PROCEEDINGS OF SCIENCE

PoS(ICRC2023)682

The Data Processing System of the ASTRI Mini-Array Project

S. Lombardi,[a,b,*] **L. A. Antonelli,**[a] **C. Bigongiari,**[a,b] **M. Cardillo,**[c] **E. Fedorova,**[a] **S. Gallozzi,**[a] **F. Lucarelli,**[a,b] **M. Mastropietro,**[a] **F. G. Saturni,**[a,b] **F. Visconti,**[a,b] **A. Bulgarelli,**[d] **M. Capalbi,**[e] **O. Catalano,**[e] **A. A. Compagnino,**[e] **V. Conforti,**[d] **S. Crestan,**[f] **G. Cusumano,**[e] **A. D'Aì,**[e] **S. Germani,**[g] **A. Giuliani,**[f] **S. Iovenitti,**[h] **A. La Barbera,**[e] **V. La Parola,**[e] **M. C. Maccarone,**[e] **T. Mineo,**[e] **D. Mollica,**[e] **A. Pagliaro,**[e] **N. Parmiggiani,**[d] **M. Perri,**[a,b] **F. Pintore,**[e] **S. Scuderi,**[f] **G. Tosti,**[g] **A. Tutone,**[a] **S. Vercellone,**[h] **L. Zampieri,**[i] **and G. Pareschi**[j] **for the ASTRI Project**[l]

[a] *INAF, Osservatorio Astronomico di Roma, Via Frascati 33, I-00078 Monte Porzio Catone (Roma), Italy*
[b] *ASI, Space Science Data Center, Via del Politecnico s.n.c., I-00133 Roma, Italy*
[c] *INAF, Istituto di Astrofisica e Planetologia Spaziale, Via Fosso del Cavaliere 100, I-00133 Roma, Italy*
[d] *INAF, Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, I-40129, Bologna, Italy*
[e] *INAF, Istituto di Astrofisica Spaziale e Fisica Cosmica, Via Ugo la Malfa 153, I-90146 Palermo, Italy*
[f] *INAF, Istituto di Astrofisica Spaziale e Fisica Cosmica, Via Alfonso Corti 12, I-20133 Milano, Italy*
[g] *Dipartimento di Fisica, Università degli Studi di Perugia, I-06123 Perugia, Italy*
[h] *INAF, Osservatorio Astronomico di Brera, Sede di Merate, Via Emilio Bianchi 46, I-23807 Merate (LC), Italy*
[i] *INAF, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[j] *INAF, Osservatorio Astronomico di Brera, Via Brera 28, I-20121 Milano, Italy*
[l] http://www.astri.inaf.it/en/library/

E-mail: saverio.lombardi@inaf.it

---

[*] Speaker

The ASTRI Mini-Array is an international project to build and operate an array of nine small-sized (4-m diameter) Imaging Atmospheric Cherenkov Telescopes (IACTs) at the Observatorio del Teide (Tenerife, Spain). The array is designed to perform deep galactic and extragalactic gamma-ray sky observations in the 1–200 TeV energy band. As part of the overall software system, the ASTRI team is developing a dedicated *Data Processing System* consisting of specific software components that implement all the functionalities needed to achieve the final science products from raw data. The main software components are the *Cherenkov Camera Pre-Processing* (for offline conversion from binary to FITS data format of Cherenkov camera raw data), the *Stereo Event Builder* (for offline stereo array trigger of Cherenkov events), the *Cherenkov Data Pipeline* (for data reduction and analysis), and the *Calibration Software* (for all required calibrations). An additional components for *Stellar Intensity Interferometry* data analysis is also envisaged. Thanks to the high-speed network connection available between Tenerife and Italy and the low data volume produced by the innovative ASTRI cameras, raw data files acquired onsite will be delivered to the ASTRI Data Center in Rome immediately after acquisition. At the Data Center, the Data Processing System will automatically perform the pre-processing, software stereo trigger, calibration, reduction, and analysis of the raw data up to the generation of science-ready data (event-lists and Instrument Response Function, to be delivered to the science community) as well as of automated science products. In this contribution we present the main features and components of the ASTRI Data Processing System and report on the status of its development and implementation.

The 38th International Cosmic Ray Conference (ICRC2023)
26 July – 3 August, 2023
Nagoya, Japan

1. Introduction

The ASTRI Mini-Array [1] is a ground-based facility of nine innovative Imaging Atmospheric Cherenkov Telescopes (IACTs) of the small-sized telescope class [2] in dual-mirror configuration. The system is under construction at the *Observatorio del Teide* (Tenerife, Spain) [3], based on a host agreement between the Italian *Istituto Nazionale di Astrofisica* (INAF) and the Spanish *Instituto de Astrofisica de Canarias* (IAC). The completion of the full array and the start of regular scientific operations are foreseen in a few years.

The ASTRI Mini-Array is designed to study in detail sources emitting at very high-energy (VHE) in the 1–200 TeV energy band, with an angular resolution of ~3 arcmin and an energy resolution of ~10% above ~10 TeV [4]. Over the first four years of observations, the array will focus on specific science topics, following a core science program [5]. The wide field of view (~10° in diameter) will enable the simultaneous investigation of multiple sources during the same pointing, making it possible to perform deep observations of complex galactic regions such as the Cygnus region or the Galactic Center. Besides the core science program, a portion of the observation time will be dedicated to the study of selected galactic [6–8] and extragalactic targets [9]. Follow-up observations of selected transient and multi-messenger astrophysics phenomena, such as GRBs, GWs, and very high-energy neutrinos, are also planned [10].

The ASTRI Mini-Array telescopes are an evolution of the ASTRI-Horn telescope [11, 12], currently operational at the INAF "M.C. Fracastoro" observing station in Serra La Nave (Mt. Etna, Italy). The first telescope, ASTRI-1, has been deployed on site in June 2022. Two more telescopes, ASTRI-8 and ASTRI-9, will be installed by spring 2024 [3]. These first three telescopes will start taking data during 2024, thus allowing us to start the verification and validation phase of the stereoscopic system and providing the first scientific observations of astrophysical targets.

In addition to the significant work being carried out to develop, verify, and deploy all hardware and control software components at the Teide site, the ASTRI Team is also working on the development of software tools necessary for pre-processing, reducing, and analyzing the data collected with the ASTRI Mini-Array [13]. These software tools are mainly part of the *Data Processing System* (DPS). In the following sections, we provide an overview of this system, describe its main features and components, and report on the status of its development and implementation.

2. The Data Processing System

The ASTRI Mini-Array *Data Processing System* (DPS) is a collection of software components that are in charge of preparing, calibrating, reducing, and analysing the raw Cherenkov data (or data level zero, DL0) and auxiliary data acquired during the observations up to the generation of high-level science-ready data products (DL3) and automated science products (DL4/5). The DPS system is also in charge of providing suitable data check and calibration products, as well as performing the reduction of the Stellar Intensity Interferometry Instrument (SI3) data.

The main software components of the DPS are:

- *Cherenkov Camera Pre-processing* (CCPP) [14]: it performs the offline conversion of the binary raw data acquired by the Cherenkov cameras of the ASTRI Mini-Array telescopes (DL0-RAW) to FITS [15] data format (DL0-FITS); besides the main quantities recorded by the cameras of each telescope, the system will also populate the primary FITS header of the DL0 files with all the main metadata and information related to the executed Observation Run[1];

- *Stereo Event Builder* (SEB) [16]: it performs the offline software stereo array trigger of the Cherenkov events recorded by the array telescopes; the inputs of the SEB system are the single-telescopes DL0 data in FITS data format; the stereo information of the events are stored in a separate FITS file (SEB0) that shall be archived and ready for use;
- *Cherenkov Data Pipeline* (CDP) [17]: it is in charge of the calibration, reconstruction, selection, and automated scientific analysis of Cherenkov data; the full scientific data reduction and analysis (from DL0 to DL5) is obtained through the `A-SciSoft` software package [18]; the automated generation of standard science products (DL4/5) is achieved with dedicated ASTRI Science Analysis Tools (included in `A-SciSoft`) and/or publicly available ones, like `Gammapy` [19] and `ctools` [20];
- *Calibration Software* (CS) [21]: it is a collection of software libraries and tools that implements all the algorithms needed to perform the calibration and monitoring of the ASTRI Mini-Array devices, at both single-telescope and array level, and to extract suitable calibration factors for the processing of scientific data; its main sub-components are: *i) Optics Calibration Software*; *ii) Cherenkov Camera Calibration Software*; *iii) Pointing Calibration Software*; *iv) Optical Throughput Evaluation/Monitoring Software*;
- *Intensity Interferometry Data Pipeline* (IIDP): it performs reconstruction and scientific analysis of SI3 data; more details can be found in [22];

For detailed pieces of information on the ASTRI Mini-Array DPS *Stereo Event Builder*, *Cherenkov Data Pipeline*, and *Calibration Software* components, we refer the reader to the companion contributions [16], [17], and [21], respectively.

2.1 Cherenkov data flow

Figure 1 depicts a schematic view of the functional design of the end-to-end Cherenkov data processing chain executed by the ASTRI Mini-Array DPS.

Thanks to the high-speed network connection available between Tenerife and Italy (10 Gbit/s) and the low data volume ($\lesssim$1 TB/night/telescope) produced by the innovative ASTRI cameras, the raw Cherenkov data files acquired by each telescope of the array are transferred immediately after acquisition to the offsite ASTRI Data Center (in Rome, Italy) and archived, along with other ancillary data [13]. A preliminary step for the proper execution of the full end-to-end data processing chain is the execution of the CCPP. This software component converts single-telescope scientific and ancillary camera raw data from binary format into the FITS standard format, adopted as common input/output (I/O) data format for each DPS subsystem. In the following step, the SEB software component performs the offline software trigger for the identification of stereoscopic Cherenkov

---

[1] An Observation Run is defined as the minimum block of array data taking that provides a self-consistent set of raw scientific data reducible up to DL5. The typical duration of an Observation Run is on the order of a few tens of minutes.

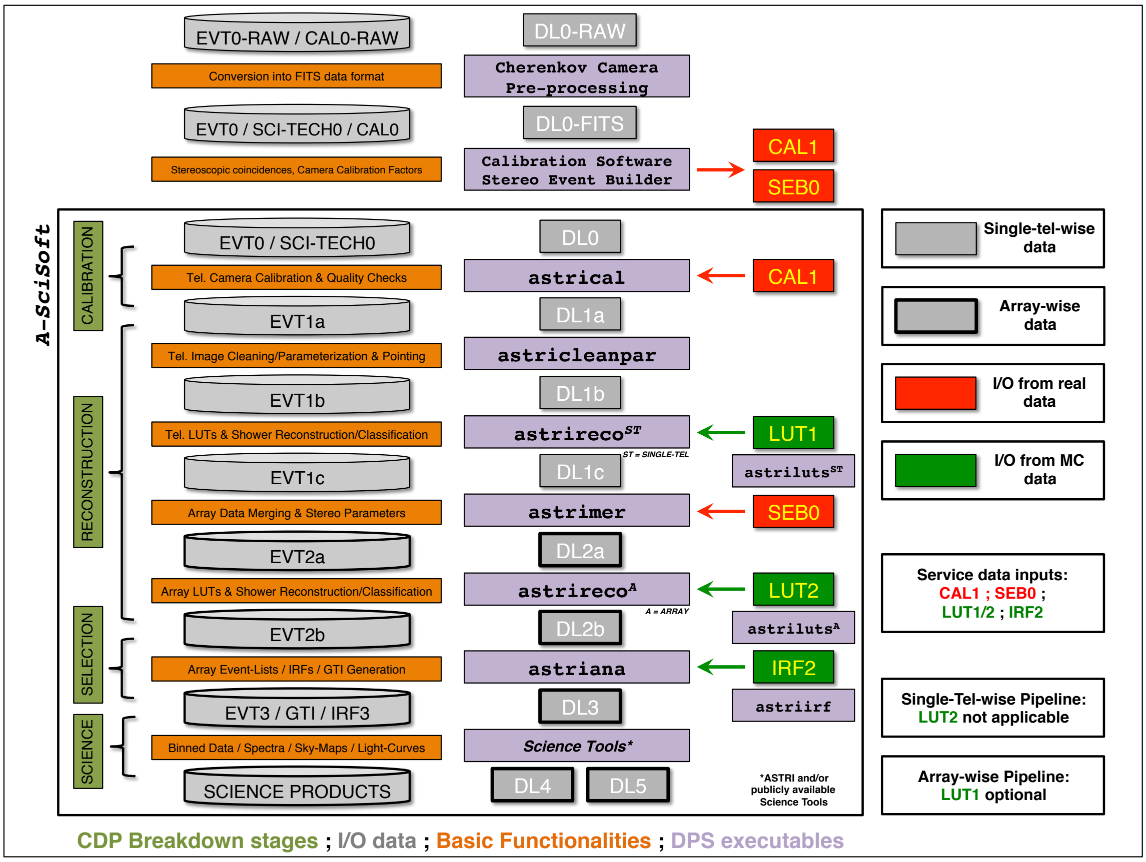

Figure 1: Schematic view of the end-to-end Cherenkov data processing chain executed by the ASTRI Mini-Array DPS, starting from the pre-processing stage down to the generation of high-level science-ready data (DL3) and automated science products (DL4/5). From left to right, functional CDP breakdown stages are shown as olive green boxes, I/O data as grey cylinders, basic software functionalities as light maroon boxes, data levels as grey boxes (thin: telescope-wise data; thick: array-wise data), and data processing executables as light purple boxes. CDP service data inputs, i.e. CAL1, SEB0, LUT1/2 and IRF2, are obtained by running the CS, SEB, and CDP (MC data pipeline) software tools, respectively.

events induced by the same extensive air shower on more than one array telescope. The next steps of the data processing are performed by the CDP software component, through the use of the `A-SciSoft` software package [18]. In this step, data are calibrated, cleaned and reconstructed separately for each telescope. The calibration factors needed for the scientific data reduction are achieved with dedicated CS software tools. Successively, the data coming from the different telescopes are merged and fully reconstructed (i.e. the parameters for gamma/hadron separation, arrival direction estimation, and energy reconstruction are made available for each Cerenkov event), taking advantage of the pieces of information extracted with the SEB software tools and suitable reconstruction array-wise Look-Up-Tables (LUTs) extracted from dedicated Monte Carlo (MC) simulations [23]. The fully reconstructed Cherenkov events are then selected through the application of optimized analysis cuts in order to get the final reduced event lists [24]. These, together with the observation-related Instrument Response Functions (IRFs) and data-filtering tables (good time intervals, GTI) constitute the so-called high-level science-ready data (DL3) that will be delivered to the Science Users. An automated scientific analysis pipeline (based on dedicated `A-SciSoft` science tools and/or publicly available ones, like `Gammapy` [19] and `ctools` [20]) is also envisaged in order to provide the Science Users with automated science products (DL4/5) for reference.

Two types of end-to-end scientific data processing levels are foreseen at the ASTRI Data Center: a *short-term* standard analysis, to be automatically run at the end of each Observation Run, as soon as the scientific raw data files are transferred to the offsite ASTRI Data Center, in order to allow a scientific quick look of the on-going observations; a *long-term* standard analysis, which shall produce consolidated high-level science-ready data and reduced IRFs for final scientific analysis and publication of results.

2.2 Software framework

For the DPS development we have adopted the best practices of software development established at the level of ASTRI SW Engineering Team [25]: *agile* software development method, minimal external dependencies, continuous integration (CI), testing, deployment and software quality assurance. We use standard `C++` for computing intensive algorithms and `Python` for event related computations and pipelining of DPS executables. Concerning the CDP, `A-SciSoft` depends essentially on four libraries included in the `HEASoft`[2] astronomical software package: `cfitsio`, `CCfits`, `hdutils`, and `hoops/ape`, needed for I/O, Calibration Database (CALDB) interface, and parameter system handling, respectively. The SEB software makes also use of these libraries. Standard `CCfits` usage caused severe data bloating and sub-optimal performance due to data duplication in internal `CCfits` data containers. For this reason, I/O critical parts of the software required the application of `cfitsio`'s low-level `fits_*_tblbytes` routines, obtaining less than half of the memory footprint when reading and writing files and about ×7 faster I/O with respect to `CCfits` [18]. Machine learning software modules (implemented for the LUTs computation) use the "pydata stack" and the `scikit-learn`[3] implementation of the Random Forest (RF) algorithm [26]. More sophisticated reconstruction techniques are already being studied for possible implementation and application [27, 28].

2.3 Development infrastructure

We leverage an INAF hosted `GitLab`[4] instance for CI and deployment of DPS software, and in particular of `A-SciSoft`. Test data and machine learning models and relative training data are managed with `dvc` (“git for data"). A `SonarQube`[5] instance is linked with the CI system statically analyzing code at each commit and producing charts and metrics of code quality and test coverage. We deploy `A-SciSoft` in both `docker`[6] and `singularity`[7] containers for the reproducibility of results and deployment on High Performance Computing (HPC) infrastructure. The compute nodes in the ASTRI Data Center are organized as an HPC cluster, issuing a set of tasks managed by the

---

[2] `http://heasarc.nasa.gov/lheasoft/`
[3] `https://scikit-learn.org/`
[4] `https://about.gitlab.com/`
[5] `https://www.sonarqube.org/`
[6] `https://www.docker.com/`
[7] `https://apptainer.org/`

queuing system `slurm`[8], whereas the automatic task orchestration for the MC data, *short-term*, and *long-term* pipelines is foreseen to be managed with `apache airflow`[9], allowing for an easy task scheduling, integration with the ASTRI Archive System [13] and continuous monitoring for data warnings and possible processing errors.

3. The ASTRI Data Center and the ICT infrastructure

The ASTRI Data Center is the main data processing and archiving offsite infrastructure for the whole ASTRI Mini-Array Project. Core system and services are hosted at the INAF Astronomical Observatory of Rome (INAF-OAR) and are made by 12 nodes with typical hardware configuration: 24-cores INTEL chipsets with 256 GB RAM each; redundant raid storage with 16 slots equipped with 16 TB SAS disks; 2 mirrored SSD disks for OS; redundant power modules. These 12 machines are managed via a virtualized environment (openHPC + Gluster) with I/O speedup by next generation NFS 4 cluster (i.e. Ganesha) building up a redundant 3PB storage cluster. Due to virtualization, the openHPC environment appears as a cluster of 20 virtual machines. For each provided service, the user authentication and authorization framework is realized by a centralized LDAP server with distributed read-only replicas. A document-oriented database cluster is also paired to the ASTRI eco-system as metadata handling a storing utility as well as archiving and logging system. As a cold-storage backup facility, a fiber-channel Oracle based Tape library with about 1 thousand data-slots with cold-store data space up to 15 PB (expandable to hundreds) is also installed in the ASTRI Data Center.

4. Summary and outlook

The deployment of the ASTRI Mini-Array is underway at the Teide Observatory (Tenerife, Spain); the first three telescopes (ASTRI-1, ASTRI-8, ASTRI-9) are expected to begin collecting data by 2024, with completion of the entire array expected within a few years. Given the scope of this commitment, the ASTRI Team is working hard on hardware and software development and implementation. Software development efforts are currently focused, besides control software activities, on providing all the software systems needed to plan, collect, reduce, analyze, disseminate, and exploit the Mini-Array science data. In this context, the ASTRI Mini-Array DPS is in charge of preparing, calibrating, reducing, and analysing the DL0 Cherenkov and auxiliary data up to the generation of high-level science-ready data products (DL3) and automated science products (DL4/5). In this contribution, we have provided a brief description of its main features and components. Preliminary testing of the DPS has already been successfully conducted using real data from the ASTRI-Horn prototype telescope and dedicated Monte Carlo simulations of the ASTRI Mini-Array. The commissioning phase of the first three telescopes will represent the first significant test-bed for the validation of all ASTRI Mini-Array DPS software components.

---

[8] https://slurm.schedmd.com/

[9] https://airflow.apache.org/

Acknowledgments

This work was conducted in the context of the ASTRI Project. We gratefully acknowledge support from the people, agencies, and organisations listed here: http://www.astri.inaf.it/en/library/. We acknowledge financial support from the ASI-INAF agreement n. 2022-14-HH.0. This paper went through the internal ASTRI review process.

References

[1] Scuderi, S. et al., Journal of High Energy Astrophysics 35, 52–68 (2022).

[2] Tagliaferri, G. et al., Proc. of SPIE 12182 (2022).

[3] Giuliani, A. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)892 (2023).

[4] Lombardi, S. et al., Proc. 37th ICRC, Berlin, Germany, PoS(ICRC2021)884 (2021).

[5] Vercellone, S. et al., Journal of High Energy Astrophysics 35, 1–42 (2022).

[6] D'Aì, A. et al., Journal of High Energy Astrophysics 35, 139–175 (2022).

[7] Giuliani, A. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)723 (2023).

[8] Tutone, A. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)965 (2023).

[9] Saturni, F. G. et al., Journal of High Energy Astrophysics 35, 91–111 (2022).

[10] Carosi, A. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)785 (2023).

[11] Lombardi, S. et al., A&A, 634, A22 (2020).

[12] Leto, G. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)729 (2023).

[13] Lombardi, S. et al., Proc. of SPIE 12189, 121890P (2022).

[14] Conforti, V. et al., Proc. of SPIE 12189 (2022).

[15] Pence, W. D. et al., A&A 524, A42, 2010.

[16] Germani, S. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)597 (2023).

[17] Mastropietro, M. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)765 (2023).

[18] Lombardi, S. et al., Proc. of SPIE 10707 (2018).

[19] Deil, C. et al., Proc. 35th ICRC, Busan, South Korea, PoS(ICRC2017)766 (2017).

[20] Knödlseder, J. et al., A&A, 593, A1 (2016).

[21] Mineo, T. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)733 (2023).

[22] Zampieri, L. et al., Proc. of SPIE 12189 (2022).

[23] Saturni, F. G. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)719 (2023).

[24] Pintore, F. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)722 (2023).

[25] Bulgarelli, A. et al., Proc. of SPIE 12189 (2022).

[26] Breiman, L., Machine Learning, 45(1), 5-32 (2001).

[27] Lombardi, S. et al., Proc. 38th ICRC, Nagoya, Japan, PoS(ICRC2023)713 (2023).

[28] Pagliaro, A. et al., Appl. Sci., 13(14), 8172 (2023).