



# CIELOS

CANARY ISLANDS DATA CENTER  
FOR ASTRONOMICAL OBSERVATIONS AND SIMULATIONS

## **Data Management Plan**

(Version 1.0)



# Canary Islands data cEnter for astronomical Observations and Simulations (CIELOS) Data Management Plan

---

## Revision History:

Revision Date	Version	Summary of changes	Changes made by
April 1 <sup>st</sup> , 2025	1.0	Document created.	J. Falcón Barroso (IAC) E. Solano (SVO) R. Gutiérrez (SVO) SIE Investigación (IAC)

## Abbreviations:

CIELOS: Canary Islands data cEnter for astronomical Observations and Simulations

DMP: Data Management Plan

ESA: European Space Agency

FTE: Full-Time Equivalent

GTC: Gran Telescopio Canarias

IAC: Instituto de Astrofísica de Canarias

ING: Isaac Newton Group

NOT: Nordic Optical Telescope

OCAN: Observatorios de Canarias

SI: Servicios Informáticos del IAC (IAC's general IT services)

SIE: Servicios Informáticos Especializados del IAC (IAC's specialized IT services)

SVO: Spanish Virtual Observatory

TNG: Telescopio Nazionale Galileo

UC3: Unidad de Comunicación y Cultura Científica (Science Communication and Outreach Unit)

## Table of Contents:

<b>INTRODUCTION</b> .....	<b>4</b>
<b>CIELOS STRUCTURE</b> .....	<b>4</b>
<b>DESCRIPTION OF THE DATA</b> .....	<b>5</b>
PROVENANCE .....	5
DATA LEVEL .....	5
DATA INTEGRITY.....	5
FORMATS AND STANDARDS FOR DATA.....	5

<b>DATA COLLECTION AND DOCUMENTATION .....</b>	<b>6</b>
METHODS FOR DATA COLLECTION AND GENERATION .....	6
METADATA STANDARDS .....	6
CALIBRATION DATA .....	7
<b>STORAGE AND BACKUP .....</b>	<b>7</b>
STORAGE SOLUTIONS .....	7
BACKUP STRATEGIES .....	7
SECURITY POLICIES .....	7
<b>OWNERSHIP AND ACKNOWLEDGEMENTS .....</b>	<b>8</b>
OWNERSHIP AND INTELLECTUAL PROPERTY .....	8
PROPRIETARY PERIODS .....	8
PUBLICATIONS .....	8
COMPLIANCE WITH OGAN USERS INSTITUTIONS' POLICIES.....	8
<b>DATA SHARING AND ACCESS .....</b>	<b>9</b>
OPEN ACCESS AND RESTRICTED ACCESS .....	9
REPOSITORIES.....	9
PUBLIC DATABASES.....	9
LICENSING AND INTELLECTUAL PROPERTY .....	9
COMPLIANCE WITH VIRTUAL OBSERVATORY POLICIES .....	10
<b>LONG-TERM PRESERVATION.....</b>	<b>10</b>
DATA ARCHIVING AND PRESERVATION.....	10
REPOSITORIES RESPONSIBLE FOR LONG-TERM STORAGE .....	10
<b>ROLES AND RESPONSIBILITIES .....</b>	<b>10</b>
<b>TOOLS AND SOFTWARE .....</b>	<b>11</b>
TOOLS USED FOR DATA COLLECTION .....	11
TOOLS FOR DATA ANALYSIS.....	11
OPEN-SOURCE PIPELINES .....	12
<b>HUMAN &amp; HARDWARE RESOURCES REQUIREMENTS.....</b>	<b>12</b>
DATA VOLUME .....	12
HARDWARE REQUIREMENTS.....	13
HUMAN RESOURCES .....	13
<b>COMMUNICATION .....</b>	<b>14</b>
TRAINING .....	14
NEWSLETTER .....	14
SOCIAL MEDIA.....	14
CONFERENCES .....	14

## INTRODUCTION

CIELOS (Canary Islands data cEnter for astronomical Observations and Simulation) aims to establish a data center dedicated to the storage, processing, and dissemination of astronomical data from the Observatorios de Canarias (OCAN). Its goal is to maximize the scientific impact and accessibility of both observational data and numerical simulations—produced by IAC-led projects or international collaborations.

In partnership with the Spanish Virtual Observatory (SVO), CIELOS will create a specialized unit within the IAC to develop tools and infrastructure for data acquisition, processing, and scientific exploitation. This initiative supports research across all areas of astrophysics, from fundamental particles to the large-scale structure of the universe, and ensures broad dissemination of results from OCAN and major IAC-linked projects.

CIELOS also aims to build a cross-disciplinary team to address the challenges of managing the increasing data volumes from OCAN and global astronomical surveys. By promoting data reuse, it fosters new discoveries, collaborations, and ensures research quality through robust archiving and version tracking—extending its impact beyond astronomy into other data-intensive fields.

This Data Management Plan (DMP) outlines the policies, technical architecture, and governance model for CIELOS. It is designed for both professional astronomers and policymakers, and includes sections covering the full data lifecycle—from collection to long-term preservation—along with sustainability and governance recommendations.

## CIELOS STRUCTURE

### Observatorios de Canarias



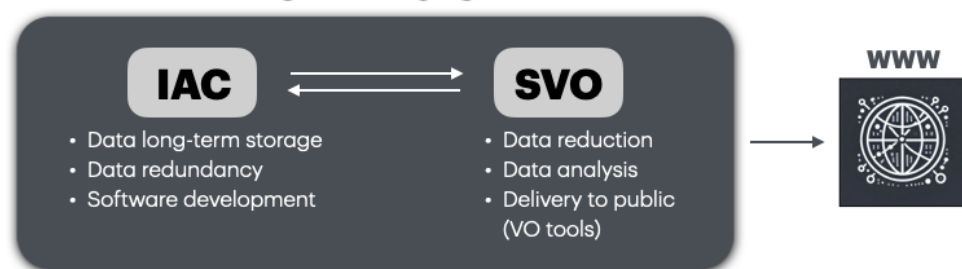
- Data source
- Local repo

### Numerical Simulations



- Data source
- Local repo

## CIELOS



## DESCRIPTION OF THE DATA

### Provenance

- Data are sourced from a wide range of instruments located at the Observatorios de Canarias (Observatorio del Roque de los Muchachos and Observatorio del Teide), including optical, infrared, solar, radio and Cherenkov telescopes. Additionally, data are generated through high-fidelity simulations and models developed both by the IAC and partner institutions.
- Observational parameters — such as geographic coordinates, timestamp, telescope orientation, airmass, and meteorological conditions — are rigorously recorded during acquisition to ensure contextual accuracy and scientific value.
- All data processing steps, from raw data collection to calibrated and derived products, are tracked via standardized processing logs. This ensures full traceability and reproducibility of scientific results.
- Version control mechanisms maintain detailed modification histories, allowing users to access previous data versions and assess changes for quality assurance and scientific comparability.
- Unique identifiers, including standardized metadata tags, are assigned to datasets to facilitate citation, referencing, and integration into research outputs.

### Data Level

Data will be systematically classified and tagged according to their level using metadata schemas, aiding data discovery, indexing, and access control. We envision 4 data levels:

- **Level 0:** Raw, unprocessed data directly from instruments, reflecting the original detector readouts and containing all inherent instrumental and environmental effects.
- **Level 1:** Calibrated datasets with standard corrections applied (bias, dark, flat-field), removing instrumental signatures and enabling consistent scientific interpretation.
- **Level 2:** Enhanced datasets with derived quantities such as spectral features, flux measurements, and astrometric solutions.
- **Level 3:** High-level science products including mosaics, catalogs, light curves, or simulation-based models, often produced by combining multiple Level 2 datasets.

### Data Integrity

To ensure data authenticity and reliability, cryptographic checksums—such as SHA-256—and file hashing algorithms will be employed to verify the integrity of data during both storage and transfer, allowing early detection of any corruption or unauthorized alterations. Validation routines will be seamlessly integrated into data processing pipelines to guarantee accuracy, completeness, and consistency at every stage of the workflow. Regular integrity audits and systematic consistency checks will be conducted to identify and resolve any discrepancies or anomalies within the archived datasets. Additionally, multiple layers of redundancy—including distributed storage systems and RAID configurations—will be implemented to safeguard against data loss and protect against tampering, reinforcing the overall resilience of the data infrastructure.

### Formats and Standards for Data

Several formats will be used to store all the acquired telescope data and simulations. Following international standards, the information will be stored using the following formats:

- **FITS (Flexible Image Transport System)** format is the primary standard for storing astronomical imaging and spectral data.
- **CSV (Comma-Separated Values)** files are used for structured datasets such as object catalogs, light curves, and simulation outputs.
- **JSON** is employed for structured metadata and configuration files due to its lightweight and machine-readable nature.
- **XML** standards enable interoperability with external observatories and facilitate structured data exchange.
- **PARQUET** is a columnar storage file format optimized for efficient data processing, offering fast data retrieval, reduced storage footprint, and support for advanced compression and encoding techniques. It is increasingly adopted as a standard in big data storage and processing applications, including by the IVOA community.
- **HDF5 (Hierarchical Data Format version 5)** is used for storing large, complex, and multidimensional datasets, such as hyperspectral cubes, simulation outputs, and time-series data. Its hierarchical structure and efficient compression make it ideal for managing high-volume scientific data in a scalable and accessible way.

All formats adhere to international astronomical data guidelines established by the IAU, IVOA, ESA, and NASA, ensuring long-term usability and global compatibility.

## DATA COLLECTION AND DOCUMENTATION

### Methods for Data Collection and Generation

Data are primarily collected through telescope instrumentation at OCAN facilities, including imagers, spectrometers, and photometers. In parallel, simulation data are produced using advanced computational models that replicate both observational conditions and underlying astrophysical phenomena. Automated data pipelines handle preprocessing, quality control, and flagging, ensuring that datasets are ready for scientific analysis. Comparative studies often incorporate external datasets from partner observatories or publicly available archives to enhance scientific context. Throughout the data acquisition or generation process, real-time logging systems capture key parameters—such as instrument settings, operator inputs, and environmental conditions—providing a comprehensive record of each observation/simulation.

### Metadata Standards

Metadata encompasses precise timestamps, telescope and instrument identifiers, filter configurations, and exposure durations, providing essential contextual information for each observation. Atmospheric parameters—such as seeing conditions, humidity, and sky transparency—are systematically recorded through observatory weather stations to complement observational data. Calibration datasets are accompanied by detailed uncertainty metrics and reference object information to ensure scientific accuracy and reproducibility. In the context of simulations, metadata consists of the input parameters and settings required to reproduce or understand the simulation results. All metadata will conform to IVOA-compliant formats, enabling seamless integration with Virtual Observatory platforms and enhancing interoperability across research infrastructures. Additionally, Persistent Identifiers (PIDs) are assigned to facilitate traceability and ensure that metadata can be reliably linked and cited across datasets and associated research outputs.

## Calibration Data

Calibration frames—including bias, dark, and flat-field images—are routinely acquired and archived alongside science data to ensure reproducibility and long-term usability. Wavelength calibration is performed using arc lamps and spectral standards, providing accurate spectral alignment across observations. Dark current corrections are applied to mitigate thermal noise, particularly in long-exposure imaging, preserving data quality. Photometric calibration is achieved through nightly observations of standard stars, which also serve to monitor instrument performance over time. To maintain consistency and precision, all calibration processes are benchmarked against independent datasets, ensuring the reliability and scientific integrity of the final data products.

## STORAGE AND BACKUP

### Storage Solutions

High-capacity storage servers located in each facility at the observatories, provide secure, high-speed access to observational data. CIELOS will draw the observed data from those facilities with the goal to support broader collaborative research efforts, with hardware infrastructure at both the IAC and the SVO premises offering scalable storage solutions and remote accessibility ensuring rapid availability for general users. A distributed storage architecture will be designed to enhance data reliability through redundancy and fault tolerance across multiple nodes, safeguarding against hardware failures. For processing data-intensive workloads, high-performance computing (HPC) clusters with dedicated storage arrays are employed, enabling efficient handling of large-scale datasets.

### Backup Strategies

Incremental backups will be performed routinely, complemented by full system backups on a monthly basis to ensure comprehensive data protection. To safeguard against local disasters, offsite backups are maintained at geographically separate locations (e.g., IAC Headquarters, SVO premises) to provide continuous availability and system resilience in the event of hardware failures. Backup validation scripts will be routinely executed to test recovery integrity and confirm the reliability of archived data. Additionally, cold storage archives will be used to preserve historical datasets, ensuring their long-term accessibility and availability for future scientific analysis.

### Security Policies

Access to data is managed through role-based authentication systems, ensuring that permissions are tailored to individual user roles and responsibilities. Sensitive data are safeguarded through encryption both during transmission and while at rest, protecting against unauthorized access. Comprehensive logging mechanisms track all access attempts and flag potential security breaches, enabling prompt response and forensic analysis. Security policies are fully aligned with institutional cybersecurity frameworks and comply with national data protection regulations (e.g., Esquema Nacional de Seguridad), ensuring a robust governance structure. Regular security audits and system updates are carried out proactively to address evolving threats and maintain the integrity and resilience of the data infrastructure.

## OWNERSHIP AND ACKNOWLEDGEMENTS

### Ownership and Intellectual Property

Data ownership is vested in the operating institution or the principal investigator (PI), in accordance with OCAN governance policies. Redistribution and licensing practices are guided by open science principles, while also respecting the proprietary rights and intellectual property of data contributors. Formal credit is given to instrument teams, data reduction specialists, and collaborating researchers to ensure proper recognition of their contributions. In collaborative projects, data-sharing agreements clearly define joint ownership arrangements and outline authorship roles, fostering transparency and accountability. Any disputes concerning data ownership or usage are addressed through established institutional arbitration protocols, ensuring fair and equitable resolution.

### Proprietary Periods

Observational data typically undergo a proprietary embargo of 12 to 24 months, allowing the principal investigator (PI) and their team exclusive access for initial analysis. The duration and scope of proprietary use are defined by institutional policies, ensuring a balance between scientific discovery and open data principles. In cases of scientific urgency, time-sensitive results may qualify for expedited public release to maximize their impact. Once the embargo period expires, datasets are transitioned to open access, accompanied by comprehensive documentation to facilitate broader scientific use. Access policies will be periodically reviewed and updated to align with evolving funding requirements and institutional mandates, ensuring continued transparency and accessibility.

### Publications

Proper citation of datasets in scientific publications is mandatory, using assigned DOIs and standardized metadata references to ensure traceability and academic recognition. Researchers are encouraged to publish in open-access journals, enhancing the visibility, accessibility, and impact of both the data and the resulting research. Establishing direct links between publications and their underlying datasets strengthens reproducibility and fosters greater transparency across the scientific process. All data usage must adhere to established ethical standards and attribution protocols, ensuring that contributors receive appropriate credit. Publication practices are designed to align with institutional guidelines, funder requirements, and international research policies, reinforcing a culture of responsible and open scientific communication. Acknowledgement of the use of CIELOS facilities must be included in every publication.

### Compliance with OCAN Users Institutions' Policies

Data policies are fully aligned with the internal governance frameworks of all OCAN-affiliated institutions, ensuring consistency across participating organizations. Data sharing practices are designed to support European and international scientific cooperation, promoting interoperability and collaborative research. Cross-institutional agreements facilitate seamless integration of datasets and foster a cohesive research environment. Adherence to funding agency requirements further ensures responsible research conduct and accountability. Above all, CIELOS prioritizes ethical and equitable access to data for researchers and the general public, reinforcing a commitment to transparency, inclusivity, and scientific integrity.



## DATA SHARING AND ACCESS

### Open Access and Restricted Access

CIELOS implements a tiered access model that distinguishes between open, institutional, and restricted datasets, ensuring appropriate levels of data availability based on user roles and data sensitivity. High-throughput data transfer protocols, such as Globus, will be supported to facilitate efficient and reliable access to large datasets. Limited guest access options could also be available to support educational initiatives and collaborative research projects. Access controls are enforced through secure user authentication, usage tracking, and quota management systems to safeguard resources and ensure responsible data usage. All access policies are designed in full alignment with the FAIR principles—ensuring that data remain Findable, Accessible, Interoperable, and Reusable for the broader scientific community.

### Repositories

CIELOS repositories (both at the IAC and SVO premises) serve as a cornerstone for long-term data hosting and robust search capabilities, ensuring that scientific datasets remain accessible and well-organized over time. Each dataset is systematically linked to associated publications and research grants, enhancing traceability and reinforcing the connection between data and scientific outcomes. The repositories will be fully integrated with national and European Union data infrastructure initiatives, promoting alignment with broader research ecosystems. Standardized metadata enhances indexing and improves external discoverability across digital platforms and scholarly networks. To ensure continued service and relevance, comprehensive sustainability plans will be established, to secure the long-term viability and operational resilience of the repositories.

The IAC repository is intended to serve as the long-term storage facility for all data produced within CIELOS (Levels 0 through 3). A duplicate copy of this data will also be maintained at the SVO facilities to enable public access through the Virtual Observatory (VO) suite of tools.

### Public Databases

Public access to data fully complies with open science mandates established by funding agencies, ensuring transparency and accountability in research. User-friendly, web-based interfaces will facilitate intuitive browsing, interactive visualization, and straightforward downloading of datasets. To maintain fair usage, download management tools will be implemented to enforce equitable access policies and prevent resource overuse. Additionally, data reuse will be actively monitored through citation tracking and usage analytics, providing valuable insights into the impact and reach of the archived datasets.

### Licensing and Intellectual Property

Datasets within CIELOS will be generally released under Creative Commons licenses, providing clear and transparent guidelines on usage rights and promoting open scientific collaboration. In cases where datasets remain proprietary, they will be explicitly labeled with appropriate restricted licenses to define access limitations. The established licensing frameworks are designed to balance collaborative data sharing with the protection of creator rights, fostering responsible and ethical reuse. All licensing practices are fully compliant with both European Union and national intellectual property regulations, ensuring legal clarity and institutional integrity.

## Compliance with Virtual Observatory Policies

Data managed by the SVO will be formatted and tagged using metadata schemas compliant with Virtual Observatory (VO) standards, ensuring seamless integration and interoperability within the broader astronomical data ecosystem. Access to CIELOS datasets will be facilitated through established VO protocols such as SIAP, SSAP, and TAP, allowing researchers to efficiently query and retrieve data across platforms. All publicly released datasets will be registered in VO repositories and registries, enhancing global discoverability and scientific reuse. To maintain long-term compatibility, datasets and associated metadata will be periodically reviewed and updated in accordance with evolving VO standards.

## LONG-TERM PRESERVATION

### Data Archiving and Preservation

Long-term data preservation within CIELOS relies on archival-grade storage media, such as LTO tapes and write-once, read-many (WORM) systems, to ensure durability and security over extended timeframes. The relevant infrastructure will be placed at the IAC's Headquarter premises. In order to prevent format obsolescence, comprehensive data migration policies will be defined, enabling periodic conversion of datasets to current and widely supported standards. Scheduled validation checks will be routinely performed to verify data integrity and preservation fidelity, safeguarding against degradation or loss. These preservation strategies are designed not only to protect the scientific record but also to guarantee open accessibility and usability well beyond the lifespan of individual projects or research initiatives.

### Repositories Responsible for Long-Term Storage

Long-term archival responsibilities within CIELOS will be managed by the designated personnel at the IAC and SVO premises. Geographic redundancy across multiple storage locations enhances resilience against regional outages or disruptions, safeguarding data availability at all times. Preservation practices strictly follow ISO-certified standards and OAIS (Open Archival Information System) models, providing a robust and internationally recognized framework for long-term data curation. Clear retrieval protocols will be in place to facilitate efficient and reliable access for future research needs, while continuous usage monitoring will help inform repository maintenance strategies and guide periodic system upgrades, ensuring sustainability and responsiveness to evolving scientific demands.

## ROLES AND RESPONSIBILITIES

Effective data stewardship within CIELOS relies on clearly defined roles and responsibilities across the data lifecycle. The structure of those are defined as follows:

- **Data management leads** are responsible for coordinating storage strategies, managing access controls, and ensuring adherence to data governance policies.  
*Responsibility: IAC's CIELOS managing team, SVO representatives*
- **Archive curators** oversee the cataloging, validation, and comprehensive documentation of datasets to maintain accuracy and traceability.  
*Responsibility: IAC's SIE, SVO staff*

- **Software architect and software engineer** in charge of the design, development, implementation and maintenance of the archive system (databases, web interface, VO protocols, programmatic access via Jupyter notebooks, and visualization tools).  
*Responsibility: SVO staff*
- **Security officers** play a critical role in safeguarding data privacy, ensuring regulatory compliance, and mitigating cybersecurity risks.  
*Responsibility: IAC's SI, SVO staff*
- **IT staff** are tasked with maintaining the underlying hardware infrastructure and software environments, ensuring system reliability and operational continuity.  
*Responsibility: IAC's SI and SIE, SVO staff*
- **User's committee** is expected to provide overall guidance and high-level requirements enhancing the scientific value and reusability of the archive.  
*Responsibility: committee to be defined by Data Management Leads*

In addition to their existing responsibilities, the IAC's SIE IT services will also oversee the development of data processing pipelines (from Level 0 to Level 3) and the integration of numerical simulations into the CIELOS framework. Meanwhile, the design, development, deployment, and maintenance of the database, along with the various data access systems—such as web portals, programmatic interfaces via Jupyter notebooks, and Virtual Observatory (VO) services—will be managed by the SVO team. (See the [Human Resources](#) section for further details and requirements)

Technical assistance will be offered to address data quality concerns, troubleshoot access issues, and resolve errors within processing pipelines will be offered by relevant IT staff. Continuous improvement of services will be encouraged through open feedback channels, allowing users to share suggestions and report challenges. For time-sensitive or critical issues, rapid response teams will be on standby to deliver prompt and effective support, ensuring minimal disruption to scientific workflows.

## TOOLS AND SOFTWARE

### Tools Used for Data Collection

Data intake at the observatories relies on a robust suite of tools and technologies to support efficient data collection and processing across its observatory network. Each instrument is paired with dedicated control and acquisition software tailored to its specific operational requirements, ensuring precise and reliable data capture. Real-time monitoring dashboards provide continuous oversight of data acquisition status, enabling prompt responses to system anomalies. The equivalent infrastructure for numerical simulations is the High-performance computers (HPC) at the local host institution used to generate all the outputs. There would be typically large super-computing centres around the world (e.g., MareNostrum, PizDaint).

### Tools for Data Analysis

The Spanish Virtual Observatory (SVO) offers a powerful and comprehensive suite of tools and services that support every stage of astronomical data analysis. It enables researchers to seamlessly discover multi-wavelength data—including images, spectra, and catalogs—across numerous archives, all accessible in standardized, interoperable formats. Through its intuitive interfaces and VO-compliant infrastructure, the Virtual Observatory provides efficient access to data from major observatories and facilitates in-depth analysis using a diverse array of tools. Imaging data analysis is supported by tools

like Aladin, while photometric studies benefit from integrated services like the Filter Profile Service and VOSA, which streamline synthetic photometry and spectral energy distribution fitting. Efficient management of data in tabular format is possible thanks to TOPCAT whereas, for spectroscopic analysis, SVO offers access to both observed spectra and theoretical models that can be visualized, compared, and used to derive physical parameters. Specialized tools such as Clusterix, for determining stellar cluster membership highlight SVO's capability to extract scientific insights from large datasets. Most importantly, all components are fully integrated within the Virtual Observatory framework, ensuring seamless interoperability and reusability across platforms. SVO's ecosystem allows researchers to move effortlessly from data discovery to processing, analysis, and visualization, significantly enhancing scientific workflows. Ultimately, the Spanish Virtual Observatory empowers astronomers to make the most of today's vast and complex data landscape through a highly organized, VO-enabled infrastructure.

### Open-source Pipelines

Dedicated data reduction pipelines will be developed to transform raw observational data (Level 0) into fully calibrated and enhanced data products (Level 1), ensuring the removal of instrumental and environmental artifacts. These pipelines will be made openly available under open-source licenses and hosted in publicly accessible Git repositories to promote transparency, community collaboration, and reproducibility. Recognizing the diversity of instrumentation across the Canary Islands Observatories, each instrument-telescope pairing will require a customized data reduction approach tailored to its specific data characteristics, calibration needs, and observing strategies. Particular emphasis will be devoted to the generation of Level 2 science ready data products. Further efforts will be dedicated to developing data analysis pipelines for the generation of Level 3 products.

To address these requirements, software development will be driven by the telescope-operating institutions—such as the IAC, GTC, ING, TNG, and NOT among potential others—each contributing domain-specific expertise to the design and implementation of robust pipelines. These efforts will be strongly supported by the SIE services at the IAC, who will provide infrastructure, version control, continuous integration systems, and support for software testing and deployment.

In addition, the development process will emphasize modularity and interoperability, ensuring that common calibration tasks and data formats (e.g., FITS compliance, IVOA metadata standards) are standardized across instruments whenever feasible. Where appropriate, pipeline components will be integrated into larger data processing frameworks compatible with Virtual Observatory tools and HPC/cloud environments, facilitating scalability for massive data sets. Documentation, usage guides, and example datasets will be included to support users, while community contributions will be encouraged through open issue tracking and collaborative code development.

Ultimately, these dedicated pipelines will ensure a consistent, high-quality transformation of raw observational data into scientifically usable products, enabling researchers to focus more on analysis and interpretation while maintaining full traceability and reproducibility in the data lifecycle.

## HUMAN & HARDWARE RESOURCES REQUIREMENTS

### Data Volume

The total data volume managed by CIELOS is expected to grow continuously as observational efforts expand and additional facilities begin using the CIELOS infrastructure for data archiving and

distribution. During the initial phase, we aim to integrate data from the Gran Telescopio Canarias, IAC80, Telescopio Carlos Sánchez, and the QUIJOTE experiment. The combined data accumulated by these facilities over the past decade already requires approximately 500 TB of storage. With the introduction of new instruments and the generation of additional data products, we project a 2-fold increase in storage demand at the data center in a 5-year timeframe.

Numerical simulations also represent a significant storage challenge. Over the past decade, simulation efforts at the IAC have consumed roughly 1 PB of storage. Given the scope of ongoing and upcoming projects, we estimate that an additional 1PB will be needed to accommodate new datasets and catalogues.

Considering both observational and simulation data needs, we establish a baseline storage requirement of 4 PB to ensure the system can support current demands and future growth for the next 5 years.

### Hardware Requirements

Based on the projected data volumes and anticipated processing demands in the coming years, we outline the following hardware as the minimum required to support the initial phase of the CIELOS project:

- **Parallel computing system.** This will be the infrastructure used to run tasks that require high computational power. At this first stage, we propose a system consisting of a cluster of identical servers, each equipped with at least 40 CPUs and 192 GB of RAM, accessible via a separate node from which processing jobs can be launched.
- **Mass storage system.** In this initial phase, we aim to equip the data center with at least 4 PB of storage capacity. This will ensure adequate capacity to host the data generated by the OCAN facilities and simulations.
- **GPU system.** This system will allow us to optimize the use of Artificial Intelligence techniques for the scientific exploitation of the data hosted within the data center. For this phase, we plan to acquire several GPUs specifically designed for this purpose.

This infrastructure will be installed at the SVO premises.

In addition to that, an archival-grade storage media, with at least 4PB, will be installed at the IAC premises to ensure the security of all data (Levels 0, 1, 2, and 3) generated in CIELOS.

### Human Resources

CIELOS will rely on a dedicated support team to carry out and oversee the activities outlined in the preceding sections. We envision an organizational structure in which each participating project, telescope, or instrument is assigned at least one staff member—appointed by the respective institution—responsible for integrating their data into the CIELOS infrastructure and for developing and maintaining the basic data reduction pipelines needed to generate high-level data products.

SVO activities will focus on (i) the VO-compliance of the data products generated by the OCAN facilities and simulations integrated into CIELOS, and (ii) the development of the archive system (database and data access systems), (iii) the long-term preservation of the archives and services at SVO premises. These tasks require at least 3 FTE.

The IAC will contribute to CIELOS through its SIE IT services. The current SIE team, composed of five support staff members, will dedicate 20% of their time to CIELOS-related activities. Their main responsibilities will include supporting OCAN institutions in the development of data reduction pipelines and leading the production of high-level data products (Levels 1, 2, and 3) using the HPC resources located at the SVO facilities. They will also be responsible for maintaining the long-term data archive hosted at the IAC. In addition, the IAC's SI IT services will oversee the maintenance and support of all hardware infrastructure located on-site.

## COMMUNICATION

### Training

CIELOS will organize training workshops and hands-on sessions to help users become familiar with the contents and services offered by CIELOS, following a similar approach to the user engagement and capacity-building activities carried out through the [SVO Schools](#). These sessions would provide practical guidance, demonstrations, and interactive exercises aimed at enhancing users' understanding and effective use of CIELOS resources.

### Newsletter

CIELOS will maintain a regular newsletter to disseminate updates about data releases, infrastructure developments, research highlights, and upcoming workshops or community events. The newsletter will be distributed electronically to all participating institutions, project collaborators, and registered users. It will serve as a critical communication tool to foster engagement and keep stakeholders informed.

### Social Media

To increase outreach and engagement with the broader scientific and public community, CIELOS will maintain an active presence on social media platforms. Social media will be used to share real-time updates, promote new data access tools, announce open science initiatives, and highlight collaborative projects. This will enhance the visibility of CIELOS activities and foster community interaction. We will seek help from the IAC's UC3 office to assist in this activity.

### Conferences

CIELOS will actively participate in the Astronomical Data Analysis Software and Systems (ADASS) conferences (<https://adass.org/>) to present advancements in data infrastructure, software tools, and data management strategies. Engagement with the ADASS community will ensure that CIELOS remains aligned with emerging standards, fosters collaboration with other international observatories, and shares best practices in astronomical data handling.