Cherenkov Telescope Array: From Experiment to Observatory

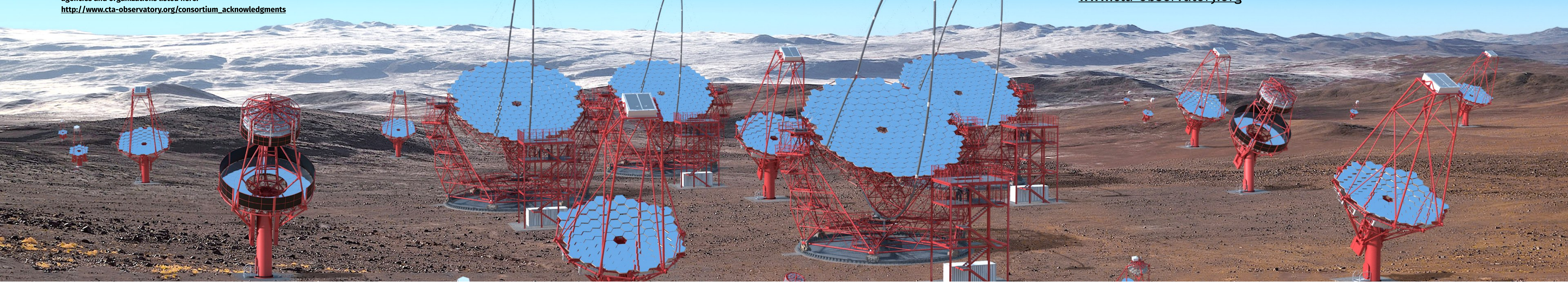
cta cherenkov telescope array

the observatory for ground-based gamma-ray astronomy

This work was conducted in the context of the *CTA Consortium*

We gratefully acknowledge financial support from the agencies and organizations listed here: http://www.cta-observatory.org/consortium_acknowledgments

Karl Kosack†, For the *CTA Observatory* and *CTA Consortium*
www.cta-observatory.org



Very-High-Energy gamma-ray astronomy is now moving from a set of independent "experiments" managed by closed collaborations to an open observatory model. The Cherenkov Telescope Array (CTA) will be the first ground-based gamma-ray **observatory** for gamma-ray astronomy at very-high energies and will produce **10s of PB of raw gamma-ray data per year,** which must be processed into a much smaller volume of **scientific data** that must be **publicly accessible** to the worldwide astronomical and particle physics communities. Here we present our plans for processing this data and how we expect to disseminate the final products to users by using and adapting **existing astronomy standards**.

What is CTA?

The Cherenkov Telescope Array (CTA) is a next-generation ground-based gamma-ray observatory, sensitive to photons in the **tens of GeV to 100s of TeV energy range**. CTA will have a sensitivity roughly **10x better** than existing ground-based gamma-ray instruments like H.E.S.S., MAGIC, and VERITAS. CTA uses the stereo *imaging atmospheric Cherenkov technique* to detect gamma rays: gamma ray photons are converted into ***extensive air showers*** in Earth's atmosphere and CTA uses an array of optical telescopes to detect the faint UV-blue **Cherenkov light** generated by particles in these air showers and to reconstruct the energy and direction of the original photon.

more info at ***www.cta-observatory.org***

Challenge 1: *Change of Culture:* Particle Physics → Astronomy

The currently operating 3rd generation Imaging Atmospheric Cherenkov Telescopes (IACTS, namely H.E.S.S., MAGIC, and VERITAS) were all born of the particle and astro-particle physics community. They operate as ***experiments*** driven by *closed collaborations* of scientists from around the world. In particular, they:

- use **proprietary data formats.** In some cases even within the same collaboration, different analysis chains produce incompatible outputs.
- The **data are not publicly available:** only members can access data, (other than a small subset of very high level data from published papers)
- Members are used to **full control**: access to all levels of data, and ability to change analysis code at will.

Currently Operating Imaging Atmospheric Cherenkov Telescope Arrays

Joining the Astronomy Community
There is momentum in all three current IACTs toward data standards and open analysis tools, heavily borrowing from the larger **astronomy community**. This includes the export of medium-to-high-level data into FITS and open VO formats, from event-lists and instrumental response functions, to data cubes, spectra, and light-curves.

CTA Observatory will both benefit from this work, and is working to drive the standardization process for the future!

Challenge 2: *Open Data and Analysis Tools*

Open Science Tools

- Analysis in *Jupyter notebooks*, focusing on reproducibility and traceability
- **Open-source** and **Open-development** with modern development practices (currently Github + Gitlab, continuous integration, code quality checking, unit test coverage testing, code reviews)
- The low-level pipeline software that process raw data into science products will also be open-source (for the current prototype, see github.com/cta-observatory/ctapipe)

Open Formats

- Event lists and Instrument Responses: in **FITS formats** inspired by X-ray and HE Gamma ray (Fermi-LAT) communities → see github.com/open-gamma-ray-astro/gamma-astro-data-formats
- Catalogs distributed to the community in FITS format

Virtual Observatory

- **Science archive** and **Science Alert Generation** systems will be **IVOA compliant** → easy data discovery with tools like Aladin, transient distribution via VOEvent
- Strong participation in defining IVOA ***provenance* data model** to meet CTA data provenance tracking requirements.

γπ A **Python** package for **gamma-ray** astronomy

ctools cherenkov telescope array

Two prototypes being developed for open science tools for CTA: **GammaPy** (gammapy.org) and **ctools** (cta.irap.omp.eu/ctools/). Both process FITS event list data to produce high-level data products, inspired by science tools from existing astronomical observatories but tailored to gamma-ray science.

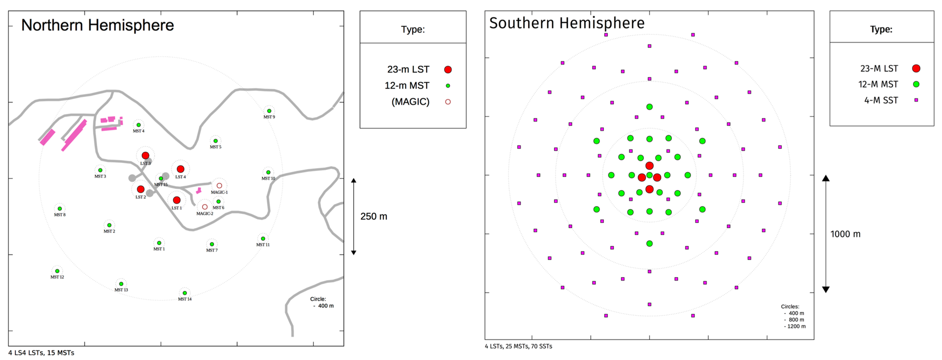
ctapipe

A prototype for the low-level data processing pipeline framework, ctapipe, developed openly (github.com/cta-observatory/ctapipe)

IVOA

Challenge 3: *An Observatory*

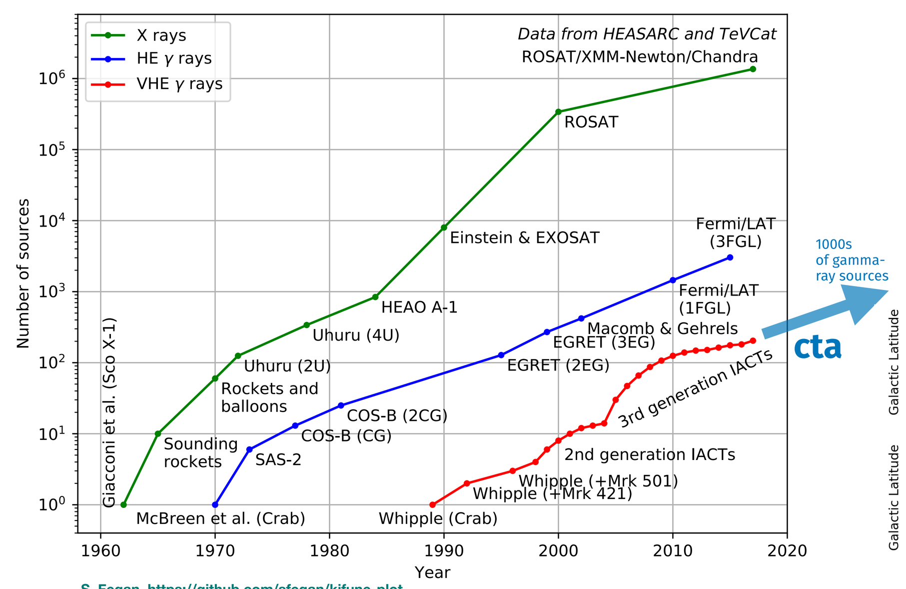
CTA will be the first ground-based gamma-ray observatory open to the world-wide astronomical and particle physics communities as a resource for data from unique high-energy astronomical observations. CTA will be operated as an open, proposal-driven observatory for the first time in very-high-energy astronomy.



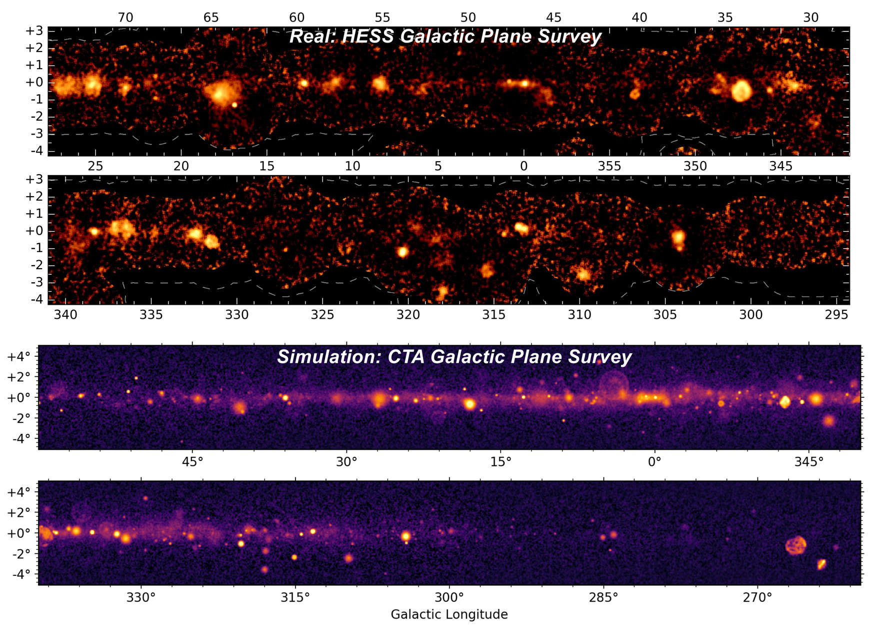
CTA Observatory full array configuration: **118 telescopes, 3 telescope classes** (large LST, medium MST, and small SST) **,** operating on **2 array sites**

Open Time: CTA will operate a Guest Observer (GO) Program, with access to proprietary observation time by submitting proposals in response to Announcements of Opportunity (AOs).
Key Science: are large observing programs that ensure that some of the key science issues for CTA are addressed in a coherent fashion, with a well-defined strategy.
Science Archive Access: all CTA gamma-ray data will be openly available, after a proprietary period.

During the first phase of operation, observation time will be split between guest observer time and KSPs, such as large-scale surveys aimed at providing legacy data sets



S. Fegan, https://github.com/sfegan/kifune-plot , original by Tadashi Kifune 1996



CTA: Zanin, R. for the CTA Consortium, PoS(ICRC2017)740
HESS GPS: H.E.S.S. Collaboration, Abdalla, H., Abramowski, A., et al. 2018, AAP, 612, A1

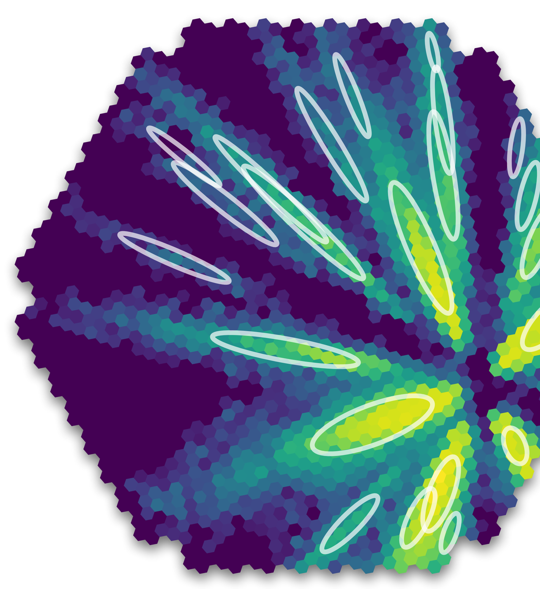
Challenge 4: Big Data

The Challenges:

- Computing resource requirements imply **distributed computing model**
- Data **volume is too big** to separate storage from computing
- we expect to **re-process *all* data** each year with new techniques, each time requiring more resources

Computing resource management
The complexity of the data processing requires a flexible system to manage computing resources and to ensure that data are processed in the same place as they are stored. It should support:

- diverse **computing centers**
- **Grid** computing (e.g. EGI, www.egi.eu)
- the addition of **cloud resources** on demand
- the management manage ***complex workflows*** with parallelization and merging/reducing



An example of raw data from CTA: this is the sum of time-integrated images (over a window of *20 ns*) of the Cherenkov light from a single gamma-ray induced air-shower, as seen from multiple telescopes. The **origin of the photon** is at the point where the axes of the ellipses cross. The **energy** is proportional to the light collected. In normal data taking, these images are recorded (with time information) at a rates around 10KHz

Bulk Data Archive:
While the final science data products are small (TBs), the raw data for CTA will include 3-4 PB/year per site along with a similar volume of simulation data needed to perform the analysis. For this reason the bulk archive must:

- Be distributed, to meet longevity and data-safety requirements
- Ensure the large volume of raw data is preserved and in an accessible format for the lifetime of CTA (30 years!)
- Support fast re-processing of old data and be tightly connected to computing resources

DIRAC THE INTERWARE
RUCIO
ONEDATA
GAMAS
XDC eXtreme DataCloud

† **Karl Kosack**
CEA Paris-Saclay
Department of Astrophysics
/ CTA Observatory

CTAO ***Data Processing and Preservation*** *Workpackage Coordinator*