CosmoHub

Exploración interactiva y distribución de datos científicos masivos

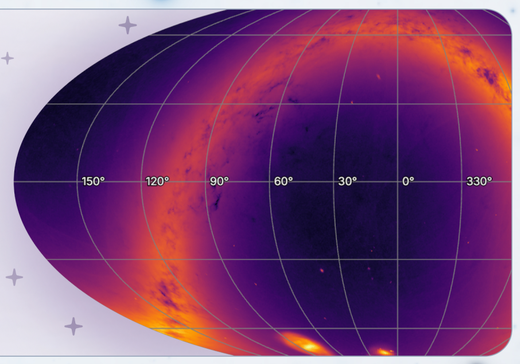
S. van den Bogaart Marzola[13], A. Luelmo Rodríguez[13], M. Santamaría[13], J. Carretero[12], P. Tallada-Crespí[12], J. Delgado[12], C. Acosta[13], R. Cruz[13], E. Gonzalez[34], C. Neissner[13]

[1] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain.
[2] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain.
[3] Port d'Informació Científica (PIC), Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain.
[4] Universitat Autònoma de Barcelona (UAB), Campus UAB, 08193 Bellaterra (Barcelona), Spain.

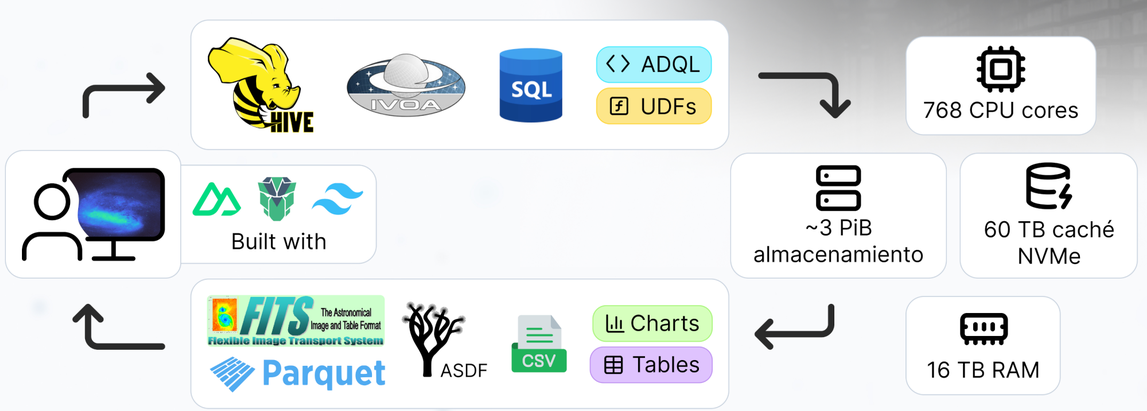
sbogaart@pic.es, aluelmor@pic.es, msantama@pic.es

Explorando el Universo a gran escala

CosmoHub es una plataforma de alto rendimiento desarrollada en el PIC para la **exploración, visualización y distribución de grandes volúmenes de datos astronómicos.** Basada en tecnologías de Big Data, permite a los investigadores **consultar y analizar catálogos científicos masivos directamente desde el navegador**, facilitando el acceso y la reutilización de datos dentro del marco de la ciencia abierta.



Arquitectura



Funcionalidades clave

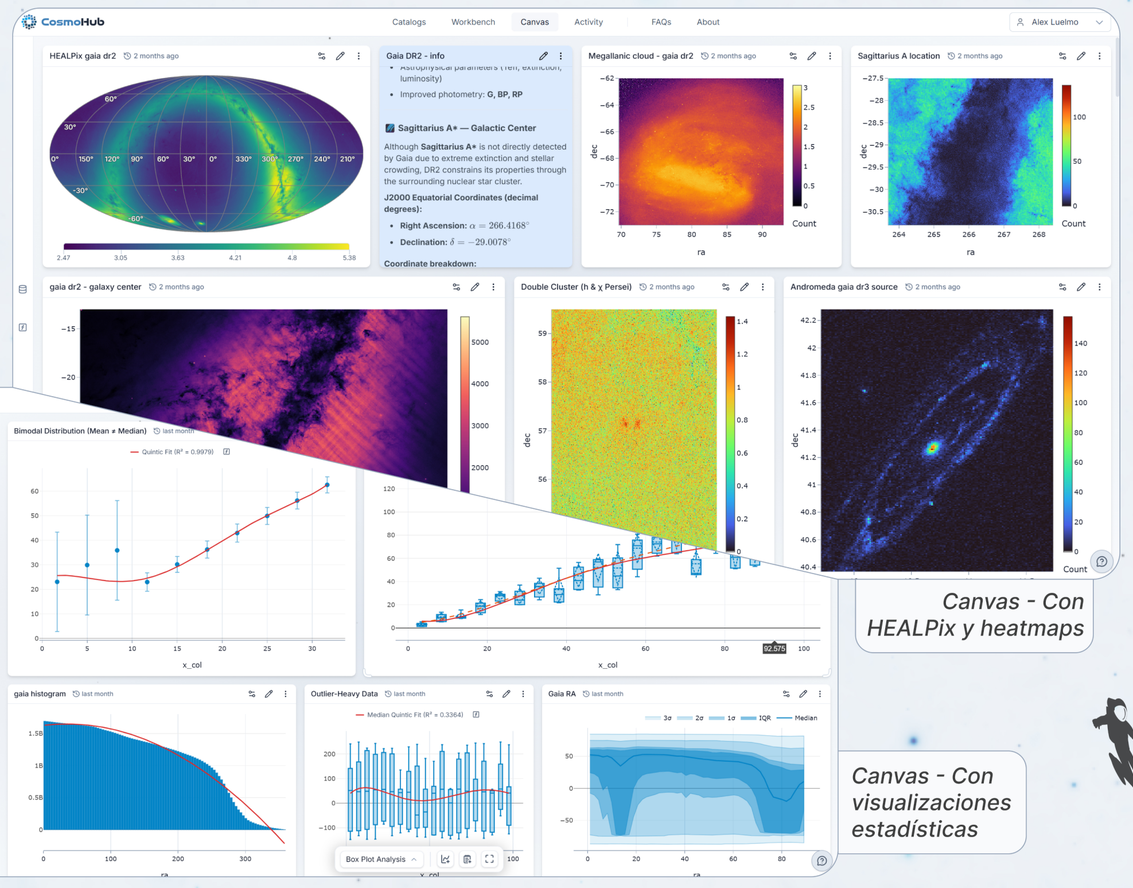
- **Gran conjunto de datos** a disposición de la comunidad, tanto públicos como embargados por los proyectos.
- **Interfaz de usuario guiada** (selección de columnas, filtros, sampling) con construcción automática de consultas SQL.
- **Visualizaciones interactivas** directamente en el navegador, incluyendo histogramas, heatmaps, mapas HEALPix, y nuevas visualizaciones estadísticas de distribuciones, incertidumbres y correlaciones.
- **Generación de catálogos personalizados** bajo demanda en **formatos estándar** (CSV, Parquet, FITS y ASDF).
- **Nueva API** (Python / Litestar) y pasarela WebDAV↔HDFS para acceso a datos.

CosmoHub en cifras

- **Objetos catalogados** > 3×10^{11}
- **Usuarios activos mensuales** ~ 100
- **Catálogos alojados** ~200
- **Catálogos personalizados** > 27 000 (75 % en < 3 min)
- **Datos alojados** > 200 TiB
- **Consultas interactivas** 80 % en < 60 s
- **Citaciones en papers** 102

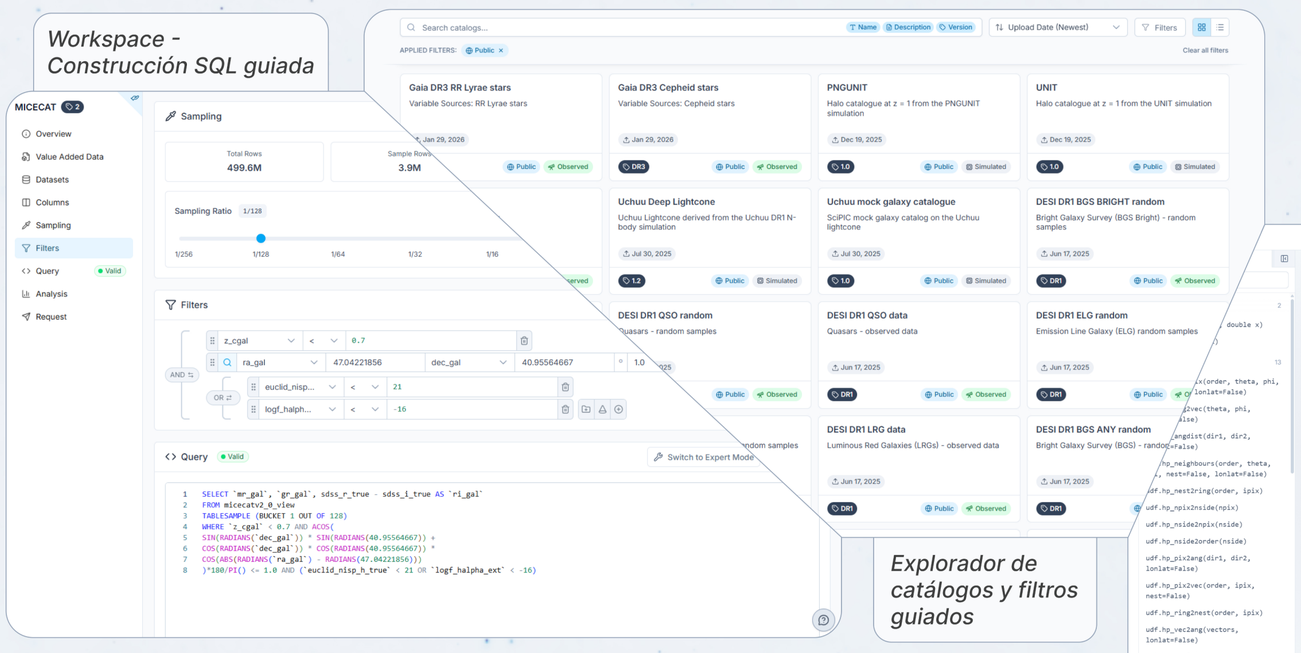
Datos de grandes misiones

euclid · gaia · THE DARK ENERGY SURVEY · DARK ENERGY SPECTROSCOPIC INSTRUMENT · LSST Legacy Survey of Space and Time



Canvas - Con HEALPix y heatmaps

Canvas - Con visualizaciones estadísticas



Workspace - Construcción SQL guiada

Explorador de catálogos y filtros guiados

Hacia una ciencia más abierta

Interoperabilidad con estándares abiertos

- Soporte de **estándares IVOA**: TAP, ADQL, VOTable, VOSpace.
- **Repositorio de datos** de investigación certificado (re3data).
- **Centro afiliado** de Gaia (ESA) y caso de uso destacado de la EOSC.

Autenticación y Autorización (A&A)

- Registro **automático** de cuentas no privilegiadas.
- **Federación de identidades**: ORCID, EduGAIN.
- Escalada de privilegios manual o automática vía **OpenID.**
- Verificaciones de membresía en segundo plano — siempre actualizado.

Referencias

https://www.re3data.org/repository/r3d100013407
https://eosc.eu/use-case/cosmohub
https://cosmohub.pic.es

CosmoHub ha recibido financiación del Gobierno de España (ayuda EQC2021-007479-P financiada por MCIN/AEI/10.13039/501100011033), del programa EU NextGeneration/PRTR (PRTR-C17.I1) y de la Generalitat de Catalunya.