

CosmoHub a big data platform for Astro-Cosmo data exploration and analysis

Friday 10 October 2025 12:36 (9 minutes)

The CosmoHub is a Big Data platform designed to enable the exploration and analysis of heterogeneous datasets within a unified environment. It was originally developed to serve cosmological surveys such as DESI DR1, DES, Euclid Q1, Gaia, Quaia, and Glade+, among others. It has since been extended to support the ingestion and cross-analysis of multi-messenger data. Recent integrations include public event catalogs from ANTARES, all-sky point sources from IceCube, gravitational-wave detections from LIGO/Virgo/KAGRA via GraceDB, as well as test datasets from MAGIC and CTAO.

The web portal and its underlying infrastructure have been enhanced to efficiently handle these diverse data streams. Through both guided and expert user modes, researchers can query, filter, and join datasets via compatible fields, leveraging Apache Hive and Apache Spark parallelization on a Hadoop-based architecture. These data selections can also serve as the basis for cross-correlation functions in counterpart searches. This approach empowers researchers to seamlessly explore correlations across multiple messengers, opening new avenues for discovery in the era of multi-messenger astrophysics.

Authors: DELGADO MENGUAL, Jordi (PIC-CIEMAT); Dr CARRETERO PALACIOS, Jorge (PIC-CIEMAT); Dr TALLADA CRESPI, Pau (PIC-CIEMAT)

Presenter: DELGADO MENGUAL, Jordi (PIC-CIEMAT)

Session Classification: Morning talks