Astronomy ESFRI & Research Infrastructure Cluster

ASTERICS - H2020 - 653477

Real-Time Data Streaming Architecture Survey

ASTERICS GA DELIVERABLE: D3.13

Document identifier:	ASTERICS_D3.13_final
Date:	**25 June 2018**
Work Package:	**OBELICS – WP3**
Lead Partner:	**INAF**
Document Status:	**Report**
Dissemination level:	**PUBLIC**
Document Link:	www.asterics2020.eu/documents/ASTERICS_D3.13_final.pdf

Abstract

ASTERICS will benefit ESFRI projects and other related major research infrastructures, including ESFRI-precursor experiments, enabling interoperability, software re-use and the

use of open standards and software libraries. The development of common solutions is a key point of this process. This survey focuses on the architectures for the real-time data streaming, already applied or envisaged by the ESFRI projects, seeking for synergies and common developments.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/WP	Date
Author(s)	Ciro Bigongiari	INAF/WP3	1 June 2018
Amendments			
Reviewed by	Giuseppe Cimo	ASTRON/WP1	12 June 2018
Approved by	Rob van der Meer	ASTRON/WP1	22 June 2018

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	1 June 2018	First draft	Ciro Bigongiari / INAF

2	12 June 2018	Suggestions for improvement	Giuseppe Cimo / ASTRON
3	22 June 2018	Final version	Ciro Bigongiari / INAF

IV. APPLICATON AREA

This document is a formal deliverable for the GA of the project, applicable to all members of the ASTERICS project, beneficiaries and third parties, as well as its collaborating projects.

V. TERMINOLOGY

AMS	Alpha Magnetic Spectrometer
ANTARES	Astronomy with a Neutrino Telescope and Abyss environmental RESearch
ASTERICS	Astronomy ESFRI & Research Infrastructure Cluster
CASA	Common Astronomy Software Applications
CTA	Cherenkov Telescope Array
CTDS	Casacore Table Data System
DADI	Data Access, Discovery and Interoperability
ELT	European Extremely Large Telescope
EGO	European Gravitational Observatory
ESFRI	European Strategy Forum on Research Infrastructures
ESO	European Southern Observatory
e-VLBI	Electronic Very-Long-Baseline Interferometry
EVN	The European VLBI Network
FITS	Flexible Image Transport System
GW	Gravitational wave
HAWC	High-Altitude Water Cherenkov Observatory
HDF	Hierarchical Data Format
H.E.S.S.	The High Energy Stereoscopic System
IACT	Imaging Atmospheric Cherenkov Telescope

IGWD	Interferometric Gravitational Wave Detector
INAF	Istituto Nazionale di Astrofisica
JIVE	Joint Institute for VLBI in Europe
KM3NeT	Cubic Kilometre Neutrino Telescope
LIGO	Laser Interferometer Gravitational Wave Observatory
LOFAR	The Low Frequency Array
LSST	The Large Synoptic Survey Telescope
MAGIC	Major Atmospheric Gamma-Ray Imaging Cherenkov
OBELICS	OBservatory E-environments LInked by common ChallengeS
SKA	The Square Kilometre Array
UCM	Universidad Complutense de Madrid
VO	Virtual Observatory

A complete project glossary is provided at the following page:
https://www.asterics2020.eu/glossary

VI. EXECUTIVE SUMMARY

ASTERICS is intended to serve research infrastructures identified in the ESFRI Roadmap (CTA, KM3NeT, SKA and ELT) as well as other major international projects, including precursor experiments, in the area astronomy, astrophysics and astroparticle physics. ASTERICS is enabling interoperability between them, encouraging cross-fertilisation and developing joint multi-wavelength/multi-messenger capabilities. To do that, one of the primary targets is to seek commonalities in data handling, data storage, etc. between those projects. This initiative also goes in the direction of the global movement for open science data.

Within ASTERICS, the work package OBELICS enables interoperability and software re-use for the data generation, integration and analysis of the ASTERICS ESFRI and pathfinder facilities. Therefore, one of its main priorities is to establish open standards and software libraries for multi-wavelength and multi-messenger data.

This report will provide an overview of architectures for real-time streaming data, already applied or envisaged by the ASTERICS ESFRI facilities, seeking synergies between experiments to enable the development of common solutions.

The architectures for real-time data streaming adopted or envisaged depend on the kind of data produced by the experiments. It is therefore easier to find synergies to develop common architectural solutions between experiments that produce the same kind of data. For this purpose, experiments have been grouped into three different categories according to the type of data that they produce: image-based, event-based and signal-based experiments.

Table of contents

1. Introduction

Data produced by nearly all kinds of detector used in the area of astrophysics and astroparticle physics usually consist of analog electric signals. Whenever such signals satisfy some pre-selection criteria (trigger conditions) they are digitized by Analog to Digital Converters (ADC) to be further processed and stored in digital form. The output of the ADC is usually read-out by the Data Acquisition system (DAQ) which is the first step in the digital-data streaming. The DAQ can simply format data and save them on some storage system or can perform some preliminary tasks like pre-calibration and pre-selection of the data before storing them.

Whereas many scientific data analyses involve “data at rest”, i.e. data saved in files in some storage system and later analysed, sometimes scientific data take the form of a nearly-continuous stream of information like in the case of continuously running experiments, such as LHC experiments, or automated observatories such as radio telescopes. In some cases, even if the experiments do not run continuously, the volume and rate of generation is so large that keeping the raw data at rest would be unfeasible and therefore data must be immediately processed into a reduced stream. Considering that the granularity (number of camera pixels, number of electronic channels) of new generation detectors is steadily growing as well as their acquisition rate the demand for a real-time or nearly-real-time data streaming is more and more widespread. Moreover, many fields of astronomy, astrophysics and astroparticle physics are characterized by transient phenomena whose duration can be as short as few seconds. The physics behind such phenomena can be investigated with more efficiency if they can be observed by many different detectors at the same time. It is therefore essential to analyse data in the shortest possible time to alert the community and trigger follow-up observations. Data acquired by the DAQ system are therefore temporarily stored on a volatile memory or directly pipelined to a simplified analysis chain which quickly extracts all the relevant information to select events that are worthwhile for issuing an alert. Such real-time or nearly real-time analysis chains are often staged in different levels of accuracy and complexity connected by fast links.

The projects included in this survey are identified in section 2. In section 3, the experiments are grouped in three categories according to the type of data that they produce. Sections 4 to 6 describe the real-time data streaming architectures that are in use or envisaged for the experiments belonging to each group. The possible synergies between experiments are discussed in section 7. Conclusions are given in section 8.

2. ESFRI projects and related pathfinders

This survey of the architectures for real-time data streaming encompasses 13 projects identified as ASTERICS stakeholders. They include ESFRI projects, their pathfinders and other linked projects, most of them being ASTERICS-OBELICS partners. The table shows the field of application of each experiment and the type of data produced as categorized in section 3.

Project	ESFRI	Field	Type of data
CTA	Yes	Cherenkov observatories for gamma-ray astronomy	Events
H.E.S.S.	Pathfinder for CTA	Cherenkov telescope array for gamma-ray astronomy	Events
MAGIC	Pathfinder for CTA	Cherenkov telescope array for gamma-ray astronomy	Events
KM3NeT	Yes	Neutrino telescope	Events
IceCube	Pathfinder for KM3NeT	Neutrino telescope	Events
ANTARES	Pathfinder for KM3NeT	Neutrino telescope	Events
ELT	Yes	Ground-based optical/near-infrared telescope	Images
LSST	No	Optical telescope	Images
EUCLID	No	Satellite mission to map the dark Universe	Images
SKA	Yes	Radio telescope arrays	Signals
e-EVN	Pathfinder for SKA	e-VLBI network for radio astronomy	Signals
LOFAR	Pathfinder for SKA	Radio interferometric array	Signals
Advanced LIGO	No	Gravitational wave detectors	Signals
Advanced Virgo	No	Gravitational wave detector	Signals

Table 1: Projects included in this data format survey. Observatories are classified according to the type of data that they produce.

3. Types of data produced by experiments

In this survey, experiments are classified according to the type of data that they produce. Although final science products are similar for all the astro(particle)physics experiments (e.g., skymaps, catalogues, spectra), raw and processed data below the science data level have fundamental differences depending on the experimental technique. As an example, imaging atmospheric Cherenkov telescopes (IACT) for gamma-ray astronomy differ fundamentally from visible-light astronomical ones. In fact the formers detect particles one by one, the so-called “events”, whereas optical telescopes collect a large number of photons to form images. Moreover, both detection techniques are very different to those employed by radio interferometers or by gravitational wave detectors, which record signals in the time (or frequency) domain to be subsequently processed. On this basis, three groups of experiments are identified: image-based, event-based and signal-based experiments.

The above types of data have different requirements for formatting and processing. This also applies to a significant part of the metadata needed to find the stored data and to link them to calibration and other auxiliary instrument data. Synergies between experiments producing similar type of data are more natural and easier to identify. Therefore, software reuse and common standards are expected to be attained to much more extent in experiments belonging to the same group. In the next three sections, these groups of experiments are reviewed and their architectures for real-time data streaming (both in use and envisaged) are discussed.

It should be noted, however, that these groups are not completely bounded and often overlap each other. This leaves open the possibility to look for common solutions for a wider range of experiments.

4. Image-based experiments

Optical/near-infrared telescopes essentially produce images. The three projects included in table 1 that belong to this group are ELT [1], LSST [2] and Euclid [3], which are under construction or in the design phase. ELT and LSST are ground-based telescopes, whereas Euclid is a satellite mission.

4.1 LSST

The rapid cadence and scale of the LSST observing program will produce approximately 15 TB of raw imaging data and millions of alerts every night [4]. Data collected by the LSST camera and telescope will be automatically processed to *data products* – catalogs, alerts, and reduced images – by the LSST Data Management (DM) system [58]. These products are designed to be sufficient to enable a large majority of LSST science cases, without the need to work directly with the raw pixels. Two major categories of data products will be produced and delivered by the DM system: the level 1 data products, designed to support the discovery, characterization, and rapid follow-up of time-dependent phenomena (“transient science”). The alerts to such events will be published within 60 seconds of observation. Level 2 data products are designed instead to enable systematics and flux-limited science and will be made available in annual data releases. The LSST data management system is architected in three layers: an infrastructure layer consisting of the computing, storage, and networking hardware and system software; a middleware layer, which handles distributed processing, data access, the user interface, and system operations services; and an applications layer, which includes the data pipelines and products and the science data archives. The applications layer is organized around the data products being produced. The nightly pipelines are based on image subtraction, a process that highlights differences between two exposures of the same field and are designed to rapidly detect interesting transient events in the image stream. The data stream from the camera will be pipeline processed and continuously updated in real time, with a transient alert triggered within 60 seconds of completing an image readout. All of these pipelines are architected to operate on very small and medium sized platforms as well as make efficient use of Linux clusters with thousands of nodes.

4.2 ELT

The different tasks supporting the operations of ESO telescopes in Chile like ELT are carried out at widely separated locations, with some of them taking place at the observatory site and many others in Europe at ESO's headquarters. A fundamental requirement to make such an operations scheme viable is a stable way of transferring the large amounts of data generated by the telescopes to Europe on the shortest possible timescale. One of the ELT requirements is that the raw scientific data, including calibrations, shall arrive at the ESO Science Archive Facility not later than 1 h after they have been acquired [5]. The average foreseen nightly data production is at the level of 1–2 TB (uncompressed) per night, with large variations depending on the instruments and modes actually used on a given night. A rough estimate of the data transfer requirements of the is thus 0.5–1 TB/night (compressed). To achieve such result a substantial improvement in the connectivity between the Paranal and Cerro Armazones Observatories in Chile and ESO Headquarters in Germany has been needed. The completion in 2010 of the FP7-funded EVALSO (Enabling Virtual Access to Latin-American Southern Observatories) project [57] and its integration into the operational information and communications technology of the observatory has provided high-speed wide-bandwidth connections to the observatories suitable for such task.

The adopted data transfer system consists of four multi-threaded Java applications: the Archive Data Organizer (ADO), which classifies incoming data and determines the transfer method, the compression method and the transfer priority; the Network Transfer Scheduler (NTS), which is in charge of transferring incoming data according to their priority, and of maximizing the transfer throughput; the Disk Transfer Scheduler (DTS), which copies to an external disk to be shipped data that cannot be transferred via network; the Archive Replicator (AR), which receives incoming data through a staging area and copies them into the ESO Scientific Archive.

4.3 EUCLID

The Euclid mission will deliver a lower data rate but still unprecedented for a space mission: about 110 GB of compressed data per day, where the daily tele-commanding and communications period will be 4 hours [3, 6]. Onboard data processing is required to reduce the data stream generated by the 4 Megapixel detectors by a factor over 100, since it is impossible to deliver to the ground all the raw detector data. The large number of scientific exposures and high-resolution images generate a high science data volume and require large on-board memory capable of hosting the 850 Gbit of daily generated. The on-board Mass Memory Unit (MMU) has a capacity of 4Tbit EoL sufficient to store 72 hrs of scientific data and 20 days of spacecraft housekeeping. The instruments deliver high volume scientific data via high speed SpaceWire links directly into the mass memory. The transfer layer protocol of the science bus is based on a cyclical communication frame at 60 Hz, maximizing the efficiency of data transfer per communication frame. Files stored in the mass memory are downloaded using the standard CCSDS File Delivery Protocol (CFDP) using the reliable transfer with acknowledges for the downlink and the simple unreliable transfer for uplink. Both the X and K band communication link can be used for the file transfer.

Due to the large number of deployed detectors and to the limited satellite telemetry available to ground, a consistent part of the data processing, conventionally performed off-line, will be accomplished on board, in parallel with the flow of data acquisitions. This has led to the development of a specific on-board, HW/SW, data processing pipeline, and to the design of computationally performing control electronics, suited to cope with the time constraints of the NISP acquisition sequences during the sky survey. The on-board pipe-line allows all the basic operations of input frame averaging, final frame interpolation and data-volume compression before ground down-link

5. Event-based experiments

Event data is produced by IACT gamma-ray observatories (i.e., CTA [13], H.E.S.S. [14] and MAGIC [15]) as well as by neutrino observatories (i.e., KM3NeT [16], IceCube [17] and ANTARES [18]). Although not included in table 1, large-aperture gamma-ray observatories (e.g., HAWC [19] and Fermi [20]) and cosmic-ray experiments (e.g., AMS-02 [21] and the Pierre Auger Observatory [22]) also belong to this group. In this context, “event data” refers to all the recorded data associated to a given cosmic particle detected by the observatory.

IACT gamma-ray observatories consist of an arrangement of several telescopes, where each telescope camera has hundreds or thousands of pixels. Raw data include the time series of the charge registered by every pixel of the triggered telescope(s) for the event as well as all the calibration, atmospheric monitoring and ancillary data needed to “build” the event. For a typical trigger rate of several kHz per telescope, this involves an enormously high data rate. Whereas H.E.S.S. and MAGIC have few Cherenkov telescopes, CTA will operate two arrays with tens of telescopes each. In particular the foreseen southern array will comprise 115 telescopes which translates into a raw data rate of around 10 GB/s [23].

5.1 H.E.S.S.

The main responsibility of the H.E.S.S. DAQ system is the operation, i.e. read out and control, of the five Cherenkov Telescopes (CTs) but it is also used for monitoring auxiliary sensors, error handling and user interaction with the array. It is a multi-machine, multi-process and multi-core system and consists of approximately 230 processes. The data rates of the Cherenkov cameras peak at 46MB/s for the primary scientific data during routine operation. To be able to cope with data rate bursts, for example due to short time-scale transient events or to read out other equipment on-site, the required maximum data rate is of the order of 80MB/s. The server farm uses for that purpose a custom-made round-robin load-balancing scheme. In this scheme, all data from all telescopes are sent to one of the nodes in the server room for four seconds, and buffered in memory. After four seconds, the data-receiving node is switched, and the Central Trigger sends the IP address of the new data-receiving node to all cameras. The node that received the data beforehand then starts the event building process and converts the raw Cherenkov data byte-stream to the common H.E.S.S. data format. Data acquired by the H.E.S.S. DAQ system are concurrently checked and analysed by a preliminary analysis. Due to the common data format adopted for online and offline analyses the same standard offline analysis software can be used for this purpose too. It actually performs a full analysis, based on the HAP TMVA multivariate data analysis toolkit [32], of the data being taken by the telescopes running in real time. The only limitation with respect to the offline analysis is that only a pre-defined camera calibration can be used online, which leads to a slightly worse background rejection. Run results are stored on disk and used to perform a near real-time analysis with data from consecutive runs on the same target. The output of the real-time pipeline are calibrated camera images with intensities in photo electrons as well as their parametrization and classification. The output of the near real-time analysis are significance maps of the region of the sky that is currently being observed and detection plots with respect to the target source position. To cope with the high-data rates the real-time pipeline is split into several different processes. Each CameraReader process has a corresponding Analyser process which subscribes to the data stream that is processed by the CameraReader. The Analyser processes all events that are generated by the CameraReader. This includes event calibration and event reconstruction as well as gamma-hadron separation. The processed data is collected by AnalysisServer processes, one for each SubArray in use. The input maps for the significance maps are filled at the Analyser and the final significance maps are created at the AnalysisServer. The latter is a time-consuming process (several minutes) and therefore happens in parallel to the input maps being received.

5.2 MAGIC

Whenever the MAGIC telescope cameras simultaneously register an image which fulfils the single telescope trigger condition, a signal is sent to the data acquisition system (DAQ) in order to register all the information recorded by the cameras and by ancillary systems. Digitalized data flow from the FADC system to the receiver boards installed in dedicated computers which pre-process them and finally store them on a SATA raid disk system (fiber channel linked) for further offline processing and archiving. The MAGIC DAQ software is a multithreaded program written in C/C++ language exploiting the *pthread* POSIX library for parallel processing. Two identical copies (except for the configuration files) of this software run on two dedicated server computers, each one appointed to the data acquisition of one of the MAGIC telescopes.

The DAQ program performs many subsequent tasks (data packet collection, event building, basic pixel-wise data analysis, integrity checks, data storing) on every event. Each of these tasks is done in parallel thanks to the multithread architecture of the program. To easily handle the data in parallel the events are temporary copied on a ring buffer structure 4000 events deep, accessible synchronously by the running processes. The MAGIC-II raw data files have a typical size of 2 GByte while the data volume nightly generated might exceed 1 TByte. The DAQ program is installed on a powerful 4-CPU 3.0 GHz server with 4 GByte RAM memory running Scientific Linux. A sustainable trigger rate of 1kHz can be achieved corresponding to a 200 MByte/s data storage rate.

MOLA (MAGIC Online Analysis) is a multithread C++ program used to obtain on-the-fly estimate of the gamma-ray flux from sources in the field of view of the telescope during MAGIC observations. The program runs simultaneously with the data acquisition software and acts as receiving client of the event information computed at the very moment the events are acquired by each telescope, MAGIC-I and MAGIC-II. The multithread program structure consists of three threads: two *reading* threads and one *analyzing* thread. The two reading threads are appointed to receive the data stream from the two DAQs asynchronously and perform the non-stereo analysis steps. The main analysis thread is instead appointed to match the events from the two streams and perform the stereoscopic reconstruction. Online analysis up to 600 Hz acquisition rate is possible, with no event loss in the data transfer. The MOLA online analysis program allows an immediate feedback in case of a rapid gamma-ray flare of an observed astrophysical source.

5.3 CTA

The ACTL (array control and data acquisition) system envisaged for CTA which will consist of the hardware and software that is necessary to control and monitor the CTA arrays, as well as to time-stamp, read-out, filter and store the scientific data. The data streams generated by the cameras after local camera triggers will dominate the overall data rate of CTA, with an estimated raw data rate of about 80 GB/s for the southern array. Along the way, inter-telescope and array-level trigger schemes and possibly an on-line filtering of events will be used to further suppress the background while selecting electromagnetic showers with high efficiency. The DAQ will comprise the camera readout, the buffering of the read-out data, the processing of array trigger decisions, the building of camera-dependent events and filtering of interesting events to reduce the overall data volume (about 20 PB/y). To cope with the very large data rate coming from many telescopes with different characteristics the DAQ need to be efficient, modular and robust. The use of modern middleware, namely ZeroMQ and Protocol Buffers, can help to achieve these goals while keeping the development effort to a reasonable level. Protocol Buffers are used as an on-line data format, while ZeroMQ is employed to communicate between processes. Protocol Buffers from Google are a way to define high-level data structures through an interface description language (IDL) and a meta-compiler while ZeroMQ is a middleware that augments the capabilities of TCP/IP sockets.

A fast reaction to unexpected transient γ-ray events is a crucial part of the CTA observatory, to trigger follow-up observations of astrophysical transients and better understand the origin of their emission. To capture these phenomena during their evolution and for effective communication to the astrophysical community, the speed is crucial and requires a system with a reliable automated trigger that can issue alerts immediately upon detection of γ-ray flares. This will be accomplished by means of a Real-Time-Analysis (RTA) pipeline. The CTA design imposes several key requirements to the RTA system: (i) scientific alerts must be generated with a latency of 30s with respect to the triggering event collection; (ii) the search for transient phenomena must be performed on multiple timescales (i.e. using different integration time windows) from seconds to hours, both within a defined source region, and elsewhere in the field of view; (iii) the sensitivity of the analysis must be not worse than the one of the final analysis by more than a factor of 3. Due to the CTA requirement that the system should react with a latency of 30 seconds and due to the limited bandwidth between the telescope sites and the Science Data Centre, the RTA should be performed on-site, i.e. with hardware and software infrastructures co-located with the telescopes;

The On-Site Analysis will have two pipelines: the Level A or Real-Time Analysis, and the Level B analysis. Both pipelines will have a dual purpose of delivering science feedback from the CTA data and monitoring the instrument performance and the data quality. Both pipelines will read raw data, apply calibration algorithms, reconstruct Cherenkov events and produce event lists; the focus of the Level-A Analysis will be to maximize the speed of the analysis and the focus of the Level-B Analysis will be instead to maximize the sensitivity.

5.4 ASTRI

Each ASTRI camera is connected, through a point-to-point fiber Ethernet cable, to a dedicated computer, the ASTRI Camera Server, that runs the Detector DAQ Software (DDS). The camera server is in charge of collecting the bulk data sent by the Camera Back End Electronics (BEE) using the TCP/IP protocol. Each Camera bulk data stream consists of structured packets with various layouts pertaining to three main classes: notification packets, housekeeping packets, and science packets. A suitable packet header identifies the Camera data and allow distinguishing the various packet layouts. The size of each packet depends on the read-out mode. In science mode, the expected over-all stream data rate for the typical event trigger rate of 600 kHz is of the order of 6-7 MB/sec.

The official software package for the ASTRI Data Reconstruction and Analysis is a collection of programs written in C++, using the standard HEASOFT CFITSI/O libraries. It includes all the necessary algorithms to transform the raw data collected during a data-taking run, into information about the physic-s parameters of the observed primary incoming cosmic-ray. The data analysis chain is divided into several tasks, each of them performed by independent programs chained one with another by a pipeline. The data processing starts with the raw data for every triggered event, consisting of binary files containing the full information available per pixel plus camera housekeeping data. The ASTRI Data Reconstruction and Analysis software has been developed using CUDA5 in order to exploit the GPU (Graphics Processing

Units) capabilities. The use of software specifically developed to take advantage of parallel computing architectures can instead reduce the calculation time for the data reduction, calibration, and final analysis by a factor of 10 to 100.

5.5 Antares & KM3NeT

Neutrino observatories comprise thousands of optical sensors attached to vertical "strings" arranged over a large volume. Similarly, to IACT observatories, event data include all the time-amplitude information registered by the optical sensors as well as calibration and other auxiliary data. On the other hand, neutrino observatories operate in continuous mode and have a much larger aperture, which results in a huge raw data volume. All these event data are combined together to derive the energy and direction of the neutrino. The data acquisition systems of ANTARES and KM3NeT, both deployed in the deep waters of the Mediterranean Sea, are based on the "all-data-to-shore" concept, meaning that all signals that pass a certain threshold are sent to shore for real-time processing [24]. In the case of KM3NeT, the infrastructure has a modular design made up by so-called "building blocks", each one constituting a 3-dimensional array of photo sensors producing a total data rate of 3 GB/s [25]. The second phase of KM3NeT is planned to have 3 building blocks, and the full telescope will have 7.

The KM3NeT data acquisition and processing system (DAQ system) is based on the experiences gained with the ANTARES system, operating for 10+ years in a stable and efficient way. The DAQ systems of both experiments have the same conceptual design which is described in the following. The data stream is produced in the detector by dedicated FPGA-based DAQ boards and sent to shore via fibre Ethernet connection; TCP for ANTARES and UDP for KM3NeT. In on-shore Linux computing farms, this data stream is first sorted to align time slices of data from the different detector parts, then filtered for interesting events with flexible, adjustable filtering algorithms, and subsequently stored to disk with a data reduction factor of about 10^5. The data is then transferred to HPC centres for offline analysis.

The communication between the software packages for real-time data collection, data filtering and data processing/analysis is based on the "ControlHost" [KM3NeTRef1] infrastructure. ControlHost provides a high-performance subscription-based messaging layer built on top of TCP/IP and allows for a flexible handling of various real-time data streams, from triggered event data down to the full raw data. Any programme can receive any of these data streams both locally at the shore-station computing centres and remotely, as the streams can be redirected via a TCP/IP tunnel to remote sites. This can be achieved in a non-blocking way, i.e. without interfering with the data acquisition of the experiments.

This flexibility in real-time data streams is exploited by systems for online monitoring of the detector, data quality monitoring and real-time reconstruction of events for alert triggering. At the same time, it allows to react on external alerts, e.g. to store all – unfiltered - raw data in case of an external Supernova or GRB alert.

5.6 IceCube

In IceCube, which is located at the South Pole, a first level of reconstruction and event filtering is done in near real time so that data can be transferred over a satellite link to the data centres located in the Northern Hemisphere [17, 26]. Digitized and timestamped to nanosecond precision data from all the optical modules are used as input to sophisticated reconstruction algorithms that determine the direction, energy, and type of the incident cosmic ray event. In order to achieve this goal, the IceCube data acquisition system merges the digital data streams from each photodetector into a single time-ordered list which is presented to online triggers that determine, in real-time, whether or not a given pattern of hits is noise or signal. At the present time, the data provided to the triggers is limited by the performance of sorting and merging algorithms: the 500 Hz raw event rate from each sensor (2.5 MHz array aggregate rate) is beyond the capability of the central sort and merge. The current solution adopted by the IceCube detector is to impose a hardware-based pre-trigger coincidence on hits emanating from the optical modules which reduces the rate by a factor of 20. In order to suppress even further the background from atmospheric muons, an online filter which uses likelihood reconstruction of muon tracks, is applied. As computing power at the South Pole is limited, computationally demanding reconstructions are only applied to events passing further cuts. After cuts on the likelihood fit quality, the event rate is reduced to 5 Hz. Afterwards, multivariate classifiers are employed to improve the neutrino purity to 90% at an event rate of 5 mHz. Only clusters significant enough to yield an alert were sent to the North immediately through a satellite link, where they are distributed to the follow-up observatories via automatically generated emails.

The system is currently transitioning to transmitting every single track-like event passing a common event selection immediately through the Iridium satellite network. For each of these events, the most important reconstructed quantities, including direction, energy, fit quality and uncertainty estimates, are sent to the North, where multiple analyses run on the events from the stream. The searches for neutrino multiplets will be run in the North, where they can be updated and improved more easily. The median delay between the neutrino interaction in the detector and receiving the event in the North is 22 seconds.

6. Signal-based experiments

In radio interferometric arrays, signals from all the individual telescopes are brought together and processed by the so-called correlator, which combines the signals to form an image of the observed radio source. Among the projects listed in table 1, SKA [39], EVN [40] and LOFAR [41] belong to this category.

6.1 EVN

The European VLBI Network (EVN) is an interferometric array of radio telescopes spread throughout Europe, Russia, South Korea, China, South Africa and Puerto Rico, that conducts unique, high resolution, radio astronomical observations of cosmic radio sources. It is the most sensitive VLBI array in the world, thanks to a number of extremely large telescopes that contribute to the network. The EVN regularly observes in conjunction with the Very Long Baseline Array (VLBA) in the United States and with the Australian Long Baseline Array (LBA), to form a truly global array.

The Joint Institute for VLBI ERIC (JIVE) was founded (under the name "Joint Institute for VLBI in Europe") to host and operate the correlator, the central data processor of the EVN, to provide user support and in general to promote the technique and use of VLBI.

Originally, the EVN network was designed to observe at a bandwidth of 128 MHz, resulting in data flows of up to 1 Gbps per telescope. These data flows were recorded on tape reels which were then physically shipped to the correlator for playback and correlation.

In the early 2000s, the price and capacity of hard disk drives had evolved to such a degree that recording on disk packs became a viable alternative to tape reels. The Mark5 system, developed at MIT Haystack Observatory, basically a Linux computer recording on shippable 8-disk packs, became the de facto standard for several VLBI networks.

The move to Linux-based recorders immediately opened up the possibility to send data via the Internet, and the first attempts to do real-time correlation (also called e-VLBI) started around 2004. Of course, in those days 1 Gbps was more than most local networks could sustain, especially in the case of astronomical observatories that tend to be built in remote locations.

Starting in 2006, through the EC-funded EXPReS and NEXPReS projects, JIVE, the EVN and various partners from the research networking and high-performance computing communities further developed the real-time capability of the network. To this purpose, control software was written that enables data transfers between a wide variety of hardware platforms, at strictly controlled speeds, using various protocols. Mostly used is UDP, for real-time lossy transfers (most VLBI data is noise after all), and UDT, for reliable post-observation transfers. Correlation at JIVE moved from the original batch-mode bespoke hardware correlator to software on a cluster of COTS Linux machines, and recording and playback were optimised to allow high-speed transfers. Local support was sought, and often found, by the participating stations to improve the connectivity to their local NREN, and via the pan-European GEANT network to JIVE.

Several live demonstrations were done to advertise the new capabilities of the EVN. One high-profile demo was the tracking and real-time correlation of one source over a period of more than 24 hours, moving from telescope to telescope around the globe as the source would rise and set, during the opening of the International Year of Astronomy in Paris in 2009.

New digital backends have opened the way to higher bandwidth observing in the EVN. Currently most EVN stations record at 1 to 2 Gbps on so-called FlexBuffs, high-end Linux servers equipped with 34 non-shippable hard disks. Every observation is transferred automatically, at whatever bandwidth is available, to similar storage at JIVE, and correlated when all data has arrived. Note that quite a few stations still have very poor connectivity, and still have to rely on shipping disk packs. This includes the telescopes of the VLBA. Throughout the year, 10 – 20 24-hour real-time observations take place, with the well-connected subset of the EVN telescopes, dubbed the e-EVN. During these e-VLBI sessions, no data is recorded at all, and all data are streamed directly from the backends at the telescopes into the correlator. These sessions are standard at 2 Gbps per telescope, although tests with 4 Gbps bandwidth have been performed.

Through the vastly improved response and turn-around time that e-VLBI provides, the e-EVN has become an instrument capable of following up transient events, like the recent GW detection of the merger of two neutron stars. Transient triggering and multi-messenger astronomy will only become more important in the future, and the e-EVN is very well positioned to play an important role in these new and exciting developments.

6.2 LOFAR

The LOFAR array, consists of about 7000 simple omni-directional antennas organised in stations containing local computing resources to perform beam-forming [42]. All these beam-formed data (about 19 GB/s for the entire array) are sent via a high-speed fibre network to the central processing facility, where they are pre-processed on-line.

The architecture of LOFAR is described in [42] (van Haarlem et al). The real-time data streaming architecture is as follows. The antennas in the stations each produce 12-bit data at a sampling rate of 200 MHz. As each antenna field consists of 48 dual dipole antennas, the raw data rate per antenna field is roughly 230Gbit/s, or for the entire LOFAR (currently ~60 stations) 14 Tbit/s. As it is not feasible to send the data at this rate to a central place, the data of the antennas of every field are combined into one or more 'station beams'. This happens on custom-built FPGA boards ('receiver units' or RCUs). The output of these RCU's is max. 3 Gbit/s per antenna field. This data is sent in real-time, over optical fiber (partly 'dark fiber', partly commercial fiber) to a central processing facility in Groningen, which is designed to receive up to 240 Gbit/s, i.e. max. 80 antenna fields. The central processing facility consists of 8 GPU machines, called Cobalt. The data streams can be up to 500 ms apart, due to difference in distance to the stations. The data is aligned using a ring buffer which can hold up to 3 seconds of data. The next step is to transpose the data from 'all channels for every station' to 'all stations for every channel'. Depending on the mode, the data is then correlated or added (beamformed) on the GPUs, and averaged. The output of Cobalt is written to a global (Lustre) file system over Infiniband, at a total data rate of max. 100 Gbit/s. From this place, the processing is performed offline.

6.3 SKA

The above two projects are recognised as pathfinders for SKA, which will operate both an array of dishes and an array of antennas grouped into stations. The SKA project is designed in two phases and, when both are complete, the observatory will consist of many thousands of connected radio telescopes located in two regions stretching up to thousands of km. For the first phase, the summed data rate is estimated to be about 3 TB/s and the expected archive data volume is 100 PB per year [47]. Data from the receivers, both antennas and dishes, are transported to the Central Signal Processor (CSP), located in a radio frequency-shielded building at the centre of the telescope array by an optical fibre network. Here, data from all receiver or station pairs are combined into visibilities by the CSP correlator. The resulting data are transported to the Science Data Processor (SDP), located several hundreds of kilometres away. The SDP produces science-ready calibrated data products for analysis by the radio astronomer, a task that is highly data intensive and is expected to require computing resources in the 100 PetaFlop range. The incoming data rate is so high that raw data are unlikely to be kept permanently. The temporary storage of raw data will need to be minimised too. This has the implication that data processing and Quality Assessment (QA) will need to be automated with little or no possibility for intervention by operators or scientists. Data from the SDP are distributed to Regional Science Centres for further analysis and dissemination. The large amount of data produced from SKA presents a computational challenge for imaging methods and calls for High Performance Computing (HPC) including hardware accelerators. Furthermore, part of the computations must be done on site, as close as possible to the detectors to decrease the cost of the data transfers. Being located in remote areas, not grid-connected, the computers have to be powered with local production systems. The power consumption is therefore a key aspect. Three technologies are presently investigated for what concerns the hardware accelerator: FPGA (Field-Programmable Gate Array), GPU (Graphics Processing Unit) and the manycore MPPA (Massively Parallel Processor Array). FPGA, known to be one of best energy efficient computation resources, are probably the best option for SKA.

6.4 LIGO & VIRGO

The two Advanced LIGO detectors in the USA [49] and the Advanced Virgo detector in Italy [50] included in table 1 are the first ones of a network of very sensitive interferometric detectors of gravitational waves (GW) working together. These detectors also produce signal-based data, but they are very different to those generated by radio interferometric arrays. The data produced by these experiments are stored in “frames” with thousands of channels, where the GW strain channel only represents a small fraction of data and all the other channels are used for auxiliary instrumental and environmental monitoring. Each interferometer has a data rate of a few tens of MB/s [51, 52].

6.5 LIGO

The LIGO Control and Data system (CDS) features a tightly coupled and highly integrated control and data acquisition system. Control of the interferometers requires many Multiple Input Multiple Output (MIMO) control loops closed both locally and across the 4-kilometer interferometer arm lengths. In addition to providing the closed loop control, the control systems front end processors act as Data Collection Units (DCU) for the data acquisition system. Data collected by these front ends and the data acquisition system must be collected and time stamped to an accuracy of 1 microsecond and made available to on-line analysis tools such as the Global Diagnostics System (GDS). Data is also sent to the LIGO Data Analysis System (LDAS) for long-term storage and off-line analysis. Data rates exceed 5 Mbytes per second per interferometer continuous. Connection between the various front-end processors and the data acquisition system is achieved using fibre optic reflective memory networks. This system employs a number of multi-core processor-based computers to perform real-time control, with connection to PCI Express Input/Output devices via fibre optic links.

6.5 Advanced Virgo

A custom fast-data network has been built in order to exchange and process in real-time the data from the sensing electronics to the driving electronics, and to provide these data to the data-collection system.

The global longitudinal and angular controls of the interferometer (ITF) are performed through synchronous fast digital loops, running at ~20 kHz and ~2 kHz, respectively. The signals are collected by the sensing electronics in different locations of the detector.

They are sent through optical fibers to real-time processing units that run the digital servo-loops and send the data to other processing units or to the driving channels of the different ITF suspensions. A set of multiplexer/demultiplexer boards is used to route the data packets in the network. The data can be delivered from any ITF device to any other ITF device with latencies usually dominated by the optical fiber propagation time. New devices, called DAQ-boxes, will be installed in AdV: these are digital motherboards that can host specific mezzanines (for ADC, DAC, digital demodulation channels, camera triggers, etc.) and manage the interface with the timing and fast-data networks. The same interface is set in the electronics used to control the suspensions. Some controls are performed locally: these are implemented by sending the data directly, on-board, from the sensing part to the driving part, where the needed filtering is performed by the associated DSP.

Most of the sensor signals are digitized by ADC at ∼1 MHz and digitally low-pass filtered and decimated on-board, down to frequencies between a few kHz and a few tens of kHz. The data are then sent to the fast-data network. Some are then used as error signals for feedback loops and others are used to monitor the different subsystems and the detector environment.

The estimated numbers of front-end channels to be acquired or driven through the fast-data network of AdV are of the order of 2200 ADC with sampling frequencies higher than 1 kHz, 1400 DAC, 100 digital demodulation inputs and 40 cameras synchronized with the timing system. The goal of the data collection is to build different data streams and provide them both for long-term storage on disk and online for commissioning, noise studies, data-quality definitions and reconstruction of the space-time strain (h(t)) before data analysis. Data are generated all over the detector: they include ITF sensing and control signals, monitoring signals from all the detector subsystems and environmental-monitoring signals. They are acquired through the fast-data network (∼400 Mbytes/s) and merged in different streams at the front-end of the data collection pipeline. Data are then reduced (decimation, compression, image processing, etc) in order to limit the flow on the Ethernet network and to require a reasonable storage amount of disk space. Data are then provided to several online processes (detector automation and monitoring, data-quality tools, h(t) reconstruction) which enrich the streams with additional computed channels that can be shared with other online processes. The data can be reached by a data-display tool to study the detector behavior online, with a latency of the order of a second for the front-end data.

At the end of the data-collection tree, different data streams are stored on disk using the GW experiment data format [94]. The raw-data stream (2 TB/day) contains all of the channels at their nominal sampling frequency, including the channels generated by the front-end electronics and the channels generated by the online processes along the data-collection tree. Other streams are generated: a fast stream, with the channels stored at their maximum frequency, before any decimation of the data collection pipeline, in order to monitor and debug the front-end electronics; a reduced data stream (RDS), 30 GB/day, with the channels decimated at lower frequencies than nominal (typically ∼100 Hz) for commissioning and noise studies at low frequencies; trend-data streams, with statistical information computed over one second (4 GB/day) and a few tenths of seconds over all of the channels; the h(t) data stream (7 GB/day), which contains the reconstructed GW strain channel and data-quality flags, and that can be directly used by the data analysis pipelines. The data are kept on circular buffers at the Virgo site, with a depth of ∼6 months for the raw data. During science runs all the data are transferred to computing centers for permanent storage and offline data analysis.

7. Possible synergies

The development of common software solutions for different ESFRI members or stakeholders is definitely easier for experiments or telescopes whose adopted or envisaged architectures are closer. Similar architectural solutions are usually shared by experiments that produce the same kind of data. The architectures for real-time data streaming adopted by current IACT considered in this survey share many similarities, as well as those chosen by present neutrino experiments and already envisaged for the Km3Net neutrino telescope. Both MAGIC and H.E.S.S. adopt their standard analysis packages for real-time data processing too but greatly reduce the time needed by skipping the calculation of calibrations parameters and applying a sub-optimal calibration procedure based on predefined parameters. The background rejection is also based on predefined strategies to avoid time-consuming strategy optimizations. To accelerate even further real-time data analysis, they moreover use a multi-threated analysis to process data acquired by each telescope separately and then merge the outcomes for the stereoscopic reconstruction. Km3Net envisages to adopt the "all-data-to-shore" approach, already followed by ANTARES, which consists in applying a very loose selection of signals underwater and send all the surviving events to computer farms on shore for further processing. They share therefore the need for very reliable selection procedure to be deployed underwater and for high-bandwidth connection to on-shore station to cope with the huge data-flow.

We notice that both the Optical/Near-Infrared telescopes and the Imaging Air Cherenkov Telescopes produce primarily images that have to be calibrated and analysed. Even if the final products are quite different, the images itself for the optical and infrared telescopes against the recorded data associated to a given cosmic particle for Cherenkov telescopes, the calibration and analysis procedure share quite some commonalities.

Highly parallelized calibration and analysis based on GPUs and other hardware accelerators could be adopted by both classes of experiments. The ASTRI prototype's software is already able to use GPUs for calibration and cleaning process and can run on low power architectures such as Nvidia TX1 as well.

A very attractive perspective is about the usage of deep learning techniques in the reconstruction process. Generally speaking, the process can be described as providing a software with an image input as obtained from the telescope (or maybe calibrated first) and let the software extract a number of features that are of interest for the scientific community, for example: gamma-hadron discrimination, energy reconstruction, direction of arrival. There is a lot of effort in industry and academia to improve this kind of machine learning, with new hardware available each year and software evolving more and more rapidly (see for example Google's Tensor Processor Units [57]).

In CTA's framework, many attempts have been made so far in using these techniques, giving rise to a separate working group named ASWG-ML (Analysis and Simulation Working Group -

Machine Learning) that have regular monthly discussions where many interesting results have been presented [58, 59].

They have been used in different network's architectures - mainly Convolution Neural Networks - on different hardware solutions.

These techniques have the interesting characteristic to be "portable" among different problems: the same network can be used to recognize dogs images from cats images as well as gammas from hadrons. Quite surprisingly, a network trained on dogs and cats images, behave very nicely on CTA images. This can lead to a joint effort in designing a network architecture specific for the astronomy domain.

8. Conclusions

The demand for real-time or nearly real-time data streaming is increasing rapidly in many scientific areas for different reasons. This is particularly true in the area of astronomy, astrophysics and astroparticle physics. The number of the acquisition channels of detectors adopted by upcoming observatories is steadily growing as well as their acquisition rate leading to huge flows of data that cannot be stored for long time. It is therefore necessary to analyse the incoming data stream in real-time to greatly reduce the amount of data to be transferred and stored. This is even more important for scientific facilities running in remote areas, like many future observatories, or satellite-borne detectors. Moreover, nowadays many scientific facilities are extremely sensitive to steering parameters that need to be continuously optimized by analysing acquired data. Finally, some research fields are characterized by transient phenomena whose study greatly benefits from the simultaneous observations by different detectors. Real-time analysis of data can provide the opportunity of issuing alerts to other research groups maximizing the overall scientific return.

In this survey we analysed the different approaches adopted by many ongoing research facilities or envisaged for the upcoming ones. The panorama emerged is quite varied, but some common trends seem to arise. The architectures for real-time data streaming surveyed are usually staged in many levels with lower levels, the ones closer to the detectors, very specific with very little chances to be replaced by common solutions. Higher levels instead share more commonalities, both hardware and software, that could be exploited to develop solutions that could be shared by different facilities. Often the online processing of data is performed on dedicated processor, like FPGA, or highly parallelized processors like GPUs which allow to accelerate the data analysis.

Another trend emerged from this survey is the increasing demand for data transfer with low latencies over long distances, for example from the observatories located in some remote areas to computing centres in North America or Europe. It is worthwhile to notice that very often present research facilities involve hundreds of scientist and engineers spread over many countries that need to access quickly data increasing the need for fast and reliable connections. Finally, an area where there will be great opportunity for sharing is in the algorithmic techniques that will be used to analyse the data. Machine Learning techniques are adopted by an increasing number of experiments and are often general enough to be shared between different research groups.

9.References

[1] https://www.eso.org/sci/facilities/eelt/

[2] http://www.lsst.org/

[3] http://www.euclid-ec.org/

[4] M. Jurić et al., The LSST Data Management System. Available at http://arxiv.org/abs/1512.07914

[5] ELT Programme – Observatory Top Level Requirements. Available at https://www.eso.org/sci/facilities/eelt/

[6] Euclid Definition Study Report. Available at http://arxiv.org/abs/1110.3193

[7] M. Jurić et al., The LSST Data Products Definition Document. Available at https://docushare.lsstcorp.org/docushare/dsweb/Get/LSE-163/

[8] http://fits.gsfc.nasa.gov/fits_home.html

[9] W.D. Pence et al., Definition of the Flexible Image Transport System (FITS), version 3.0, A&A 524 (2010) A42. Available at http://www.aanda.org/articles/aa/abs/2010/16/aa15362-10/aa15362-10.html

[10] http://www.ivoa.net/

[11] VOTable Format Definition. Available at http://www.ivoa.net/Documents/latest/VOT.html

[12] VOEvent: Sky Event Reporting Metada. Available at http://www.ivoa.net/documents/VOEvent/

[13] https://www.cta-observatory.org/

[14] https://www.mpi-hd.mpg.de/hfm/HESS/

[15] https://magic.mpp.mpg.de/

[16] http://www.km3net.org/home.php

[17] https://icecube.wisc.edu/

[18] http://antares.in2p3.fr/

[19] www.hawc-observatory.org/

[20] http://fermi.gsfc.nasa.gov/

[21] http://www.ams02.org/

[22] https://www.auger.org/

[23] G. Lamanna et al., Cherenkov Telescope Data Management, PoS (ICRC2015) 947. Available at: http://arxiv.org/abs/1509.01012

[24] J.A. Aguilar et al., The data acquisition system for the ANTARES neutrino telescope, NIM A 570 (2007) 107. Available at http://www.sciencedirect.com/science/article/pii/S0168900206017608

[25] Letter of Intent for KM3NeT2.0. Available at: http://arxiv.org/abs/1601.07459

[26] R. Abbasi et al., The IceCube data acquisition system: Signal capture, digitization, and timestamping, NIM A 601 (2009) 294. Available at: http://www.sciencedirect.com/science/article/pii/S0168900209000084

[27] https://root.cern.ch/

[28] A. Balzer et al., The H.E.S.S. central data acquisition system, Astropart. Phys. 54 (2015) 67. Available at http://www.sciencedirect.com/science/article/pii/S0927650513001655

[29] J. Aleksić, Optimized Dark Matter Searches in Deep Observations of Segue 1 with MAGIC, PhD Thesis (2013). Available at https://magicold.mpp.mpg.de/publications/theses/JAleksic1.pdf

[30] http://fermi.gsfc.nasa.gov/ssc/data/

[31] http://gamma-astro-data-formats.readthedocs.org/en/latest/

[32] https://www-zeuthen.desy.de/multi-messenger/GammaRayData/

[33] http://magic.pic.es/ (see section "Public area").

[34] S. Derrière et al., Using the Virtual Observatory: multi-instrument, multi-wavelength study of high-energy sources. Journées de la SF2A 2014. Available at http://arxiv.org/abs/1411.6514

[35] http://antares.in2p3.fr/publicdata.html

[36] [KM3NeTRef1] R. Gurin, A. Maslennikov, Controlhost: distributed data handling package, Technical report (1995) http://www.nikhef.nl/pub/experiments/antares/software/Dispatcher/ControlHost.ps

[36] https://icecube.wisc.edu/science/data

[37] https://www.hdfgroup.org/HDF5/

[38] http://icecube.umd.edu/PublicData/

[39] https://www.skatelescope.org/

[40] http://www.evlbi.org/

[41] http://www.lofar.org/

[42] M.P. van Haarlem et al., LOFAR: The LOw-Frequency ARray, A&A 556 (2013) A2. Available at http://www.aanda.org/component/content/article?id=950

[43] G.N.J. van Diepen, Casacore Table Data System and its use in the MeasurementSet, A&C 12 (2015) 174. Available at http://www.sciencedirect.com/science/article/pii/S2213133715000621

[44] https://casa.nrao.edu/

[45] https://prestodb.io/

[46] http://www.astron.nl/radio-observatory/astronomers/technical-information/lofar-technical-information

[47] SKA1 System Baseline v2. Available at https://www.skatelescope.org/key-documents/

[48] http://amiga.iaa.es/

[49] http://www.ligo.org/index.php

[50] http://www.virgo-gw.eu/

[51] LIGO Data Management Plant. Available at https://dcc.ligo.org/LIGO-M1000066/public

[52] F. Acernese et al., Advanced Virgo: a second-generation interferometric gravitational wave detector, Class. Quantum Grav. 32 (2015) 024001. Available at http://iopscience.iop.org/article/10.1088/0264-9381/32/2/024001/meta;jsessionid=E8990E15AD096C8D651D69A653E5A8AA.c4.iopscience.cld.iop.org

[53] Specification of a common data frame format for interferometric gravitational wave detectors. Available at http://dcc.ligo.org/cgi-bin/DocDB/RetrieveFile?docid=329

[54] http://amon.gravity.psu.edu/index.shtml

[55] https://github.com/open-gamma-ray-astro/2016-04_IACT_DL3_Meeting/blob/master/README.md

[56] https://www.asterics2020.eu/dokuwiki/doku.php?id=open:wp4:start

[57] http://www.evalso.eu/evalso/

[58]https://indico.cta-observatory.org/category/37/

[59] D.Nieto et al. “Exploring deep learning as an event classification method for the Cherenkov Telescope Array” https://arxiv.org/abs/1709.05889