

# Port d'Informació Científica

## Cosmology & Astrophysics

*J. Carretero on behalf of the team*



# People (April 2024)

---

- Staff Researchers
  - Jorge Carretero (CIEMAT)
  - Manuel Delfino (UAB)
  - Martin Eriksen (IFAE)
  - Christian Neissner (IFAE)
- Staff Research Software Engineers
  - Jordi Delgado (CIEMAT)
  - Pau Tallada (CIEMAT)
  - Francesc Torradeflot (CIEMAT)
- Galactic researcher
  - Kylian Mbappé
- Research Postdocs
  - Laura Cabayol (IFAE)
  - Elizabeth Gonzalez (IFAE)
- Research Software Engineer
  - Agustín Bruzzese (IFAE)
- PhD students:
  - Jiefeng Chan (IFAE)
  - Hanyue Guo (IFAE)
- Master & Grade students:
  - Javier de Mora
  - Miquel Fernández
  - Yasmin L'Harrak

# People

- Staff Researchers
  - Jorge Carretero (CIEMAT)
  - Manuel Delfino (UAB)
  - Martin Eriksen (IFAE)
  - Christian Neissner (IFAE)
- Staff Research Software Engineers
  - Jordi Delgado (CIEMAT)
  - Pau Tallada (CIEMAT)
  - Francesc Torradeflot (CIEMAT)
- Research Postdocs
  - **Laura Cabayol (IFAE) (left Academia)**
  - Elizabeth Gonzalez (IFAE)
- Research Software Engineer (RSE)
  - **Agustín Bruzese (IFAE) (left academia)**
  - **Marc Santamaria (IFAE) A&A**
  - **New RSE (IFAE) TBD**
- **New Web Developer (IFAE) new CosmoHub interface**
- PhD students:
  - Jiefeng Chan (IFAE)
  - Hanyue Guo (IFAE)
  - **ACME PhD student (IFAE) TBD**
- Master & Grade students:
  - **R. Sanfeliu (IFAE) CosmoHub + Cosmo algorithms**
  - **C. Fernández (IFAE) CosmoHub LLM interface**
  - **J. Garcia (IFAE) Streaming dCache billing to Hadoop**
  - **V. Deshpande (Erasmus+; 6 months) TBD**

# Future plans / Group strategy (April 2024)

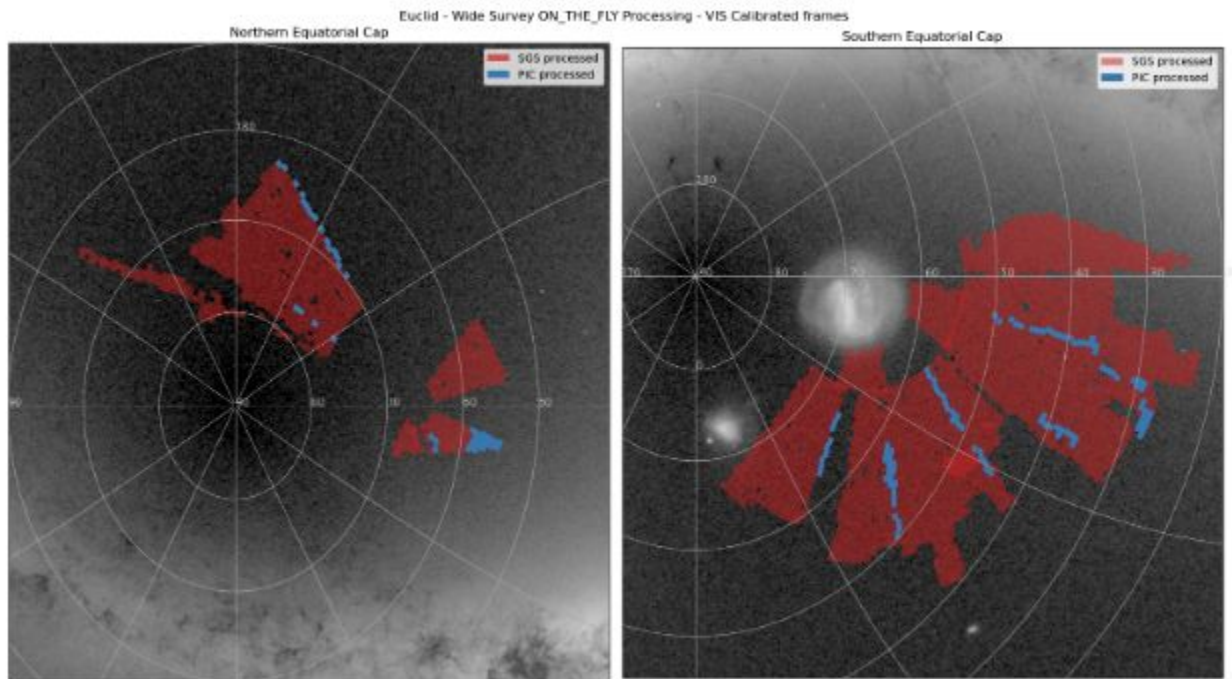
---

- Euclid (*Plan Nacional*)
  - Generation and distribution of mock galaxy catalogs
  - Photo-z
- LSST-DESC (*Plan Nacional, Equipamiento Científico-Técnico*)
  - Provide mock galaxy catalogs
  - lite Independent Data Access Center (lite-IDAC)
- Multi-messenger (*Planes complementarios, European fundings (e.g. ACME)*)
  - MM - Science portal (Optical: IDAC-LSST; Gammas: CTA, MAGIC; GW: Virgo/LIGO, ET; Neutrinos: KM3NeT)
- Applied research with Artificial intelligence (*Spanish and/or European fundings*)

- Euclid (*Plan Nacional*)
  - Generation and distribution of mock galaxy catalogs
  - Photo-z
  - SDC-ES (NOT FUNDED)
- LSST-DESC (*Plan Nacional, Equipamiento Científico-Técnico*)
  - Provide mock galaxy catalogs
  - lite Independent Data Access Center (lite-IDAC)
- Multi-messenger (*Planes complementarios, European fundings (e.g. ACME)*)
  - MM - Science portal (Optical: IDAC-LSST; Gammas: CTA, MAGIC; GW: Virgo/LIGO, ET; Neutrinos: KM3NeT)
- Applied research with Artificial intelligence (*Spanish and/or European fundings*)

- Currently applying for "Plan Nacional 2024" call
  - COSMOFUN: *Simulations and analysis of massive datasets for Cosmology with extragalactic surveys*; PI: J. Carretero; coordinated with IFT/UAM, CIEMAT and IFAE )
- Euclid
  - Generation and distribution of mock galaxy catalogs:
    - Euclid. V. "The Flagship galaxy mock catalogue" (accepted in A&A)
      - *Euclid: Photometric redshift calibration with the clustering redshifts technique (W.d'Assignies D et al. )*
      - *Euclid Preparation. TBD : Simulating thousands of Euclid spectroscopic skies 1000 Mocks (P. Monaco et al.)*
    - Great progress in Euclid Deep mock galaxy catalog (v0 already created)
      - *Including assembly bias in SciPIC recipe (See E. Gonzalez's talk: Populating a dark universe with SciPIC)*
    - CS-SWG - WP1 (mock catalogs) coordinators (E. Gonzalez & J. Carretero)
      - *Close collaboration with CS-SWG - WP3 (mock data challenges)*
    - **TreeCorr in Spark** (all NN correlations)
  - Photo-z
    - FS2 patches including Q1 DR1-like photometry (north and south scenarios)
      - *Estimation of photo-z using Deepz*
    - DNF integration into SciPIC
  - SDC-ES:
    - Fully operative and participating in the OTF and RR data campaigns
    - Contributing to ~5% of the total SGS efforts
    - Running on services developed by Euclid (IAL, DSS, EAS) integrated into the HTCondor, Ceph and dCache infrastructure at PIC
    - **Minimal funding through IFAE PN project**

# SDC-ES processing



- LSST-DESC
  - Provide mock galaxy catalogs
  - lite Independent Data Access Center (lite-IDAC)
    - **Most DP0.2 ingested into CosmoHub** (LSST@Europe6 invited talk)
    - **Authorization & Authentication, and IDAC data transfer meetings** (including UK, Brazil and Poland IDACs)
- Multi-messenger
  - **MM - Science portal** (Optical: IDAC-LSST; Gammas: CTA, MAGIC; GW: Virgo/LIGO, ET; Neutrinos: KM3NeT)
    - **ACME: HORIZON-INFRA-2023-SERV-01** EU-funded project
    - **Efficient distributed spatial matching Spark notebook**
    - Ingestion of several datasets into CosmoHub
    - **GW events data into CosmoHub**
- Applied research with Artificial intelligence
  - **DeepSky: Advancing Cosmological Insights: AI-Enabled Joint Analysis of Galaxy Surveys from Euclid, PAUS, and DESI** (Plan Nacional 2023; AYA; **PI: M. Eriksen** - IFAE)



- **ACME:** Astrophysics Centre for Multi-messenger studies in Europe
- HORIZON-INFRA-2023-SERV-01 EU-funded project:
  - “Ambitious coordinated European-wide optimization of the accessibility and cohesion between multiple leading RI, offering access to instruments, data and expertise, focused on the new science of multi-messenger astrophysics”
- **IFAE** (leading institute; PI: O. Blanch, co-PI: C. Nigro; budget: 663k€)
  - Research Infrastructure: MAGIC Telescopes
- **PIC as MAGIC Data Center** (J. Delgado, J. Carretero)
  - WP4: CosmoHub as multi-messenger portal
    - (140k€) 1 PhD student (3 years) + Operation costs

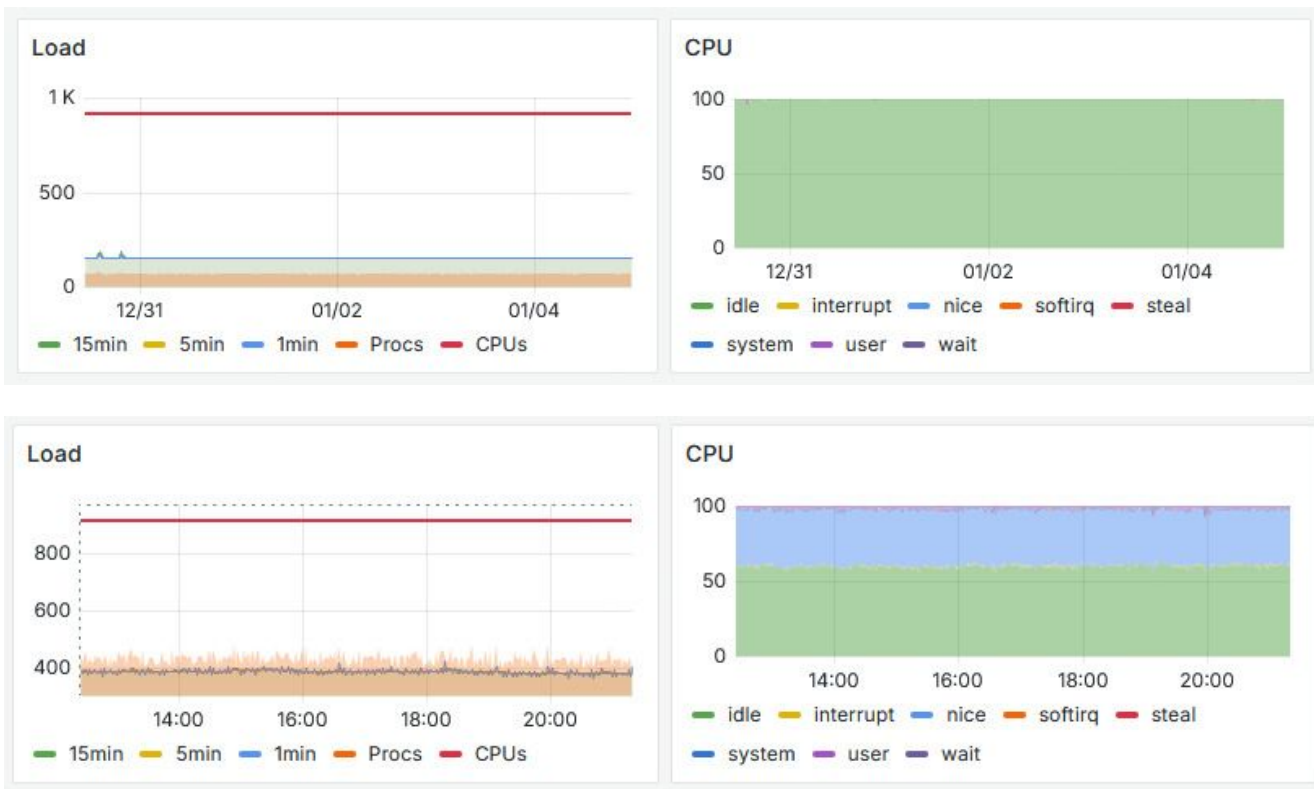
Grant ID: 101131928; Sep 2024 - Aug 2028, 14.5M€  
IPs: CNRS (France); <https://cordis.europa.eu/project/id/101131928>

# Hadoop cluster

- **Big data cluster**
  - 32 nodes (last 12 since last week)
    - Already in production (some configuration and testing still pending)
    - **768 cores, 16 TiB RAM**
    - 60 TB NVMe (for cache)
    - 4.3 PiB raw storage (**3 PiB usable**; ~300 TiB already occupied)
- Great effort to provide **backfilling** for HTCondor jobs
  - Already in production
    - **DUNE and Virgo jobs running successfully**
    - CosmoHub performance unaffected
- Staff shortage (there are no backups for the service administrators)
  - **A. Bruzzese's departure is a significant challenge for our team**
    - No immediate replacement or future funding available



# Hadoop cluster backfilling



- **Euclid**
  - Euclid Q1 data release; WIP (discussion ongoing on access permissions)
  - Flagship v2 patches to estimate photo-z using DR1-like photometry:
    - [FS2\\_sdr1\\_obs + phz](#) (Southern region: LSST + DES)
    - [FS2\\_ndr1\\_obs + phz](#) (Northern region: LSST + UNIONS)
  - [Non-standard Euclid simulations](#) (Euclid preparation paper: [Simulations and nonlinearities beyond  \$\Lambda\$ CDM. 2](#))
  - [FS2 Deep halo catalog SNAP](#) (6 snapshots,  $z = 1.8, 1.5, 1.2, 0.9, 0.5, 0.0$ )  $L_{\text{BOX}}$  size = 1Gpc/h, Part. mass =  $10^8 M_{\text{sun}}/h$
- **LSST-IDAC** (most of DP0.2 tables available but access cannot be provided yet)
- [PAUS master catalog](#) (Value-added final catalog for PAUS wide fields)
- [DES Y6 Gold catalog v2.2](#)
- **ARRAKHIS** simulations to assist with filter characterization
- **Uchuu** public dark matter simulations (8 catalogs, including DR1 and DR3)
- **Multi-messenger**
  - [DESI Legacy Surveys DR10](#) (description pending)
  - **CTA** Science Data Challenge (Events, Obs. and IRF tables)

# CosmoHub in EOSC Macro-roadmap

- **Euclid**

- Euclid Q1 data release; WIP (discussion ongoing on access permissions)
- Flagship v2 patches to estimate photo-z using DR1-like photometry:
  - [FS2 sdr1\\_obs + phz](#) (Southern region: LSST + DES)
  - [FS2 ndr1\\_obs + phz](#) (Northern region: LSST + UNIONS)

CosmoHub has been selected as a success use case in the [EOSC Macro-roadmap](#) thanks to its contributions to promoting open science, fostering interdisciplinary research, and advancing data-driven innovation in Astronomy and Cosmology (Dec 2024; not yet published on the web)

- **LSST**

- **PAUS**

- **DES Y**

- **ARRAKHIS** simulations to assist with filter characterization

- **Uchuu** public dark matter simulations (8 catalogs, including DR1 and DR3)

- **Multi-messenger**

- [DESI Legacy Surveys DR10](#) (description pending)
- **CTA** Science Data Challenge (Events, Obs. and IRF tables)

# Thanks for your attention!

---

## PIC services information

**Credits** to: E. Acción, V. Acín, C. Acosta, A. Alou, J. Carretero, J. Casals, J. Chen, R. Cruz, M. Delfino, J. Delgado, M. Eriksen, D. Graña, H. Guo, J. Flix, E. Gonzalez, G. Merino, C. Neissner, A. Pacheco, A. Pérez-Calero, E. Planas, J. Priego, R. Sanfeliu, M. Santamaría, P. Tallada, F. Torradeflot

[www.pic.es](http://www.pic.es)

### Acknowledgements

The authors of this work express their gratitude to the PIC and CIEMAT teams for their support in these studies and for deploying novel cache services for the CMS experiment in the Spanish region. This project is partially financed by the Spanish Ministry of Science and Innovation (MINECO) through grants FPA2016-80994-C2-1-R, PID2019-110942RB-C22 and BES-2017-082665, which include FEDER funds from the European Union. It has also been supported by the Ministerio de Ciencia e Innovación MCIN AEI/10.13039/501100011033 under contract PID2020-113614RB-C21, the Catalan government under contract 2021 SGR 00574, and the Red Española de Supercomputación (RES) through the grant DATA-2020-1-0039.

CosmoHub has been developed by PIC (maintained by IFAE and CIEMAT) in collaboration with ICE-CSIC. It received funding from the Spanish government (grant EQC2021-007479-P funded by MCIN/AEI/10.13039/501100011033), the EU NextGeneration/PRTR (PRTR-C17.I1), and the Generalitat de Catalunya.