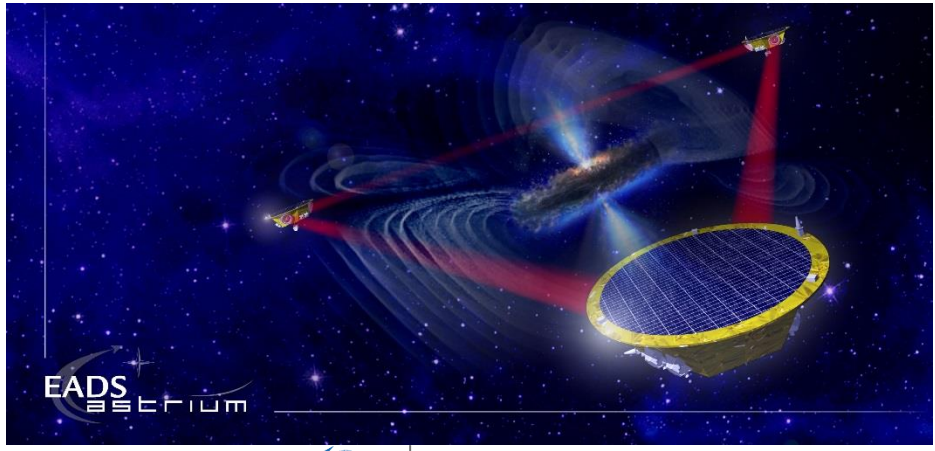
Einstein Telescope (ET):

- Third-generation GW observatory, expected ~2035
- 3 nested detectors, 10km arms, underground, cryogenic parts
- Now part of ESFRI roadmap
- Envisaged contributions from ICCUB:
  - Science case and data analysis
  - Outreach
  - **"E-Infrastructure" (Computing & Software)**:
    Contributions to the general computing model and architecture,
    efficient data handling, cloud and Big Data technologies, software engineering…
    Expected ~1 FTE Q2'2022



LISA:

- Space-based GW observatory, expected ~2037
- 3 detectors, 2.5 million km arms
- ICCUB contributions still being defined:
  probably data challenge / data analysis activities



Other projects

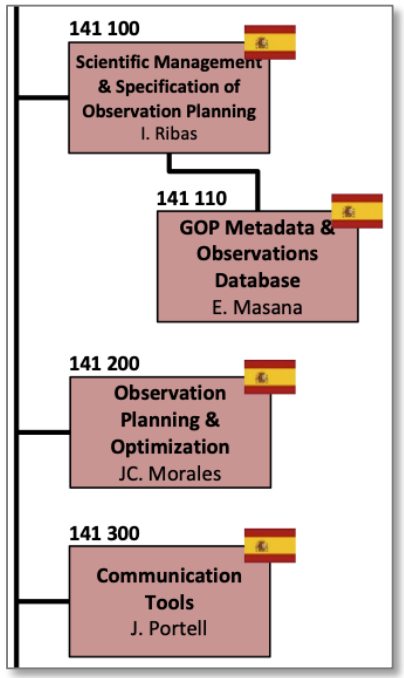
PLATO

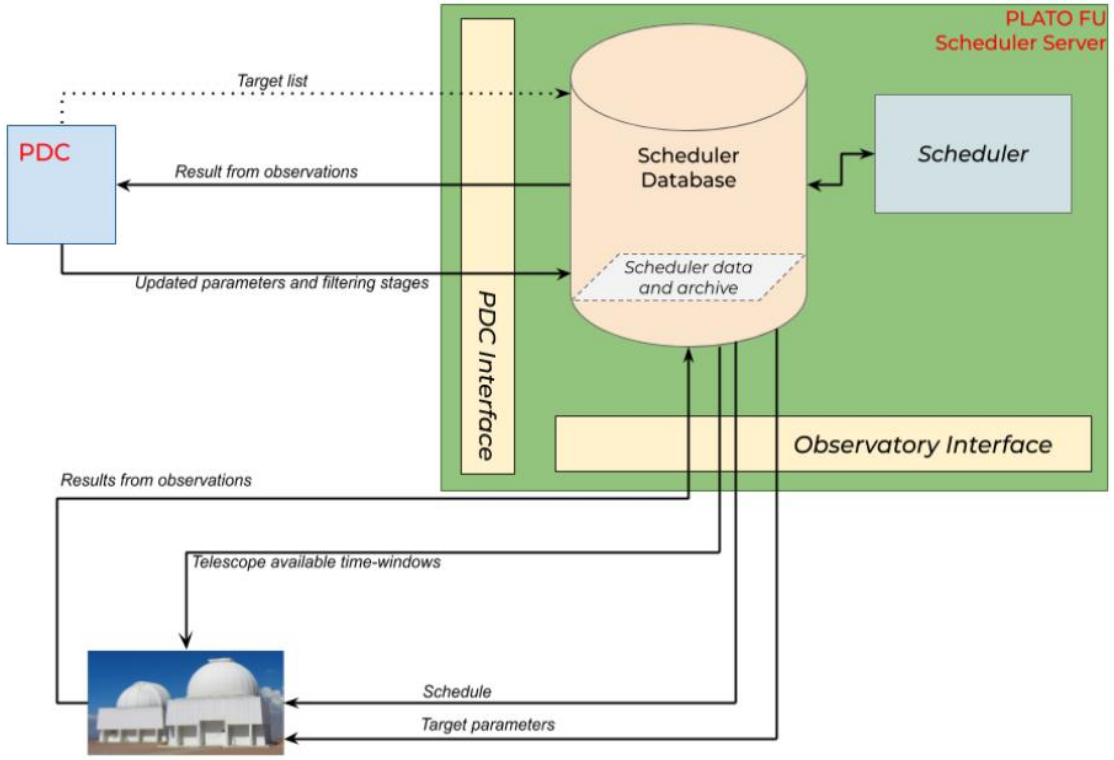
ESA mission, launch expected ~2026.

Ground-based follow-up of exoplanet candidates **(GOP, Ground-based Observation Programme)**:

- Definition of overall architecture and requirements
- Interfaces and protocols between PLATO Data Center and Observatories
- Database and metadata
- Observational constraints
- Software implementation

Now ramping up,
~1 FTE at ICCUB expected Q2'2022





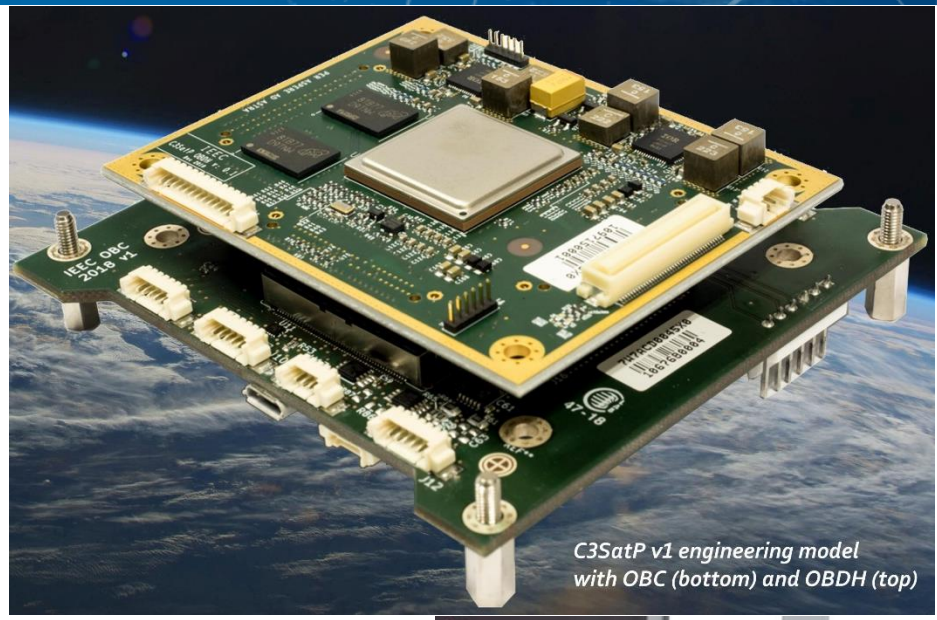
Nanosatellites

See talk by Chema Gómez!

Remarkably (on the software side):

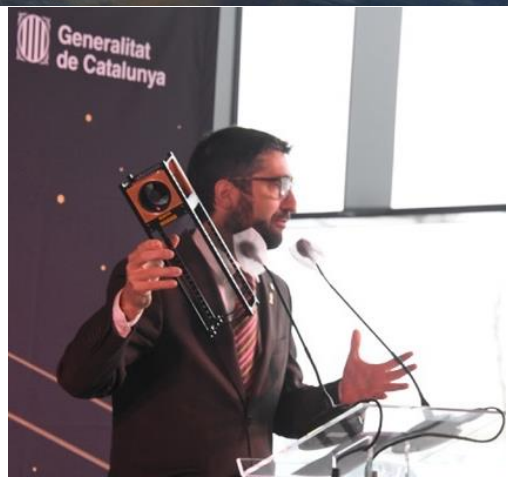
- IEEC's **C3SatP platform**:
  CCSDS packetisation stack,
  Reed-Solomon error detection and correction,
  efficient data compression (FAPEC)
  - Feasibility to compress payload data even on a low-end OnBoard Computer (OBC)
  - High-throughput software-based data compression
  - Collaboration with our spin-off (DAPCOM)

- IEEC's **PhotSat**:
  Support to mission definition and requirements



C3SatP v1 engineering model with OBC (bottom) and OBDH (top)

IEEC
Institut d'Estudis Espacials de Catalunya



ICCUB
Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA
EXCELENCIA MARÍA DE MAEZTU 2020-2023

Other projects and activities

- **GaiaNIR**:
  - Support to initial definition
  - On-board data handling, on-ground raw data processing and initial data treatment, …

- **Euclid**:
  - Specific engineering tasks still being defined.

- **Jasmine**:
  - Also being defined: perhaps support to some PSF models, simulations, and eventually data processing/analysis

- Additional future activities to be funded by *Planes Complementarios*:
  - **Data fusion** of Gaia data with other catalogues: JPAS/JPLUS, WEAVE, Euclid, LSST...
  - Support to other projects: WEAVE, LSST, DESI, Lattice...

Summary

Main activities:

- Gaia
  - DPAC (data processing and validation), Cloud, Big Data, light pollution
- Virgo
  - Computing, pipelines, data analysis

Ramping up:

- PLATO
  - Ground-based followup
- Nanosatellites
  - Onboard software, data compression, mission design
- Other GW projects
  - ET (computing), LISA

Future projects:

- Euclid, Jasmine, GaiaNIR
- Growing soon! (Planes Complementarios)

Thank you

Jordi Portell (jportell@icc.ub.edu)

on behalf of the ICCUB-Tech Computing Division