New tools for better-characterizing astroparticle physics sites and detectors for the Latin American Giant Observatory

Antonio Juan Rubio-Montero[1], Raúl Pagán-Muñoz[1], Alfonso Pardo-Diaz[2], Álvaro Taboada[4], Iván Sidelnik[3], Rafael Mayo-García[1] and Hernán Asorey [4,5,*]

on behalf of the LAGO Collaboration[5]

[1] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), [2] Centro Extremeño de Tecnologías Avanzadas (CETA-CIEMAT), [3] Departamento de Física de Neutrones (CNEA/CONICET), [4] Instituto de Tecnologías en Detección y Astropartículas (ITeDA, CNEA-CONICET-UNSAM), [5] Departamento Física Médica (CNEA), [6] lagoproject.net

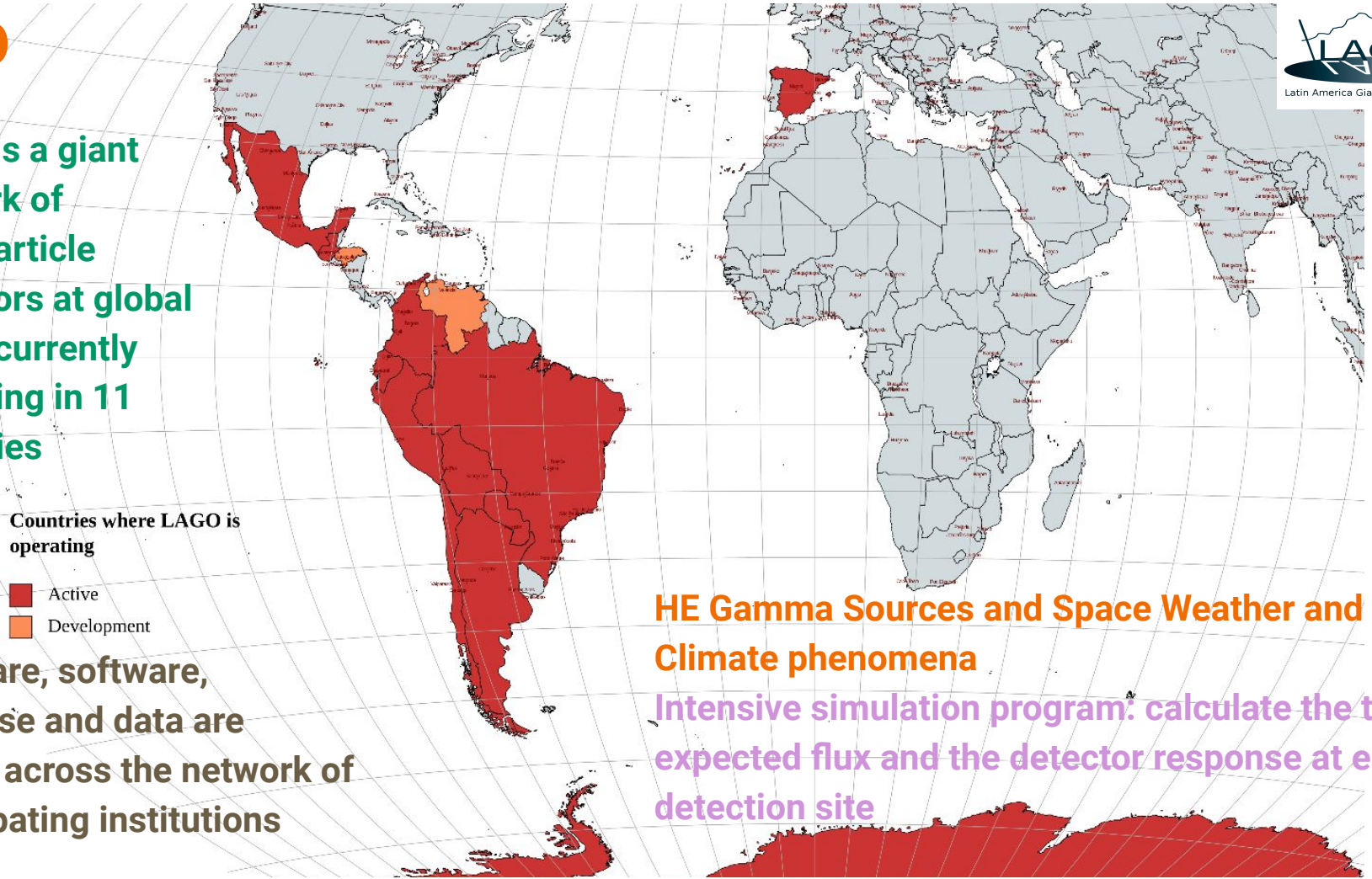
*hernan.asorey@iteda.cnea.gov.ar, presenter

Horizon 2020 RI project 857647

LAGO

Latin America Giant Observatory
lagoproject.net

LAGO is a giant network of astroparticle detectors at global scale, currently operating in 11 countries



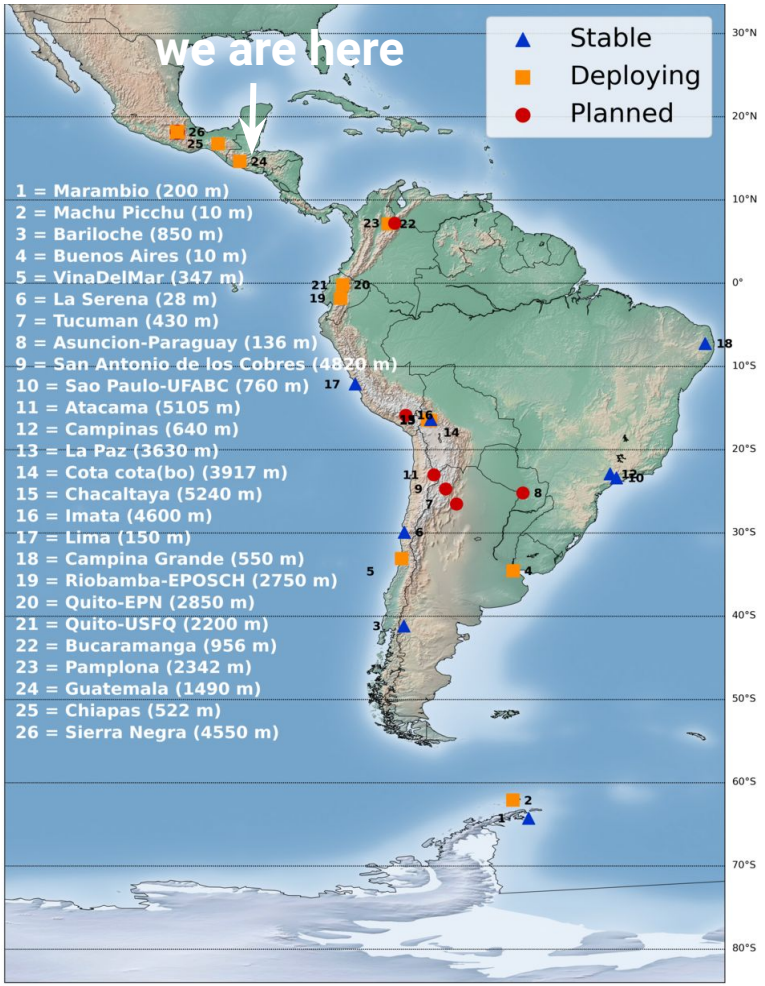
Hardware, software, expertise and data are shared across the network of participating institutions

HE Gamma Sources and Space Weather and Climate phenomena

Intensive simulation program: calculate the total expected flux and the detector response at each detection site

The Latin American Giant Observatory

Latin America Giant Observatory
lagoproject.net

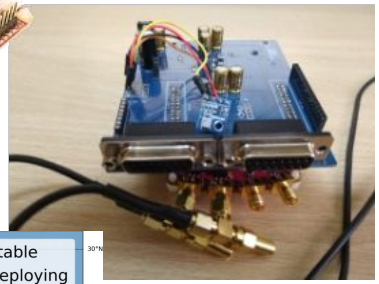


LAGO is an extended astroparticle observatory at continental scale: from Mexico to Antarctica

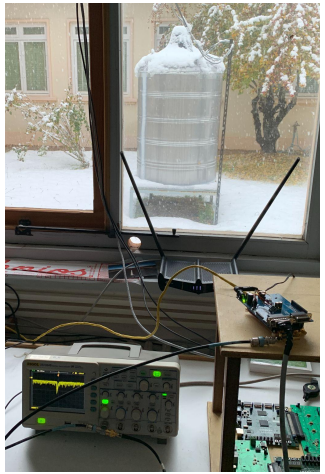
- **Scientific goals**
  - Astroparticle physics to study the extreme universe
  - Transient and long term space weather phenomena trough Solar modulation of Cosmic Rays
  - Measurement of background radiation at ground level
- **Academic goals**
  - We are a HEP and astro-ph seedbed for Latin American students
  - Build a Latin American network of Astroparticle and Cosmic Rays researchers

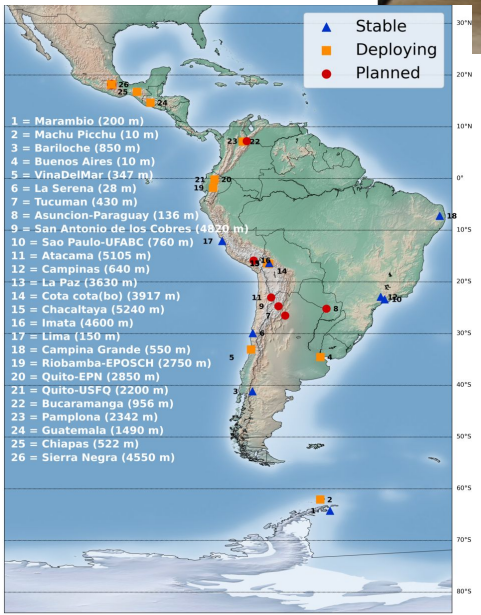
*I. Sidelnik & H. Asorey, NIM A **876** 173, 2017; H. Asorey et al., PoS ICRC2015 247, 2015*

The Latin American Giant Observatory



Own designed electronics based on RedPitaya SteamLab. Provides services for other experiments





1-10 m^3 WCD deployed at different altitudes and geomagnetic coordinates

LAGO WCDs have a single PMT and are based on SBC and COTS. They are autonomous, reliable, simple, smart, and inexpensive.

H. Arnaldi et al., IEEE CAE48787, 2020; H. Asorey et al., PoS ICRC2015 247, 2015

LAGO Capabilities: Multi-spectral analysis

- Simultaneous measurements of secondaries at ground level
- Intensive simulation and data analysis frameworks

Connections

CR Flux	Astrophysics transients →	Modulated flux $\cdots$
$\cdots$ Modulated flux	Geomagnetic field →	Primaries $\cdots$
$\cdots$ Primaries	Atmospheric conditions →	Secondary particles $\cdots$
$\cdots$ Secondary particles	Detector response →	**Signals**

Synergy

Flux variation of signals at detector level⇔Transients

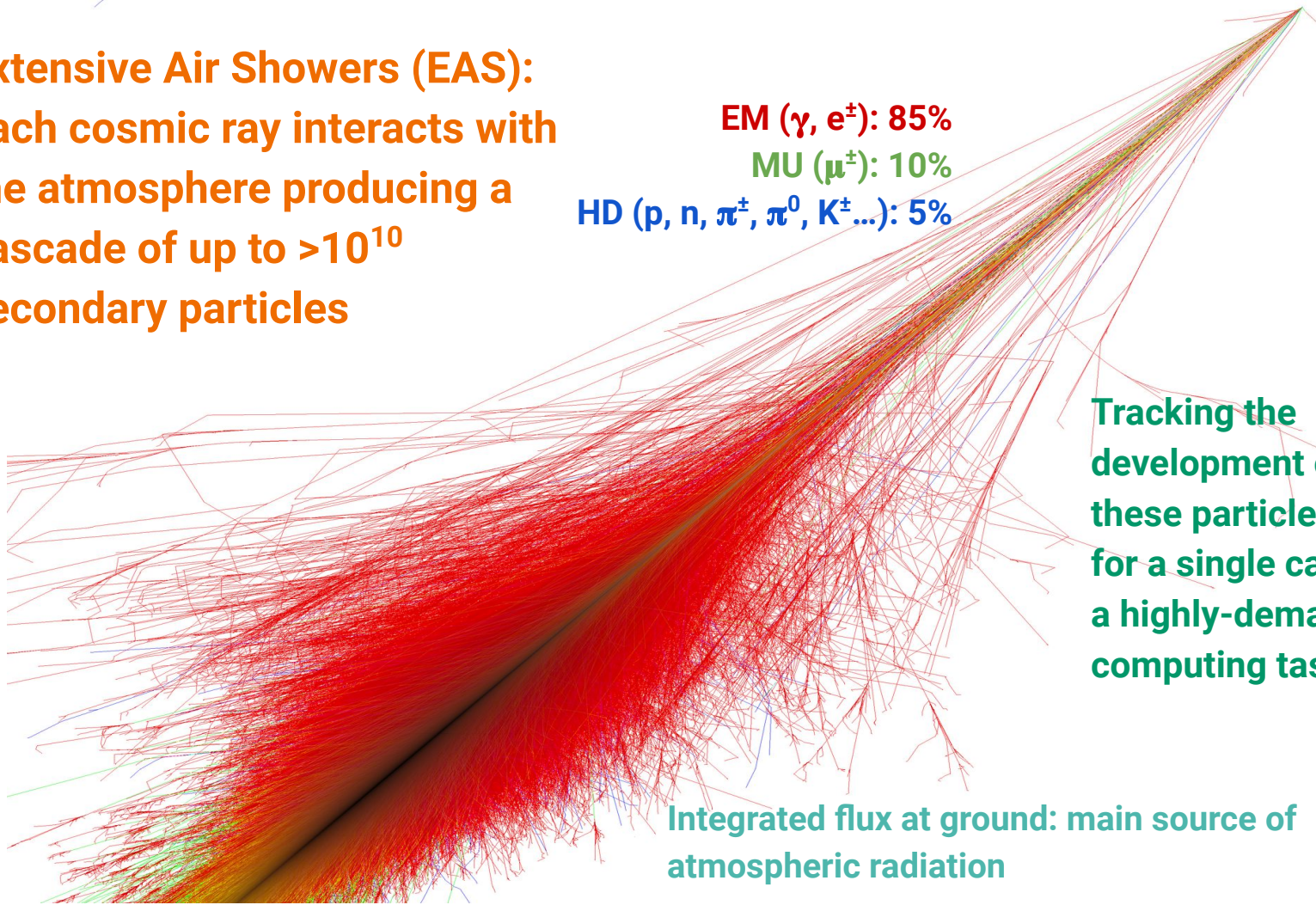
H. Asorey et al., PoS ICRC2015 142, 2015

Extensive Air Showers (EAS): each cosmic ray interacts with the atmosphere producing a cascade of up to >10^{10} secondary particles

EM (γ, $e^{\pm}$): 85%

MU ($\mu^{\pm}$): 10%

HD (p, n, $\pi^{\pm}$, π^{0}, $K^{\pm}$…): 5%



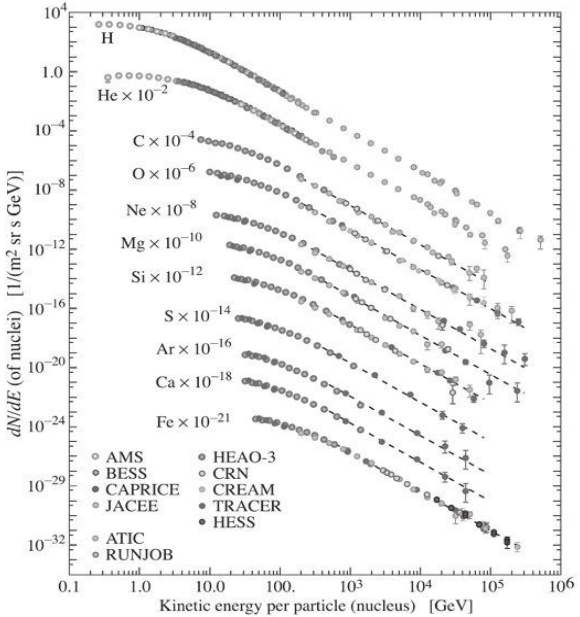
Tracking the development of all of these particles, even for a single cascade, is a highly-demanding computing task

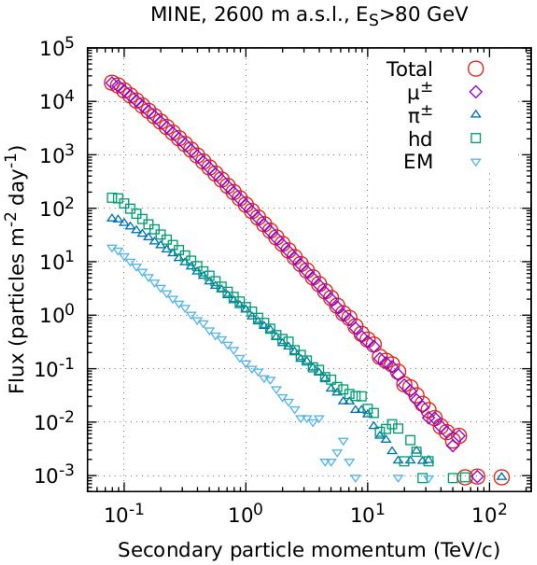
Integrated flux at ground: main source of atmospheric radiation

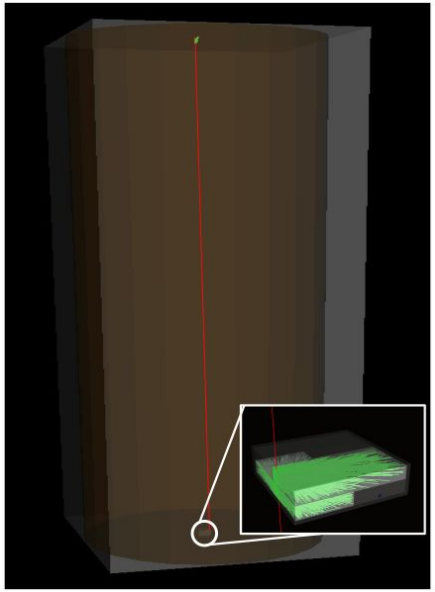
Latin America Giant Observatory
lagoproject.net

ARTI: The LAGO Simulation framework

GCR Flux → (Heliosphere) → Modulated Flux → (Magnetosphere) → **Primaries** → (Atmosphere) → **Secondaries** → (Detector) → **Signals or Doses**



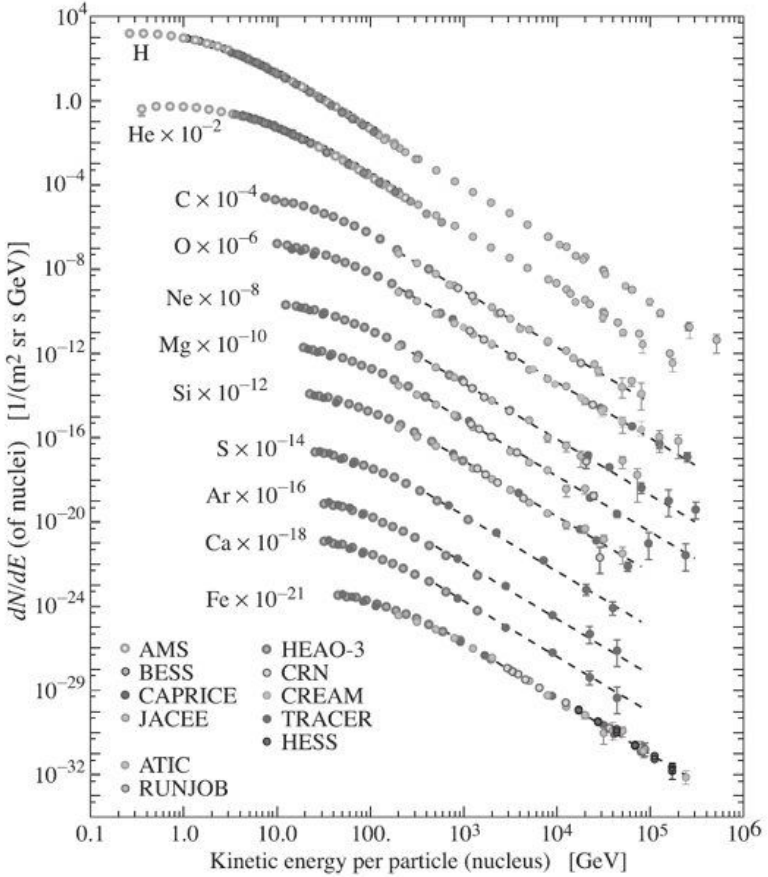




ARTI + CORSIKA | **ARTI** | **ARTI + MEIGA (Geant4)**

C. Sarmiento Cano et al., EPJ C in press, 2022, arXiv:2010.14591

Primary flux integration



For each primary, we need to integrate its spectrum to get the expected number of primaries at the top of the atmosphere

$$N_{t,S} = \int_t \int_S \int_\Omega \int_{E_p} j_0(E_p, Z_p)^{\alpha(E_p,Z_p)}\, dt\, dS\, d\Omega\, dE$$

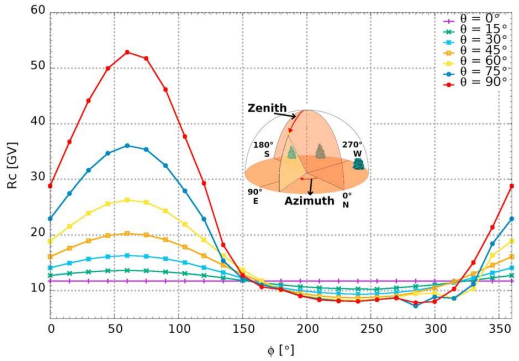
We integrate:

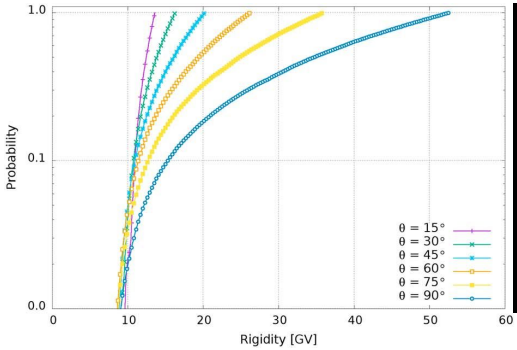
- full spectra, $1 < Z < 26$
- hemisphere, $0 \leq \theta \leq \pi/2, -\pi \leq \phi \leq \pi$
- energy range, $(R_C \times Z_p) < E/GeV < E_{\max}$

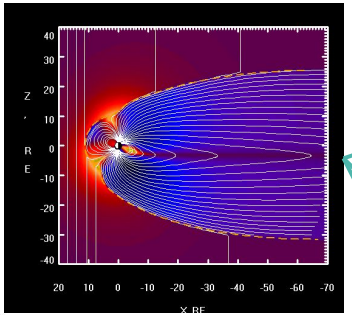
R_C is the local, time-dependent, geomagnetic rigidity cut-off
E_{max} depending on application

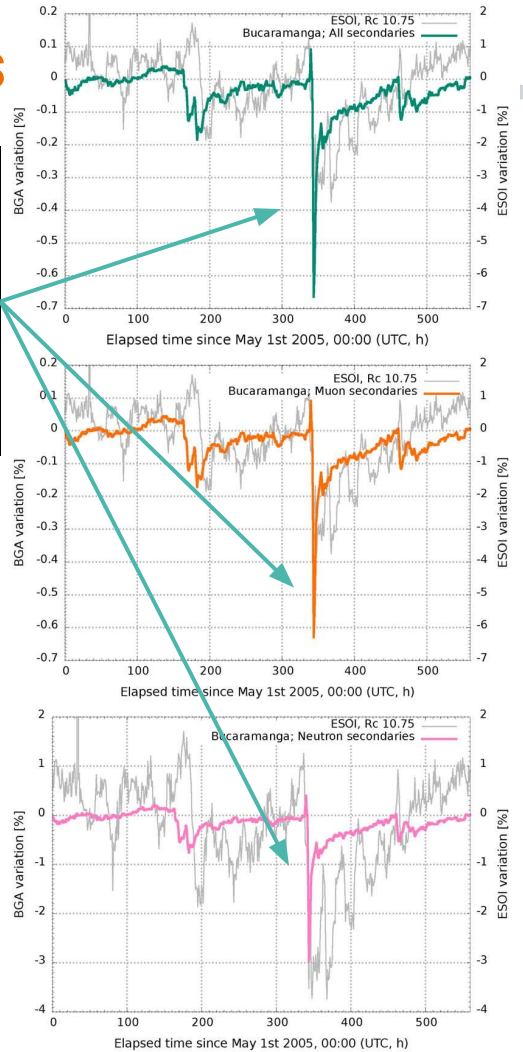
M. Suárez-Durán, H. Asorey and L.A. Núñez, Space Weather, (2018), DOI:10.1002/2017SW001774

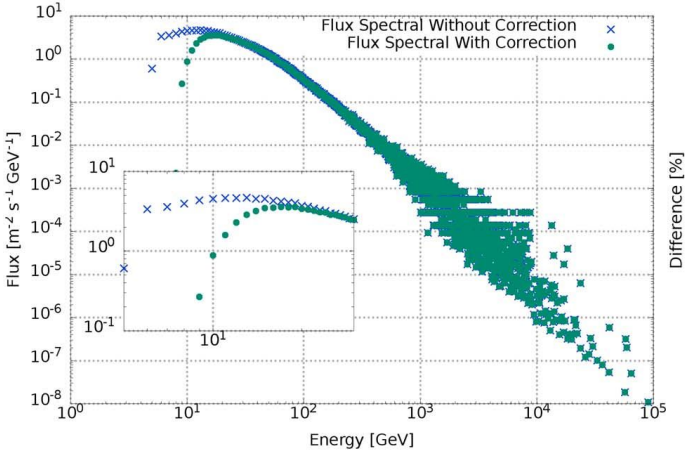
Time-dependent local geomagnetic effects

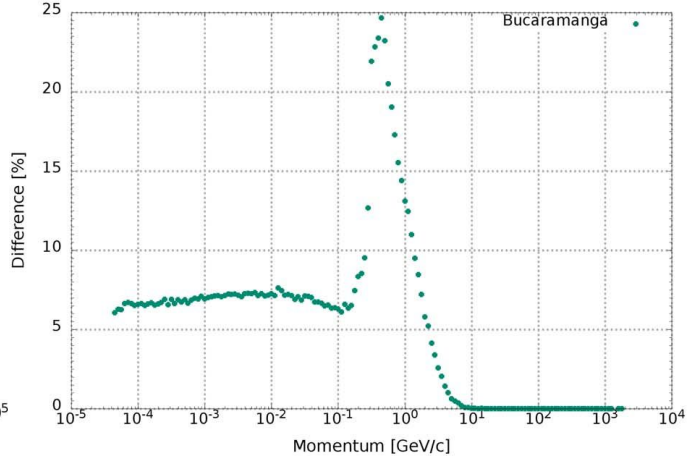








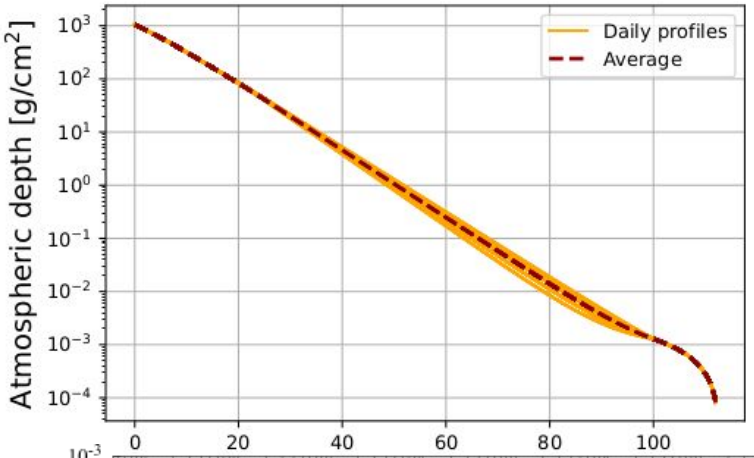


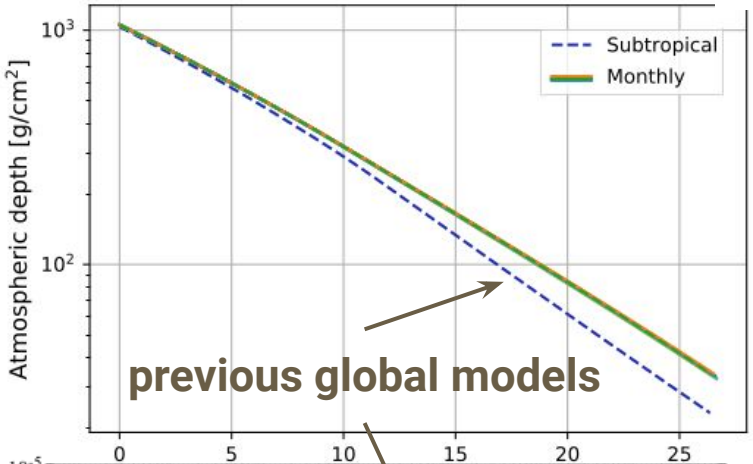


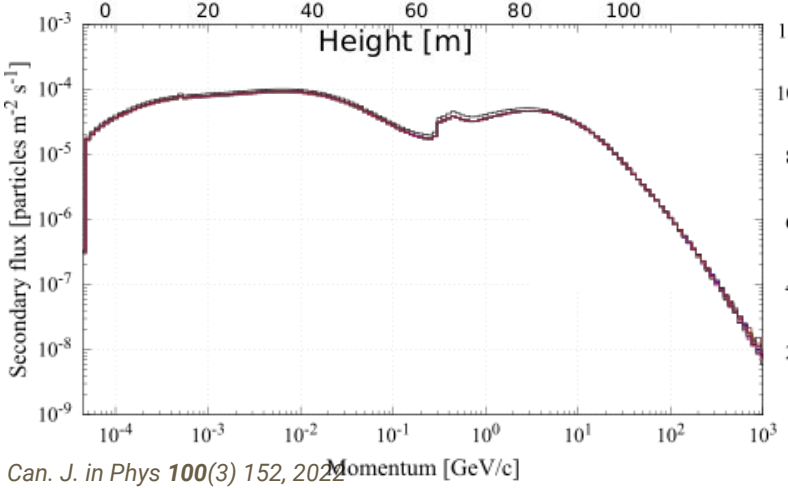
M. Suárez-Durán, H. Asorey and LA Núñez, Space Weather, (2018), DOI:10.1002/2017SW001774

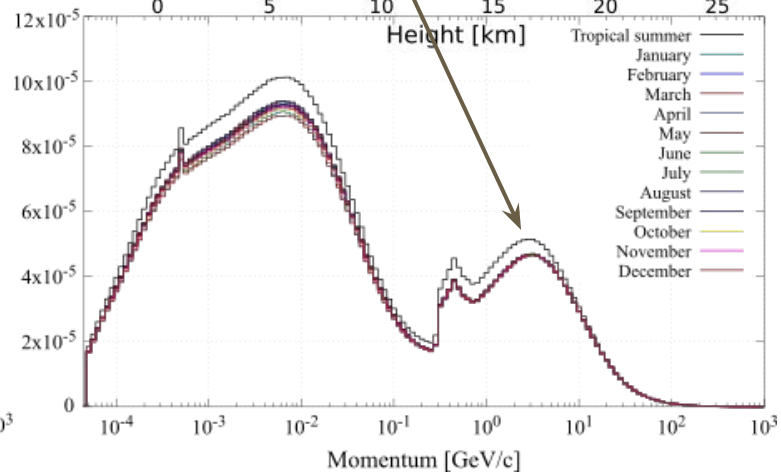
Local atmospheric effects

Monthly-averaged or instantaneous local atn profiles from GDAS





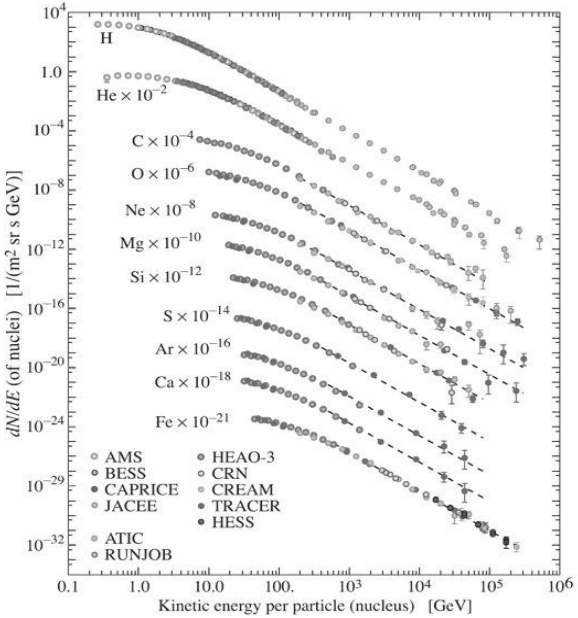


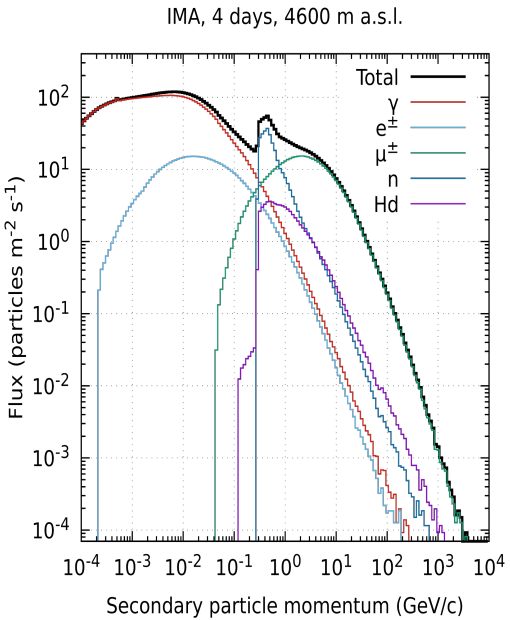


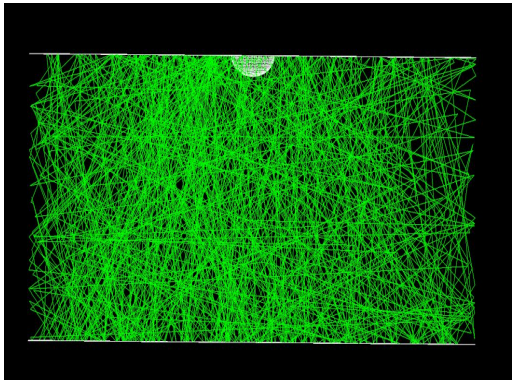
J. Grisales et al., Can. J. in Phys ***100****(3) 152, 2022*

ARTI: The LAGO Simulation framework

GCR Flux → (Heliosphere) → Modulated Flux → (Magnetosphere) → **Primaries** → (Atmosphere) → **Secondaries** → (Detector) → **WCD complete simulation**







ARTI (CORSIKA) **ARTI** **ARTI+MEIGA (Geant4)**

R. Mayo-García, this symposium, *C. Sarmiento Cano et al., EPJ C in press, 2022, arXiv:2010.14591*

Challenge: profit from the effort of every collaborator

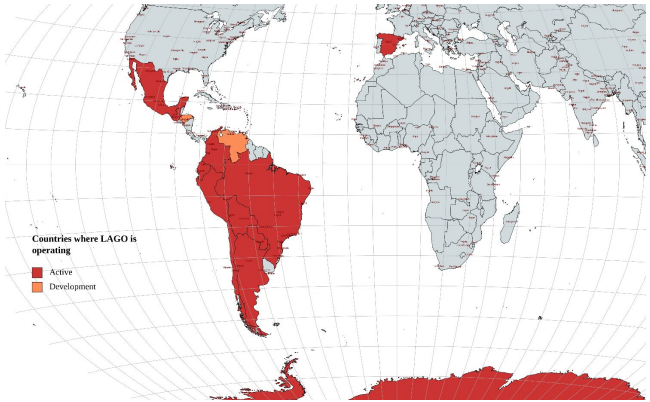
- Non-centralized alliance (30 institutions, 11 countries)
- >100 researchers grouped into autonomous research units
- Externals & students are join the project for short periods of time
- Computation is supported by the heterogeneous resources of each institution (private HPC facilities)

Latin America Giant Observatory
lagoproject.net

Every result (simulations, measurements, code or designs) is of interest to other members and external actors (industrial, government...)

Need of effective mechanisms for:

- **curation and data publishing**
- **reproducibility**
- **coordination of the development**
- **sharing computational resources**



The FAIR paradigm

Latin America Giant Observatory
lagoproject.net

Findability

discoverable, identifiable and locatable by means of a standard identification mechanism

PiD, Metadata

Accessibility

always available and obtainable; even if the data is restricted, the metadata is open

Cloud storage

Interoperability

syntactically parseable and semantically understandable, allowing data exchange and reuse

Vocabulary

Reusability

described and shared with the least restrictive licences, allowing the widest reuse possible

Harversters, copyleft

Our Approach: “FAIRification” of every result

Latin America Giant Observatory
lagoproject.net

Data, software, materials should be considered **as Linked Open Data**:

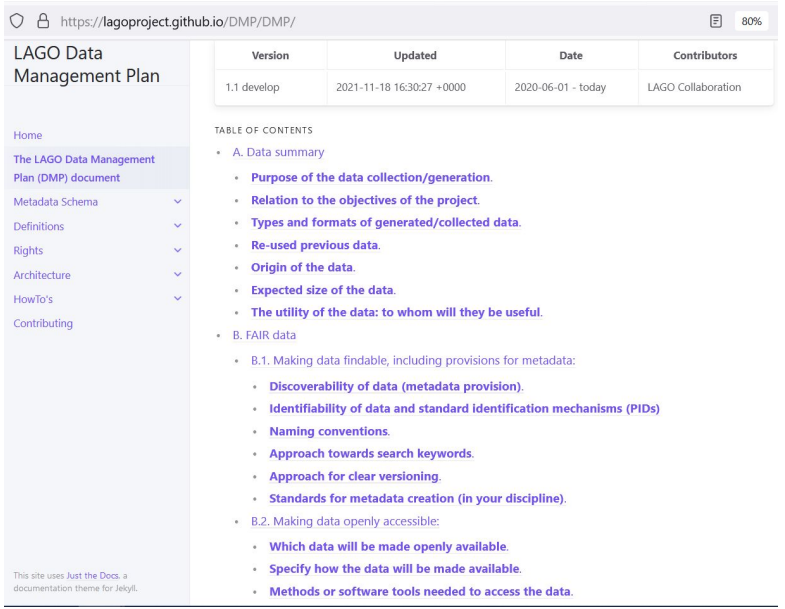
- Univocally referenced with **PiDs and/or IRIs**
- **Metadata** in current standard formats for linked-data
- **Provenance (IRIs in metadata)**
- Version control (CVS commits)
- Open policies and licences
- **Structured** and open-formatted (if possible)

All generated and available on **Open platforms**:

- **Computing, storage, and provisioning tools**
- **Virtual Organisation management**
- CD/CI, quality assurance tools

Whole described in an extended **DMP document**:

- not only a “*Data*” Management Plan
- comprises whole docs: how-to’s, team rules, **workflows**
- is a living document
- is another result (with PiD/IRI, version control, public)



AJ Rubio-Montero et al., IEEE WSC2021 21706697, 2021

Expanding the capacity and capabilities of European Open Science Cloud by leveraging the experience, effort and resources of national publicly-funded digital infrastructures

EOSC-Synergy Horizon 2020 RI project 857647

Objectives aligned to Open Science

+

Coordination with other EU Projects

=

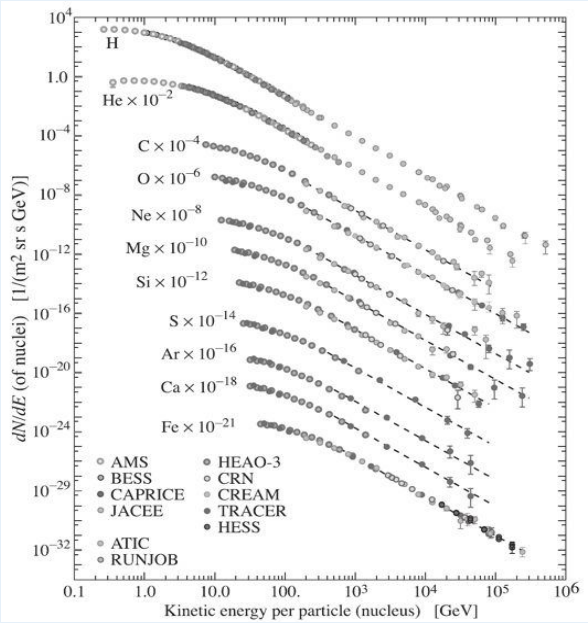
Open Platforms & Tools for:

- Computing & Storage & Publishing
- Community Management (IdP...)
- CI/CD & quality assurance
- Expertise

A. Calatrava et al., Comp. Sci. Rev, submitted, 2022

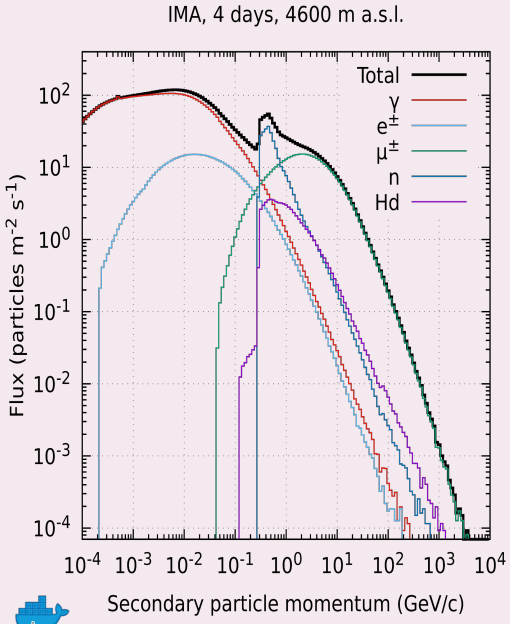
Standardizing: data-sets & encapsulates pipeline steps

GCR Flux → (Heliosphere) → Modulated Flux → (Magnetosphere) → **Primaries (S0)**



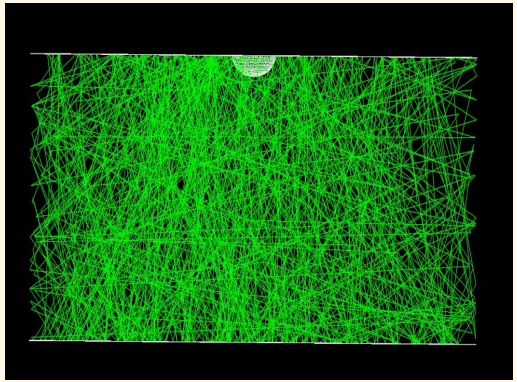
docker **onedataSim-S0**

Atmosphere → **Secondaries (S1)**



docker **onedataSim-S1**

Detector → **Signals or Doses (S2)**



docker **onedataSim-S2**

See R. Mayo's talk (this conference). A. J. Rubio- Montero et al., PoS ICRC2021 261, 2021

Standardizing: workflow and some figures

Latin America Giant Observatory
lagoproject.net

Transient astrophysical phenomena
GRBs, Solar Activity…
Inputs

Flux of CR

$4.5 \times 10^{8}\ \text{day}^{-1}\ \text{m}^{-2}$

Secondaries

Secondary propagation and detector response
type, geometry, materials...
GEANT4 detector models

Primaries

$4 \times 10^{8}\ \text{day}^{-1}\text{m}^{-2}$ @4600 m asl

EM : MU : NE : HD= 0.875 : 0.065 : 0.050 : 0.010

Site characteristics
location, altitude, geomagnetic field, atmosphere…
time-evolving conditions:
MAGNETOCOSMICS (IGRF13&TSY), GDAS and CORSIKA

Signals or Doses

See R. Mayo's talk (this conference). C. Sarmiento Cano et al., EPJ C in press, 2022, arXiv:2010.14591

Standardizing: workflow and some figures

Transient astrophysical phenomena

GRBs, Solar Activity, …

Flux of CR

4.5 x

ion and

se

rials...

dels

cker

Too many parameters and files: must ensure the provenance and reproducibility defining the data and metadata structures

Signals or Doses

location, altitude, geomagnetic field, atmosphere…

time-evolving conditions:

MAGNETOCOSMICS (IGRF13&TSY), GDAS and CORSIKA

C. Sarmiento Cano et al., EPJ C in press, 2022, arXiv:2010.14591

Schemas for data and metadata

Latin America Giant Observatory
lagoproject.net

Metadata:

- Language syntax: **JSON-LD 1.1** (W3C). Enabling linked-data.

- Main vocabulary: **DCAT-AP2** (), a profile of DCAT2 (W3C), for government, research centers & funded projects.

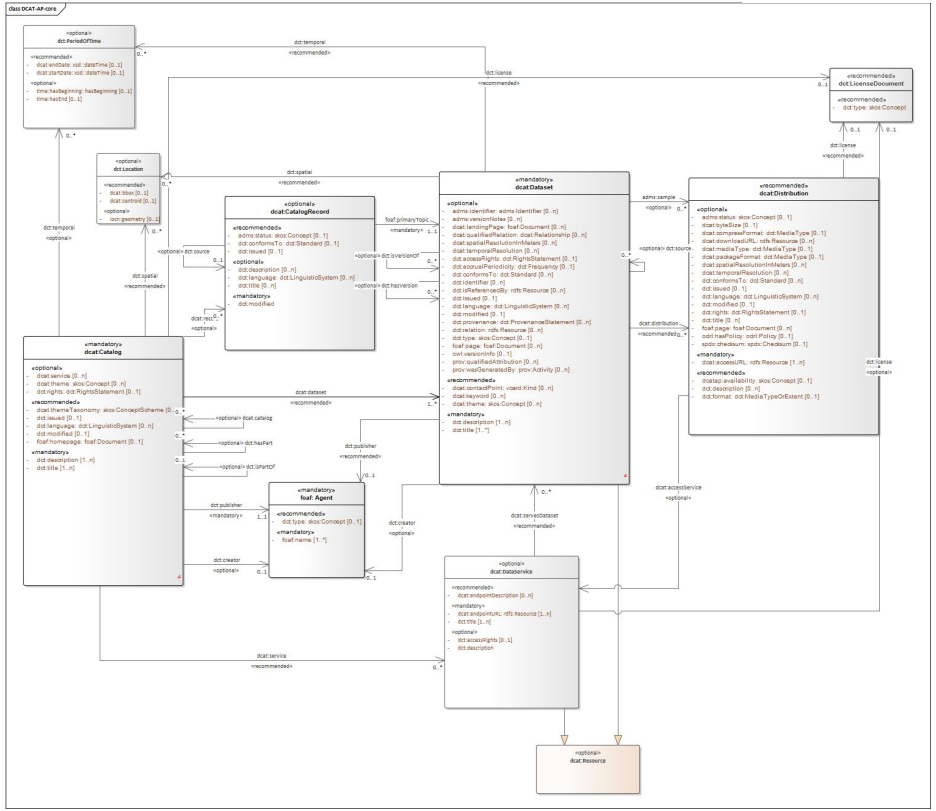
- The **LAGO vocabulary: re-profile of DCAT-AP2**, extending classes & properties needed (geomagnetic locations, expected atmospheres, software...).

Data conventions compatible with DCAT-AP2:

- A super-Catalog comprises whole repository

- A Catalog is any complete simulation (S0, S1, or S2)

- A Dataset is any file belonging one simulation



ISA² is now interoperable europe

European Commission

AJ Rubio-Montero et al., IEEE WSC2021 21706697, 2021

How does it work?

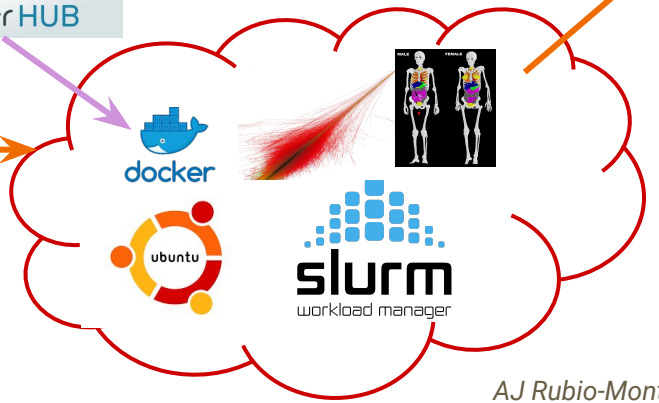
HPC provider assigns cloud resources to EOSC:
n Nodes, *r* GB/TB of RAM, *d* TB local storage

Infrastructure Manager

web-based service with templates for distributing to available resources in virtual clusters. Installs OS, in our case, ubuntu 20.04 + slurm manager

docker containers with our codes are deployed from docker HUB in the virtual cluster

virtual cluster



FAIR: catalogs are findable, accessible, interoperable, and reusable

Results are stored at cloud-based storage services. Access trough personal tokens

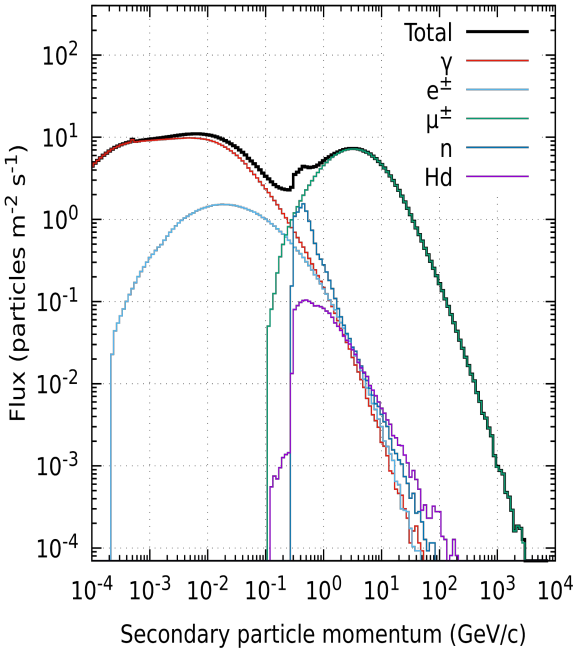
PID (Persistent IDentifiers) are assigned for each data catalog

AJ Rubio-Montero et al., IEEE WSC2021 21706697, 2021

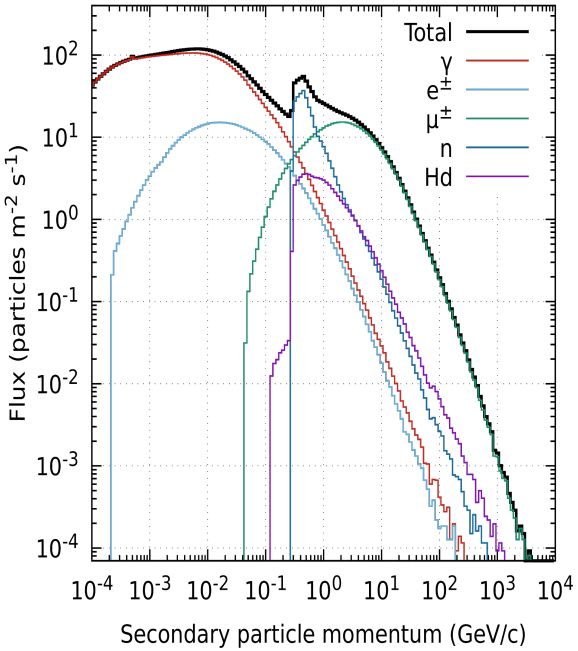
Expected secondary flux

Detailed flux of of secondary particles at any location **around the World.**

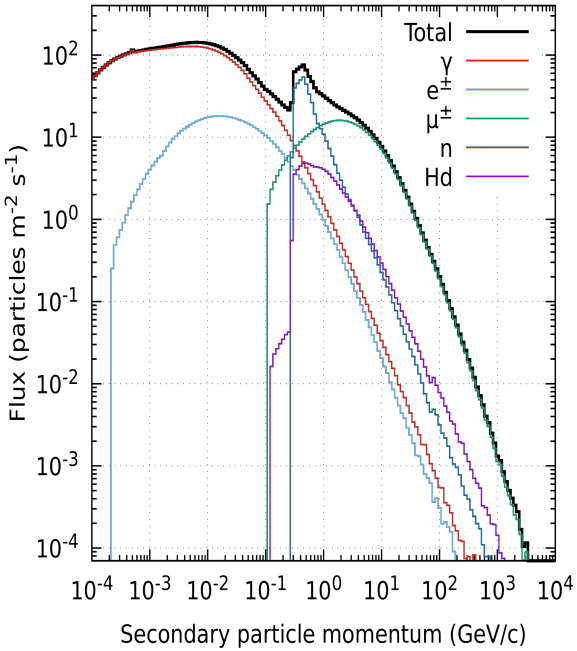
LSC, 3 days, 28 m a.s.l.



IMA, 4 days, 4600 m a.s.l.



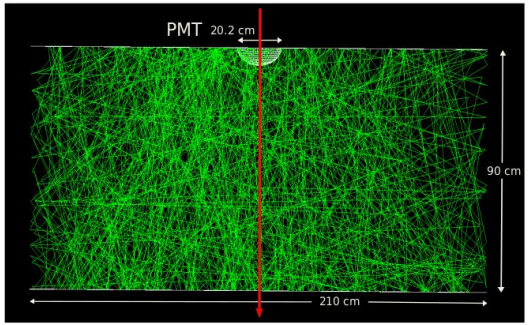
SAC, 7 days, 4500 m a.s.l.

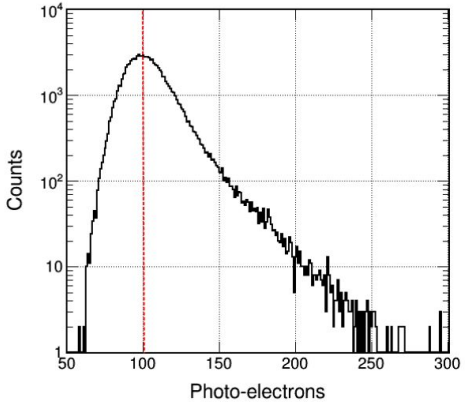


Secondary flux in 4 LAGO sites: La Serena (CL), Imata (PE) and San Antonio de los Cobres (AR)

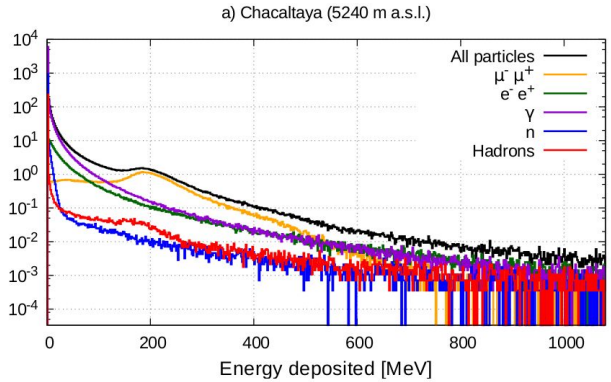
AJ Rubio-Montero et al., (ICRC 2021), DOI:10.22323/1.395.0261

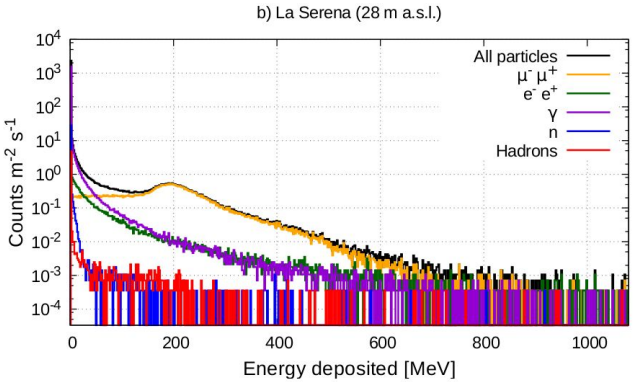
Signals simulations: deposited charge histograms

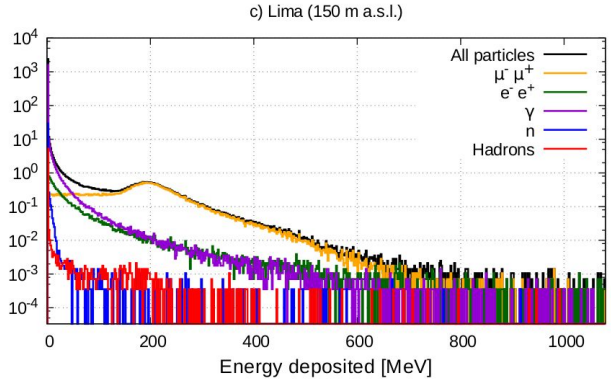


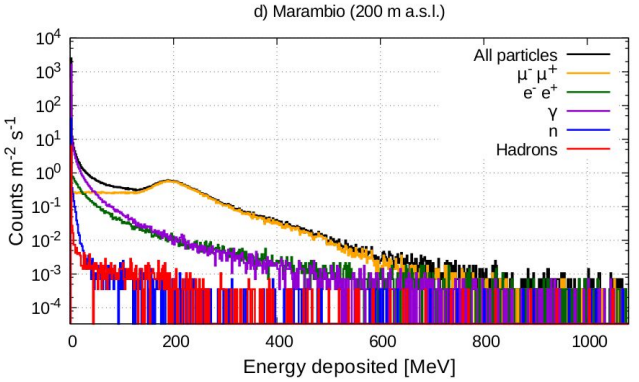


Vertical Equivalent Muon (VEM)







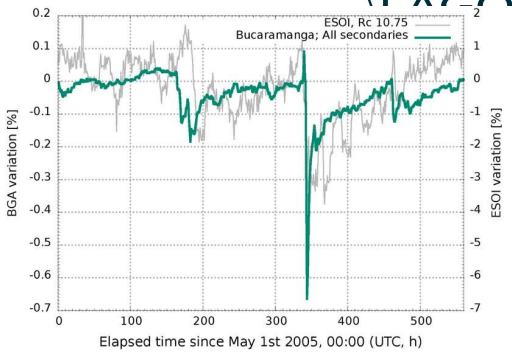


Expected histograms at Chacaltaya (BO), La Serena (CL), Lima (PE) and Marambio (ANT)

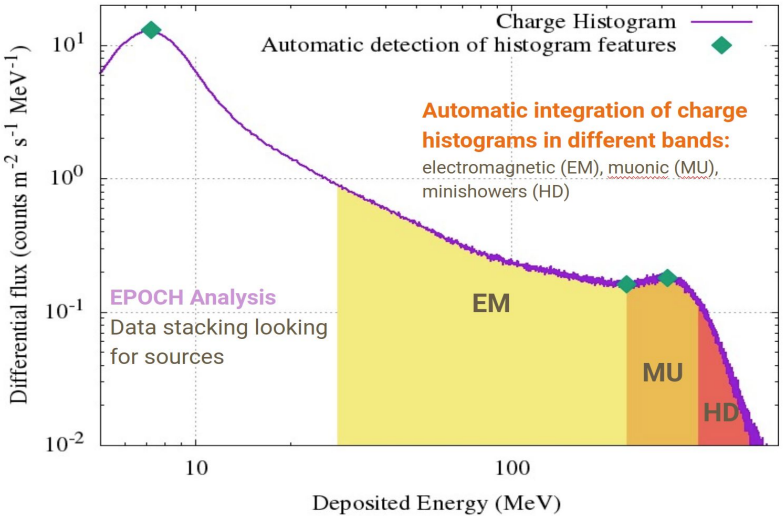
C. Sarmiento Cano et al., EPJ C in press, 2022, arXiv:2010.14591

LAGO Space Weather

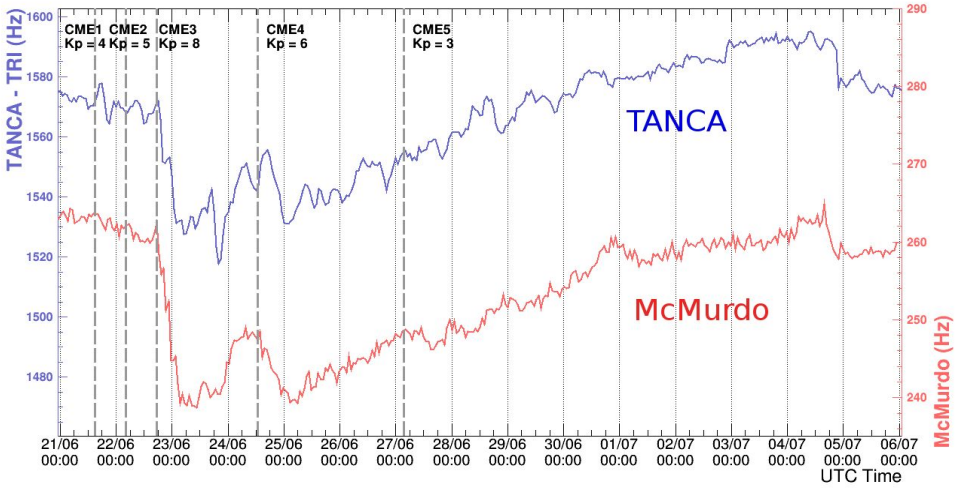
LAGO studies Earth-Sun connection by measuring the time-evolving secondary signals from ground level. Atmospheric conditions are continuously monitored



Multispectral analysis of the charge histogram at the WCD

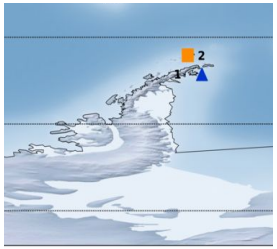


Forbush Decrease Observation at Campina (BR)



H. Asorey et al., PoS ICRC2015 142, 2015; de Aguiar et al., PoS ICRC2021 267, 2021

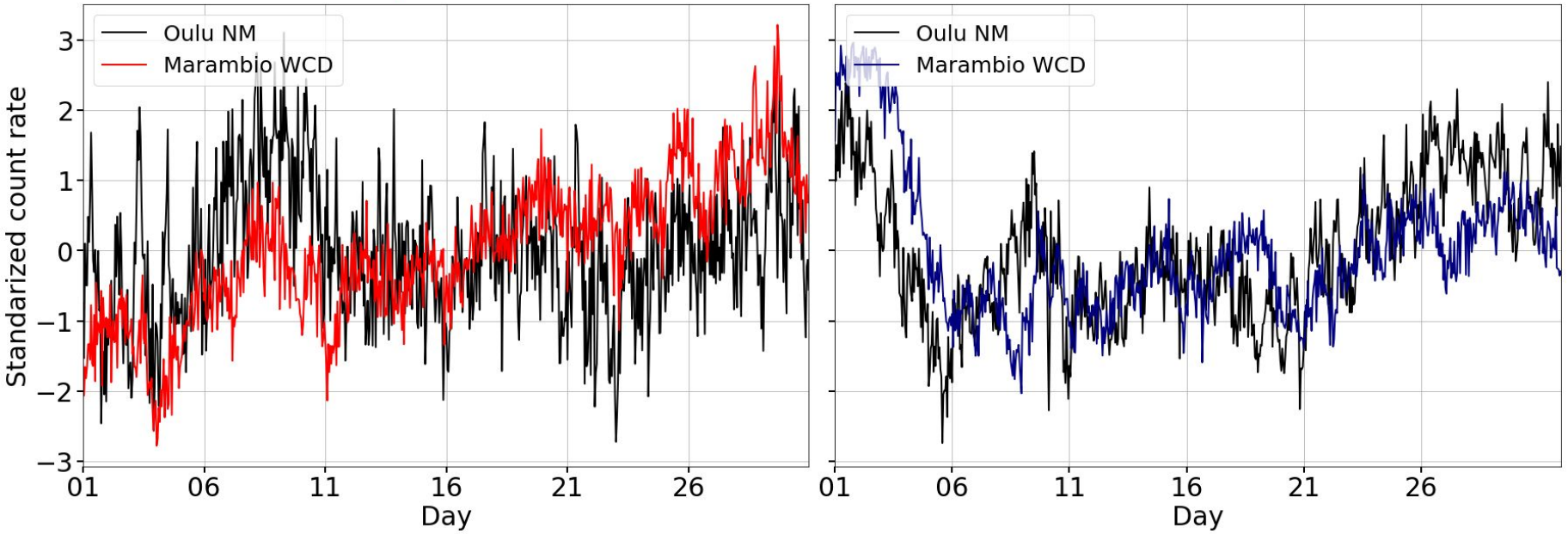
LAGO Space Weather



Long term analysis for space climate at low rigidity sites: LAGO observations at the Antarctica Peninsula

April 2019

December 2020



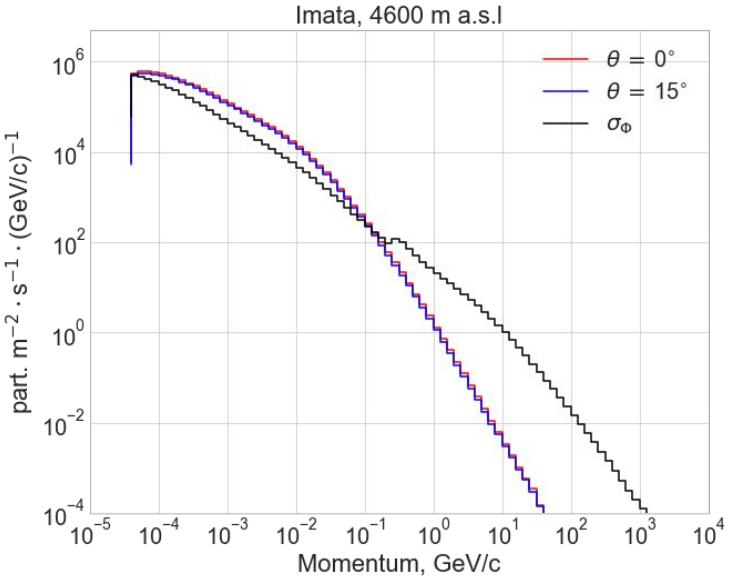
N. Santos et al, (ICRC2021), DOI:10.22323/1.395.0304

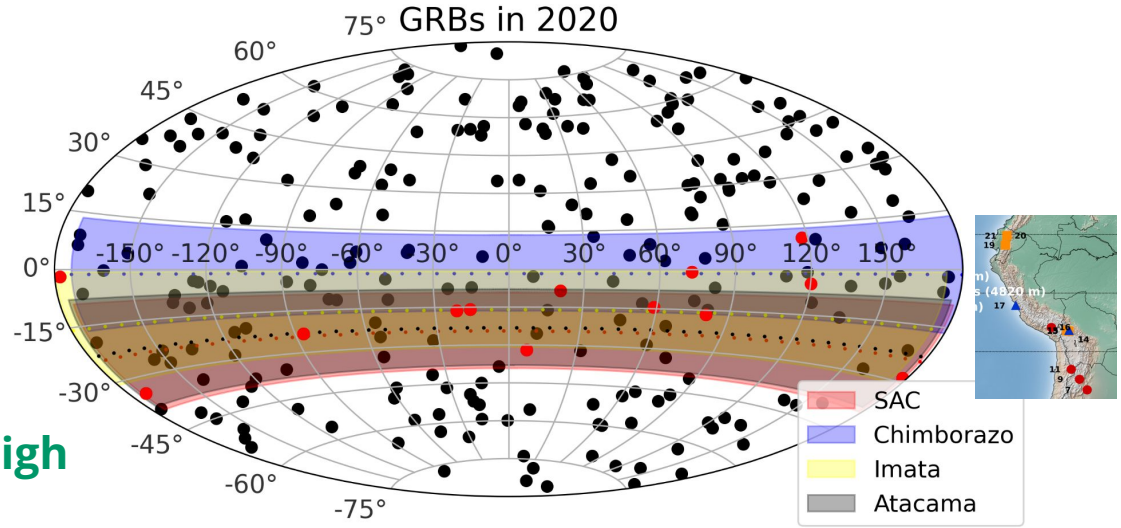
LAGO HE program: GRBs

Small arrays of WCD at very high altitude sites (<4500 m asl)

FOV overlapped by design for simultaneous measurements

muon decay for signal-to-noise ratio improvement

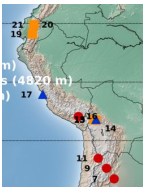


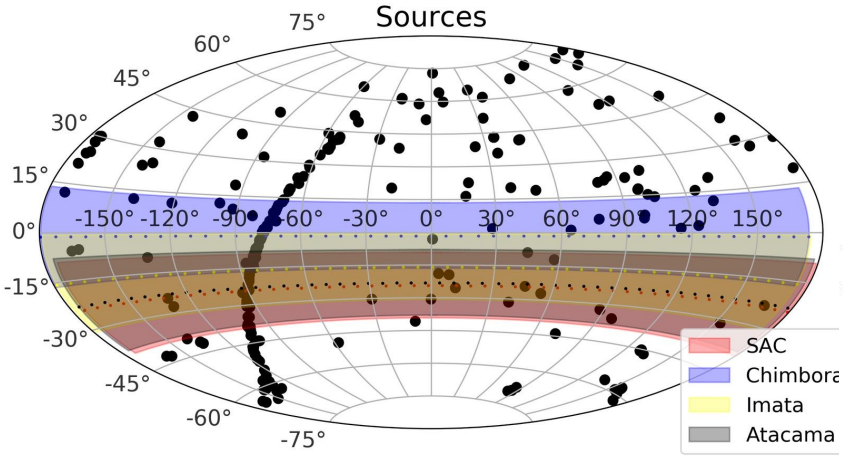


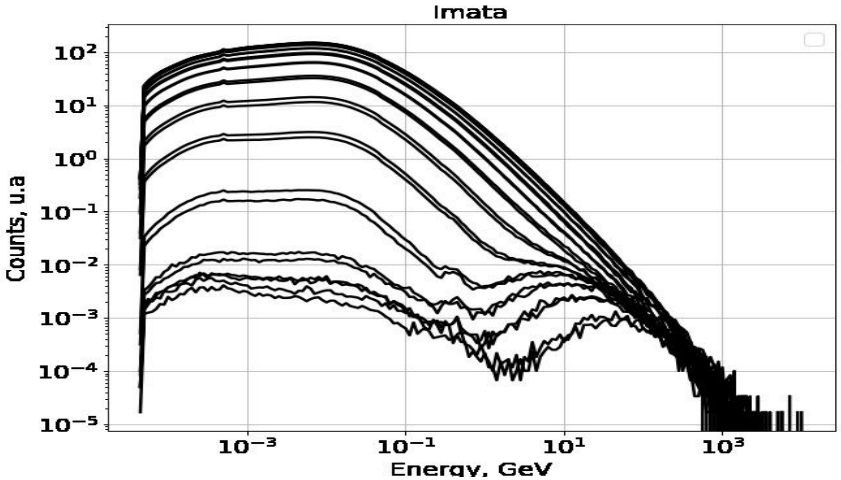
LAGO Capabilities for detecting high energy component of GRBs and Gamma Sources

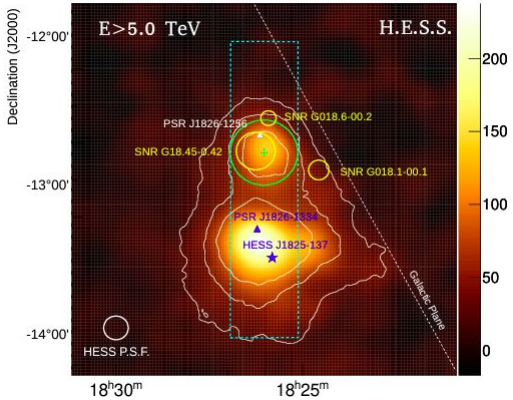
C. Sarmiento-Cano, et al. (ICRC2021), DOI:10.22323/1.395.0929

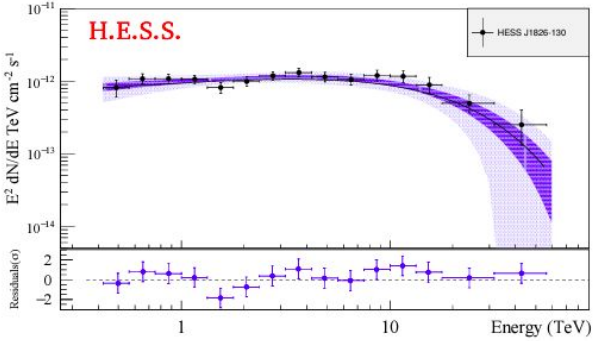
LAGO HE program: Steady γ sources: HESS J1826-130

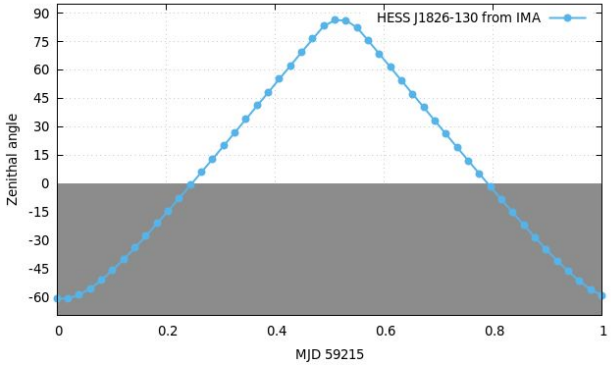






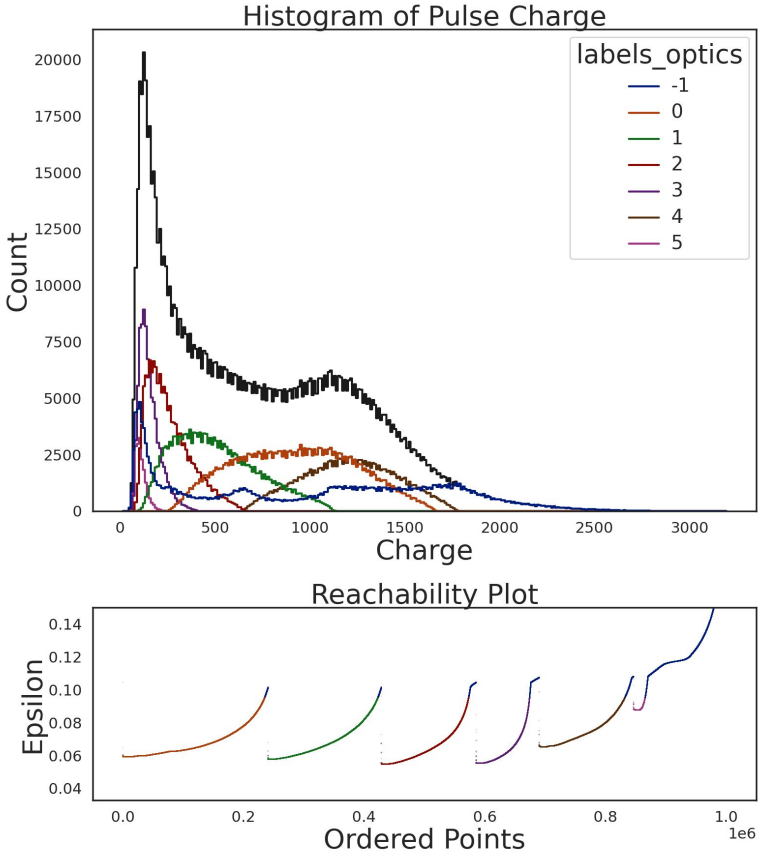






A&A644, A112 (2020), I. Sidelnik et al, RICH'22, next week at Edinburgh

Particle classification using ML algorithms

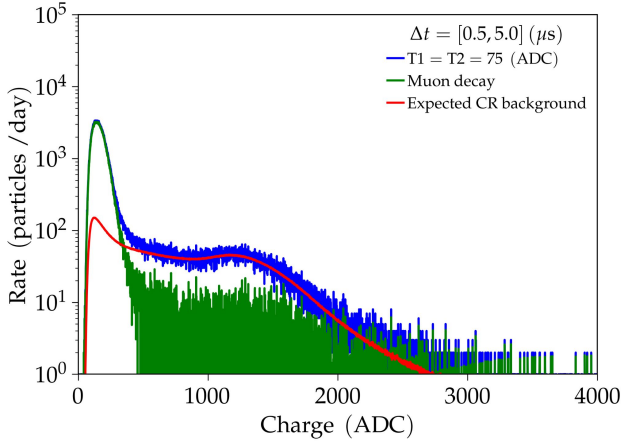


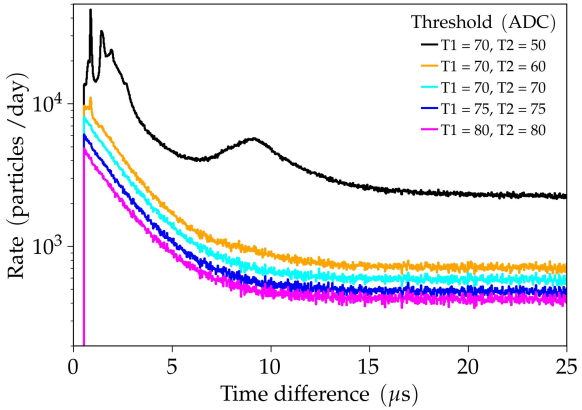
What kind of particles produce the observed features in the charge histogram?

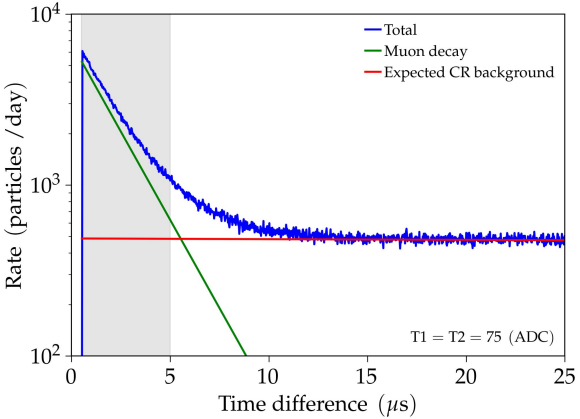
- OPTICS clustering algorithms
- Expected features observe
  - EM (photons and electrons), muon and multi-particle contributions are clearly observed
- **Multi-spectral analysis allowed**
  - Gamma/hadron separation
  - Space Weather spectroscopy
  - Extensive Air Showers analysis: LAGO sensitivity up to the knee

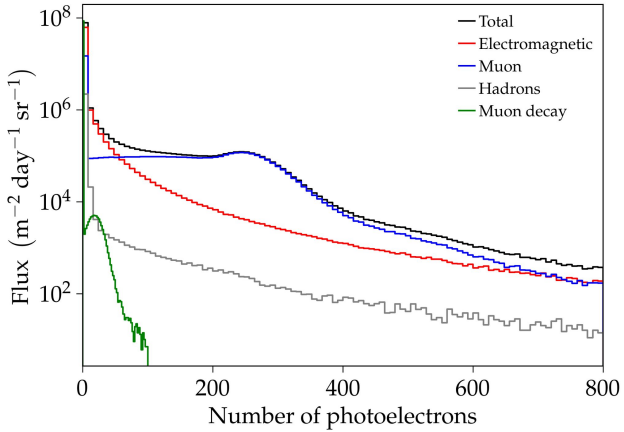
T. Torres-Peralta et al., RICH'22, next week at Edinburgh

Measurement of the Muon Lifetime & Michel Spectrum



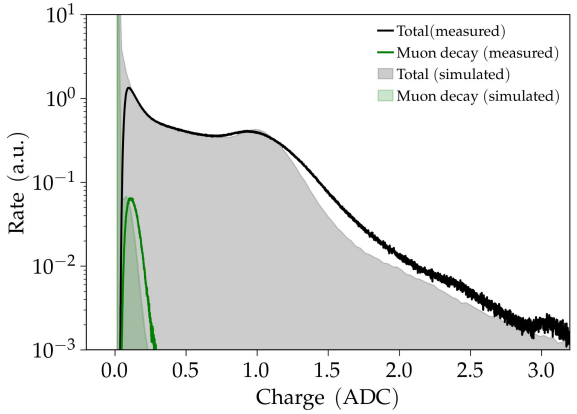






Muon decay measurements

- Low-energy muons decay within the water volume
- Observed in all our WCD
- **Large improvement in the signal-to-noise ratio and energy calibration**
- **An educational tool for associated Universities**



L. Otiniano et al., RICH'22, next week at Edinburgh

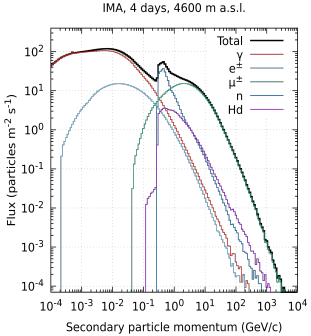
Conclusions: ARTI/OneDataSim

>10^{13} simulated EAS in 1.6

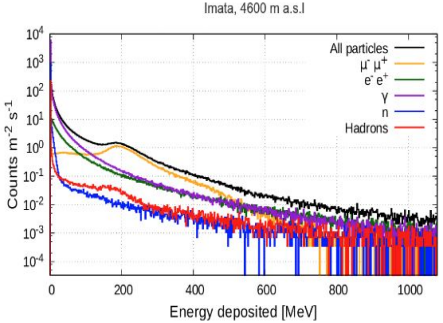
Latin America Giant Observatory

lagoproject.net

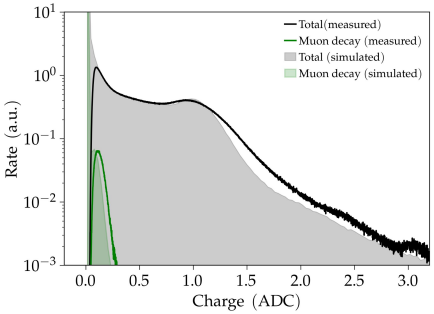
Functional, fully-integrated, and publicly-available cloud-based FAIR-compliant implementation



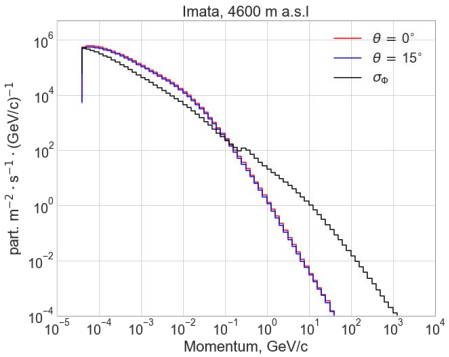
Secondaries at ground



WCD charges histograms



Muon decay and Michel spectrum



Site characterization for Gama sources

Expected flux of secondaries and the corresponding signals at any particle detector installed in any place in the World and under realistic atmospheric and geomagnetic conditions.

ARTI/OneDataSim in astroparticle physics: sites characterization, new detector designs, signal-to-background ratios…

ARTI/OneDataSim in other fields: volcano muography, mining prospecting, underground labs, errors at HPC, doses in flights, personal and safeguard protection, space exploration… (see R. Mayo-García, this conference)

ARTI/OneDataSim is publicly available: LAGO GitHub repository and in the EOSC Marketplace

hernan.asorey@iteda.cnea.gov.ar

Thank you for your attention

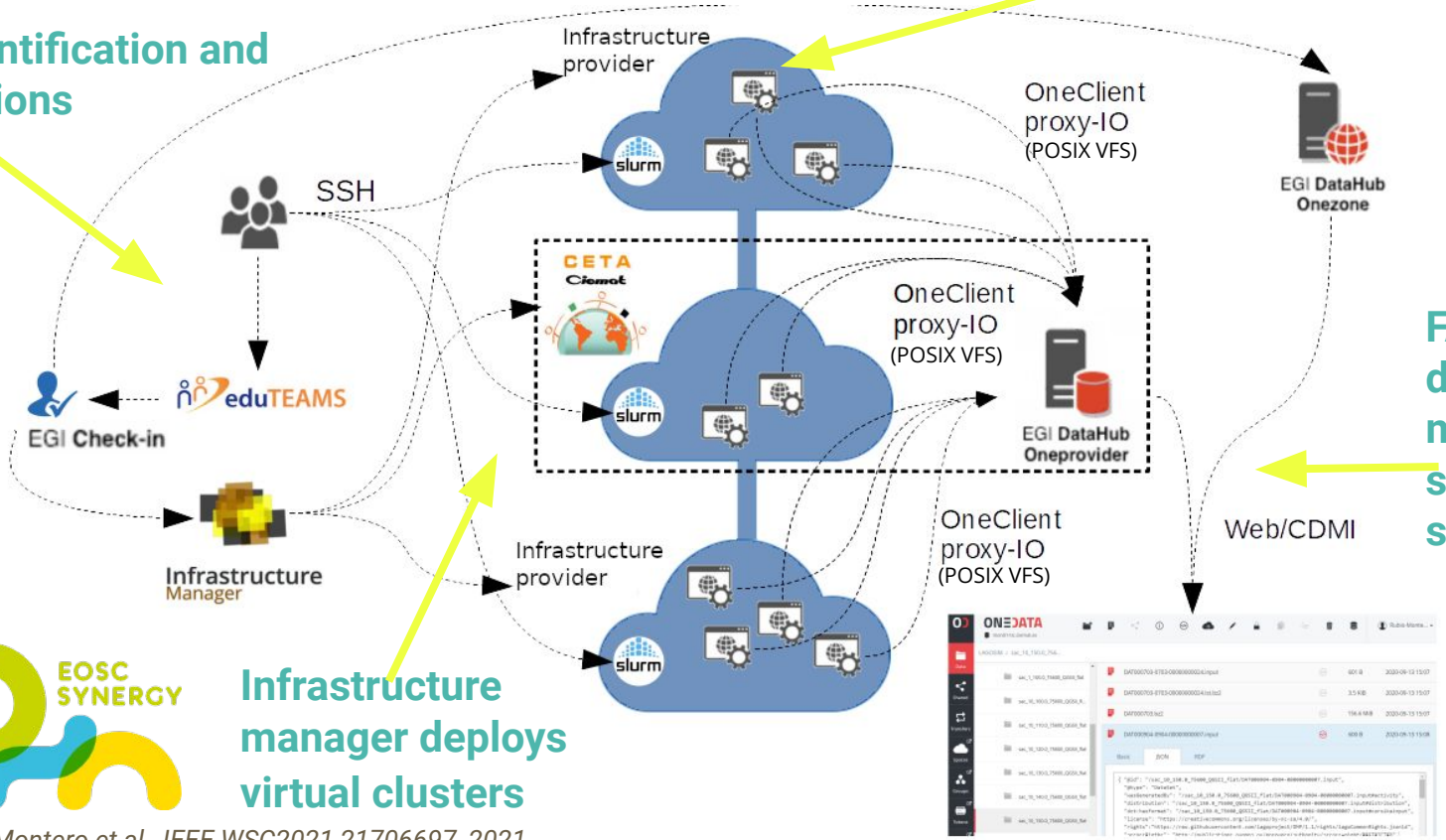
Simulations in EOSC resources

onedataSim images pulled from DockerHub and executed on nodes

User identification and permissions

FAIR compliant data and metadata stored at cloud servers

Infrastructure manager deploys virtual clusters



EOSC SYNERGY

AJ Rubio-Montero et al., IEEE WSC2021 21706697, 2021

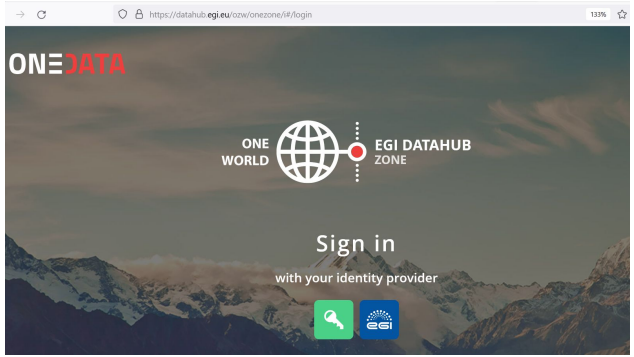
Where are the meta & data stored?

ONEDATA is an object-based distributed FS focused on cloud and HPC:

- offers unified data access and roles across globally distributed environments;
- manages metadata and expose them to harvesters.

EGI/EOSC supports the “DataHub” instance that:

- allows extending the capacity by adding providers around the World;
- asks for (Handle.net) PiDs;
- enable pan-European IdPs and translate Virtual Organisation roles.



https://datahub.egi.eu