BOF Astronomy & Astroparticle Physics

MT DMA ST1 Synergy Workshop

Nov 8-10 2023

!BOF?

BOF = special interest sessions?

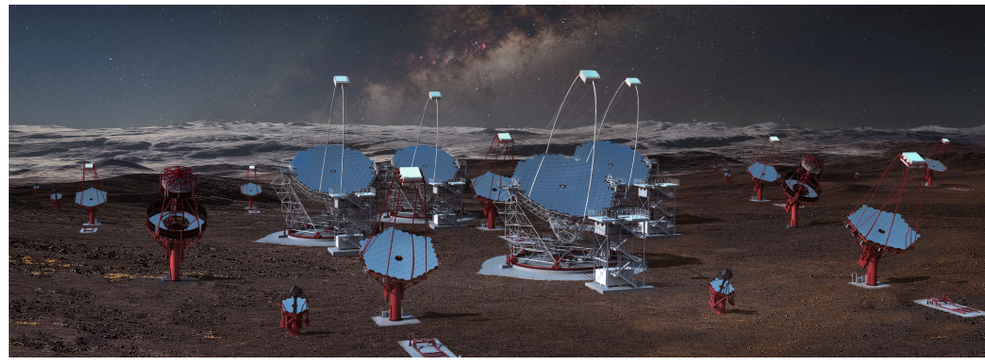
Gernot Maier

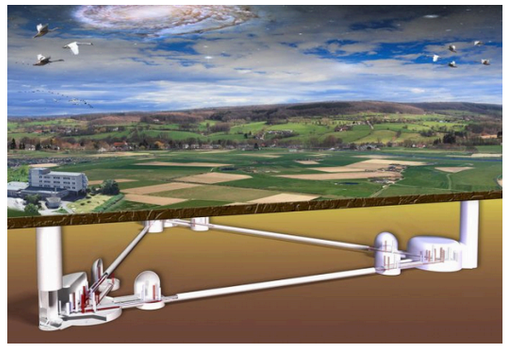
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

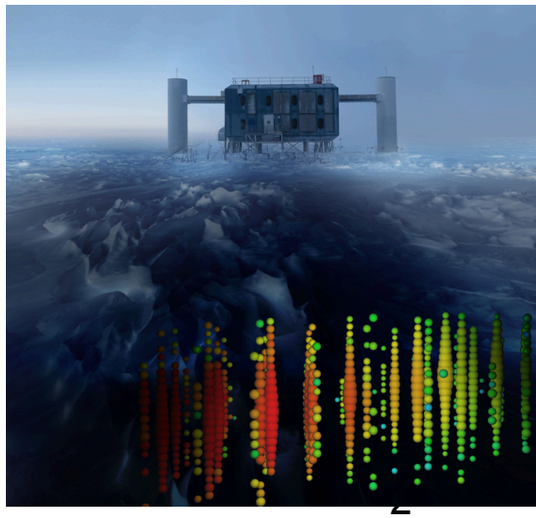
DESY.

Astronomy Large Infrastructures

- **Cherenkov Telescope Array (CTAO)**
  (gamma rays: La Palma, Chile)
- **Square Kilometer Array (SKAO)**
  (radio; South Africa, Australia)
- **LSST / Vera Rubin Observatory**
  (Optical; Chile)
- **Pierre Auger Observatory**
  (Cosmic Rays; Argentina)
- **Einstein Observatory**
  (Gravitational waves; “Europe”)
- **IceCube, Lofar, MeerKAT, H.E.S.S., MAGIC, VERITAS, Ultrasat, …**

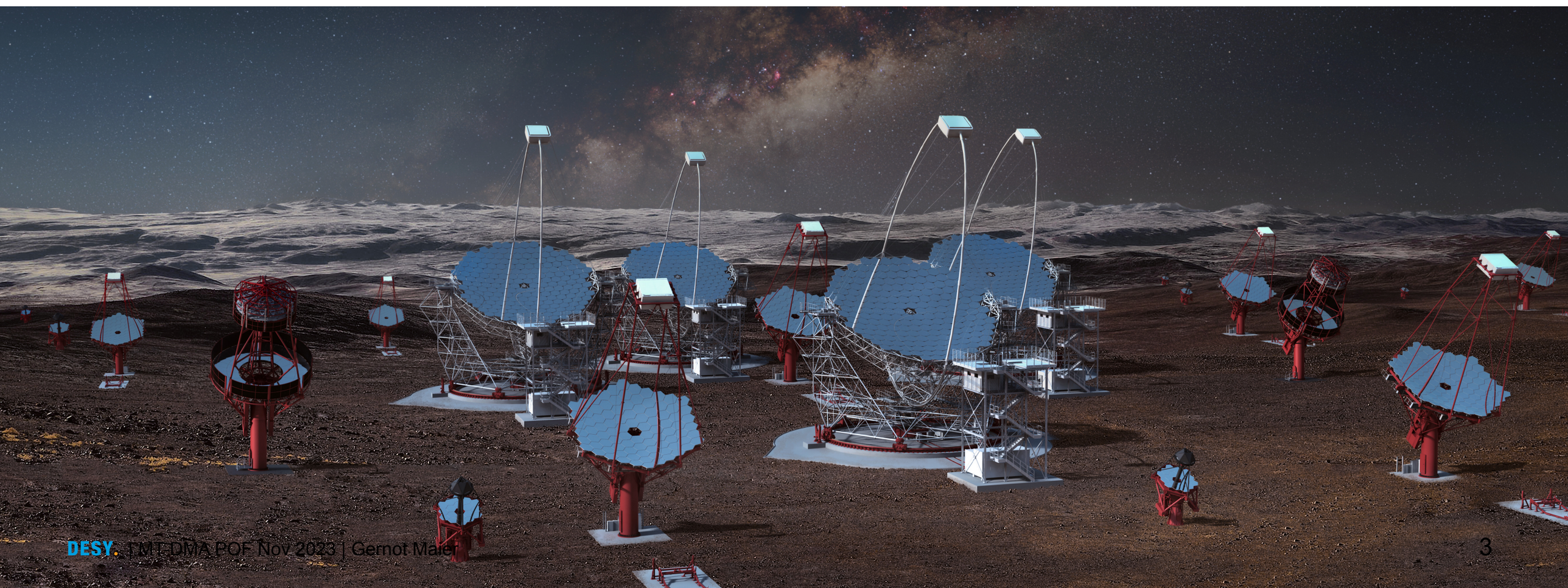






CTAO

- next-generation gamma-ray observatory
  (5-10x increase in sensitivity; broader energy range 20 GeV to 300 TeV; precision)
- two observatories (La Palma / Chile)
- open observatory serving user community (proposal driven)
- follows FAIR principles (open software, data formats)
- full operations 2028 (30 year life time)
- 6 PB / year + 20 PB simulation data



LSST / Vera C. Rubin

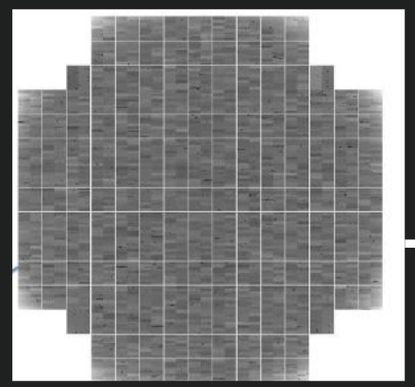
VERA C. RUBIN OBSERVATORY

Legacy Survey of Space and Time

See talk by Fabio Hernandez in Data Management track earlier today

Bolton, CHEP 2023

Survey runs 2025-2035



raw images

alerts

science-ready images

astronomical catalog

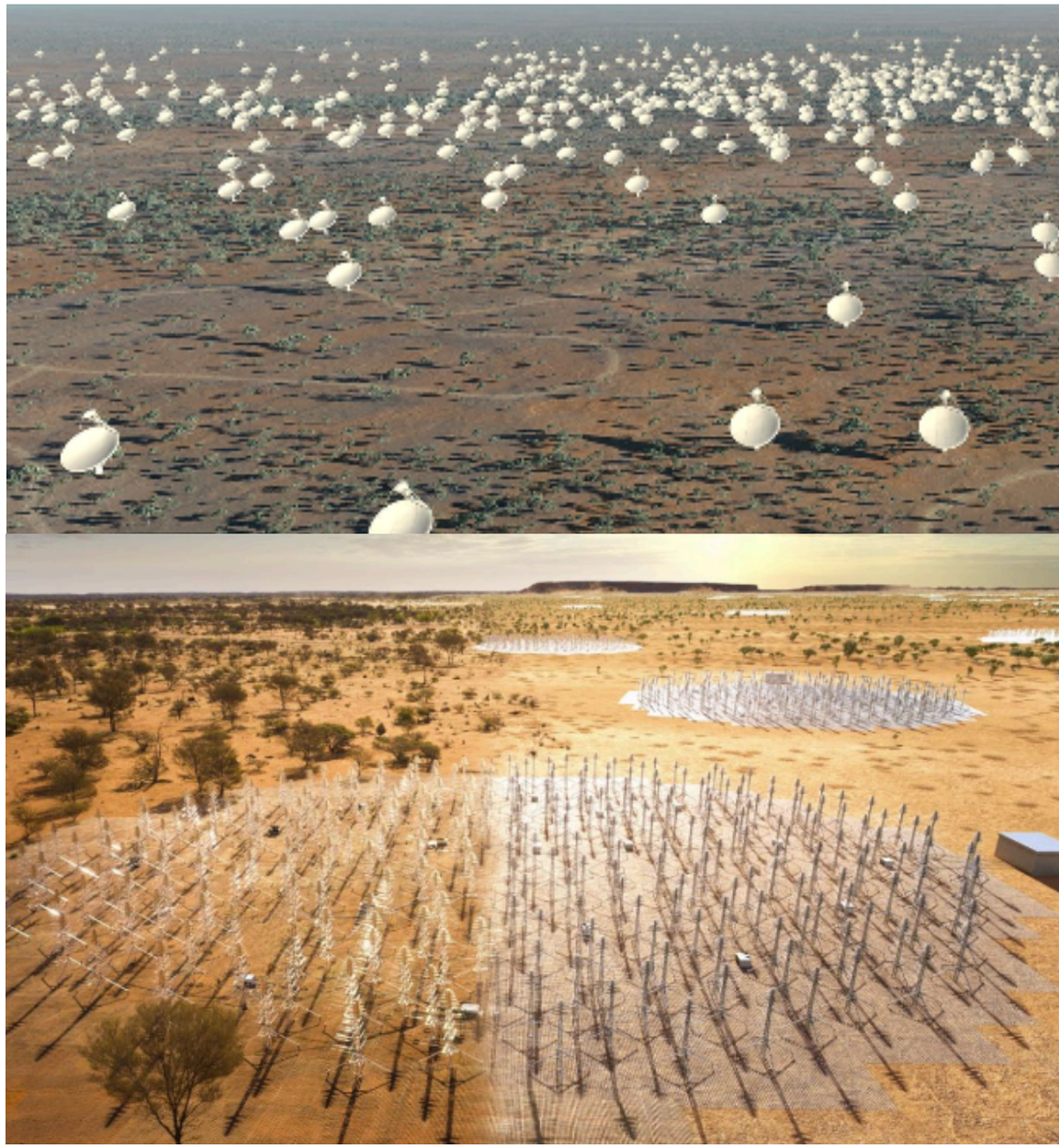
science collaborations

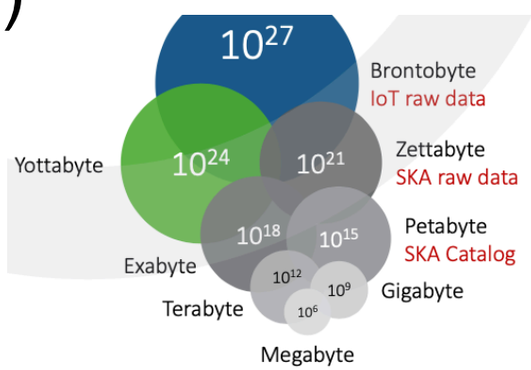
LSST aims to deliver a catalog of 20 billion galaxies and 17 billion stars with their associated physical properties

Survey strategy developed by science consortia;
Not open for proposals.

SKAO

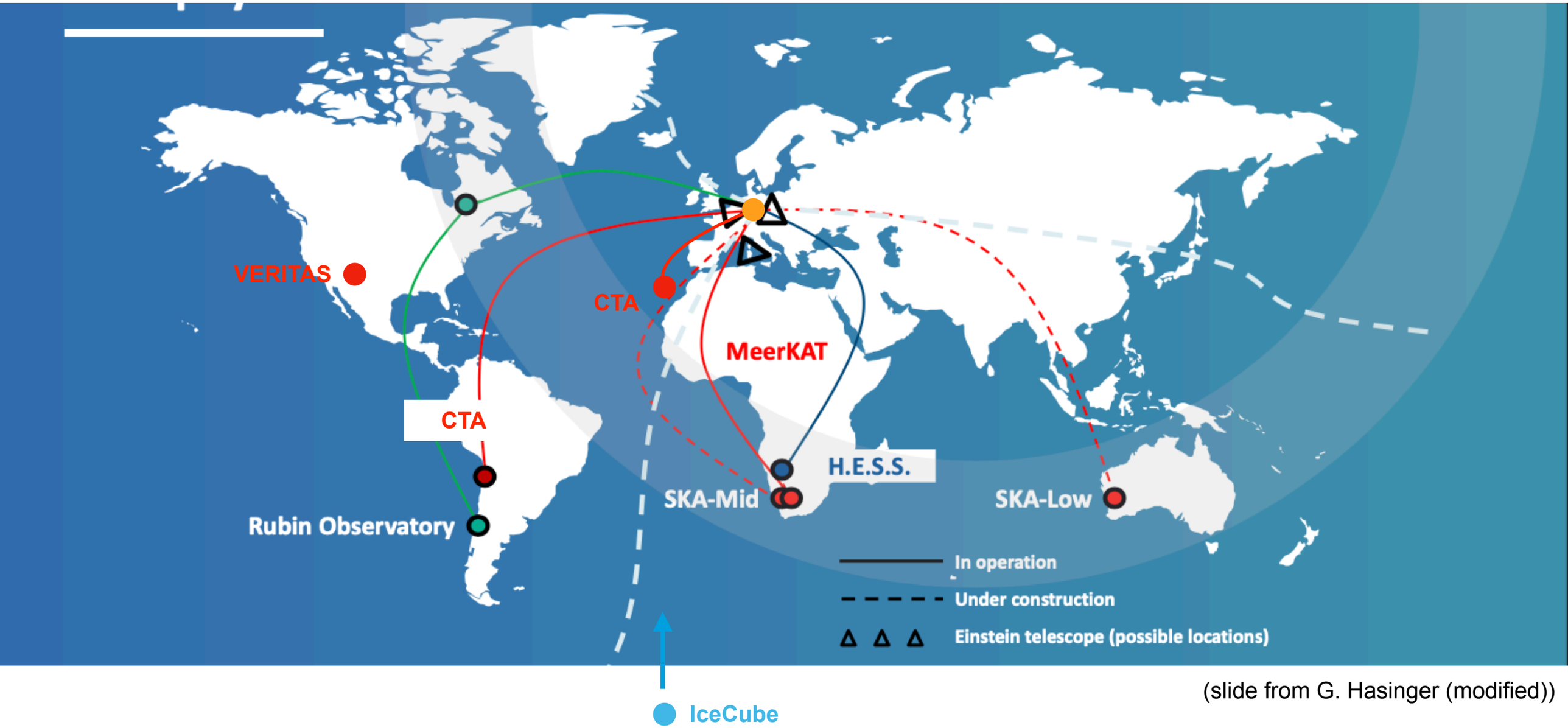
- largest radio observatory
  - 200 dishes (~GHz) spread in Karoo desert (South Africa)
  - 1300,000 antennas (100 MHz) in Western Australia
- operational 2026-2028
- open to scientists from all member states (proposals)
- science archive
  - ~600 PB/year (observatory)
  - ~200 PB/year (user generated)





From Sites to Data Centers

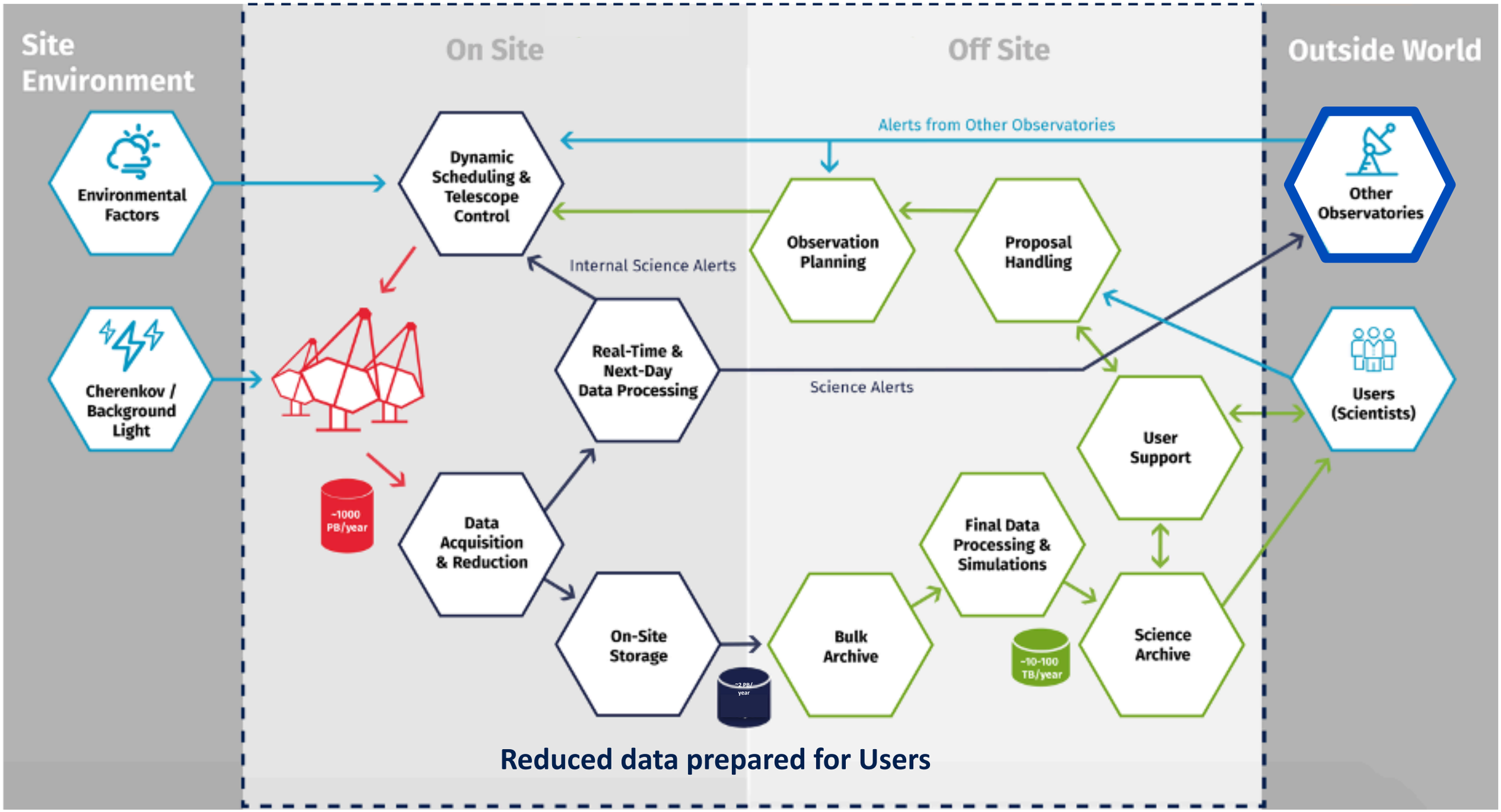
Data acquisition in “remote place”: mix of onsite computing plus (tiered) off-site computing.



(slide from G. Hasinger (modified))

Science Operations Flow - example CTAO

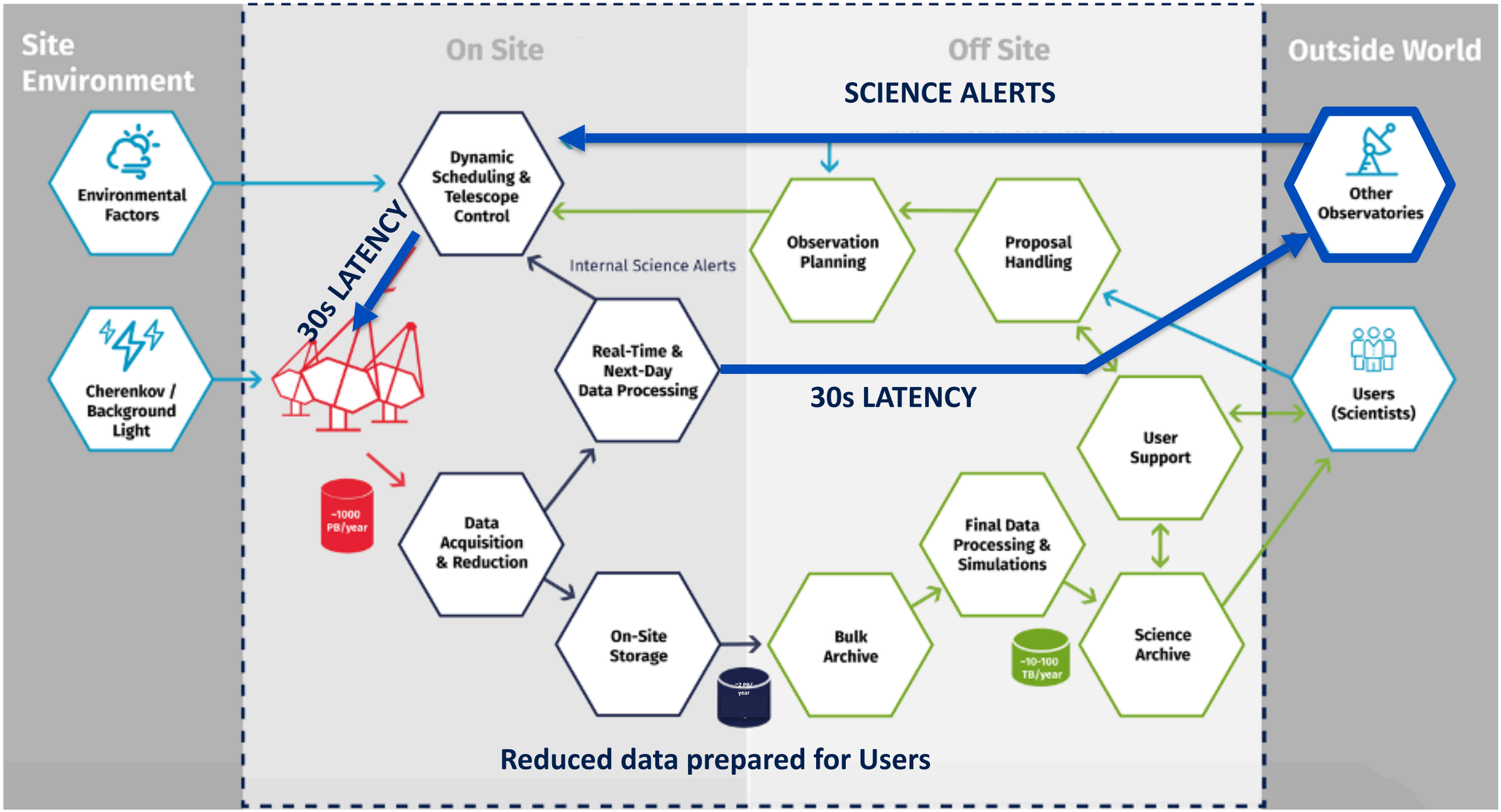
CTAO



Almost all computing executed by CTA Observatory staff.

Science Operations Flow - example CTAO

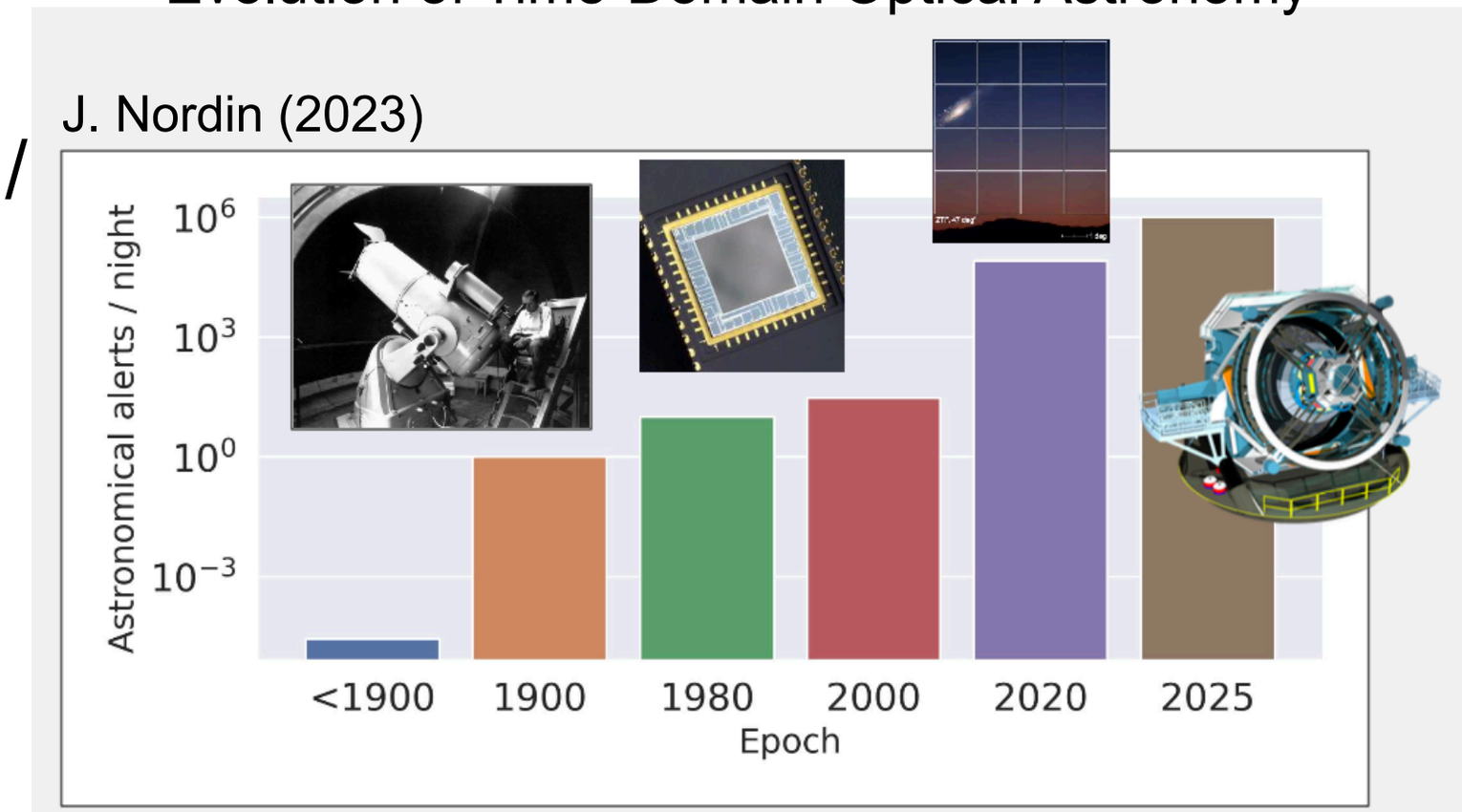
CTAO



Almost all computing executed by CTA Observatory staff.

High-throughput Observatories

- enormous increase of alerts from variable / transient objects
- require follow ups on sec to days time scale
- challenge of selection
  - expensive observing time should not be wasted
  - HPC using modern statistical methods (ML)
  - systematic processing and filtering of (non-) selection
  - dynamic (real-time) filtering of huge data streams
- challenge of combining data from many observatories

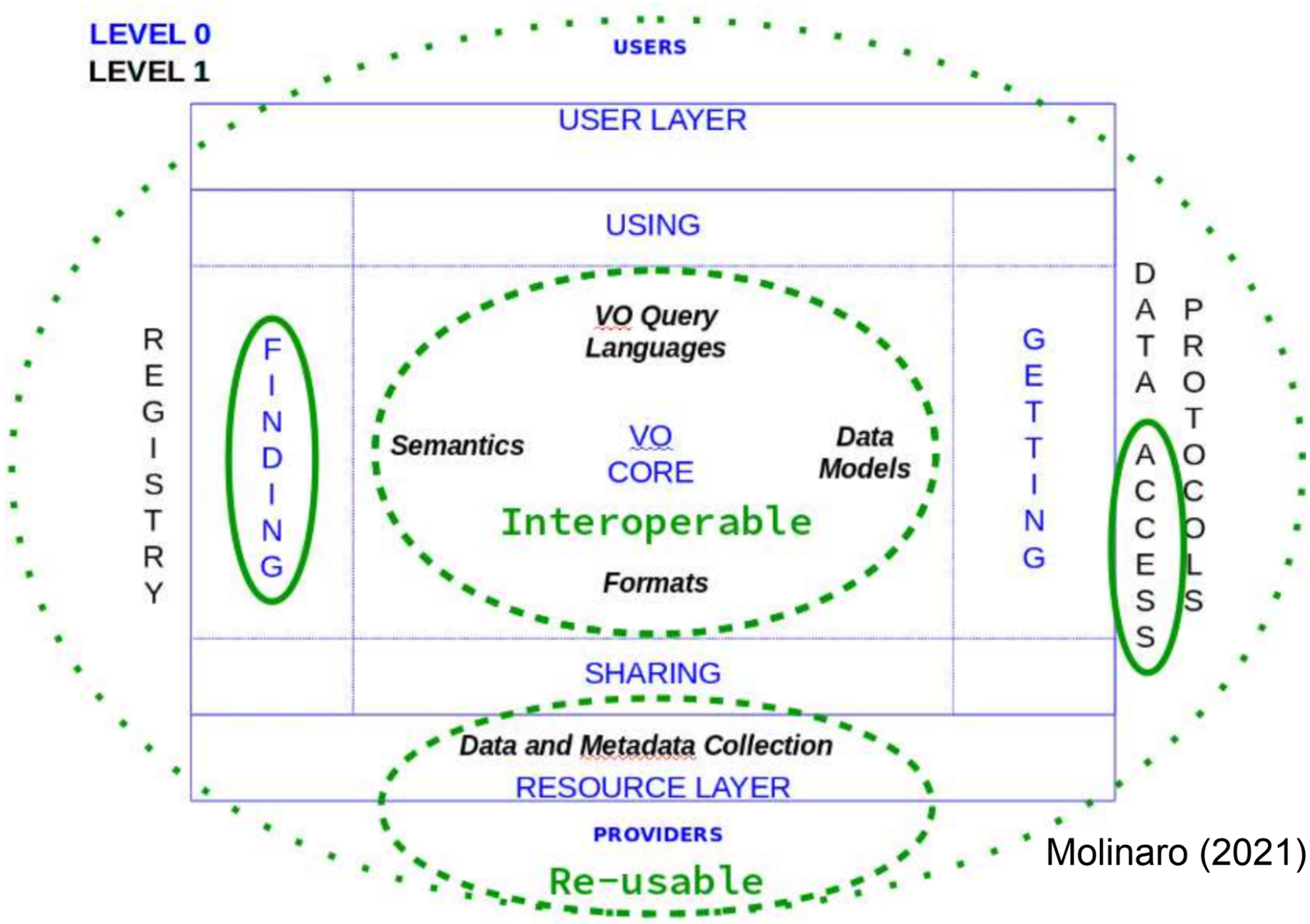


AMPEL

International Virtual Observatory Alliance

“The Virtual Observatory (VO) is the vision that astronomical datasets and other resources should work as a seamless whole.” (IOVA website)

Technological **standards** to find and access astronomical datasets

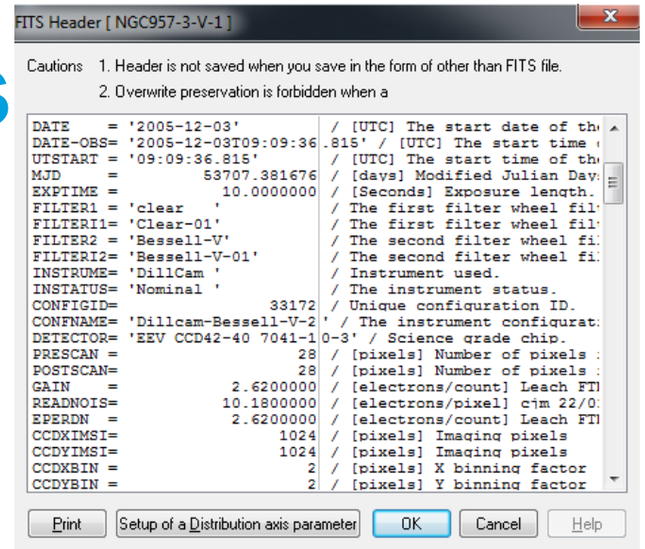


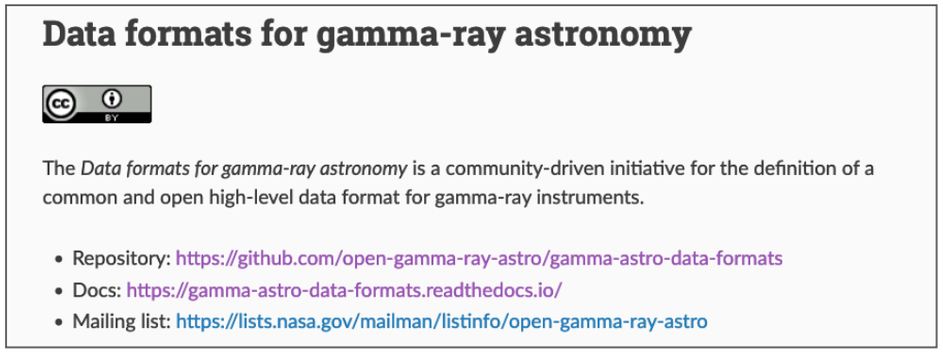
Molinaro (2021)

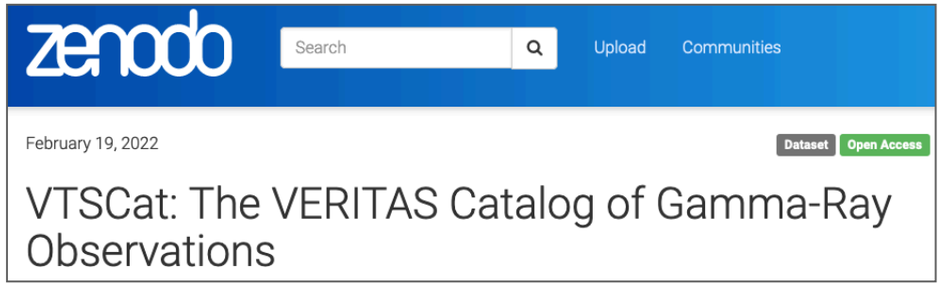
IVOA / EOSC connection established as ESCAPE project

Standardised Data Formats / Models

- combination of data requires standards
- advantage of “simple” formats
  - csv wherever possible (rare events)
  - FITS
- open data models developed in the high-energy community based on IVOA standards
  - “game changer”, as it provides long-term accessibility to data from existing observatories
- base of open-data releases
- machine readable catalogs of physics results
- (remark on time-consuming (meta-)data model development)

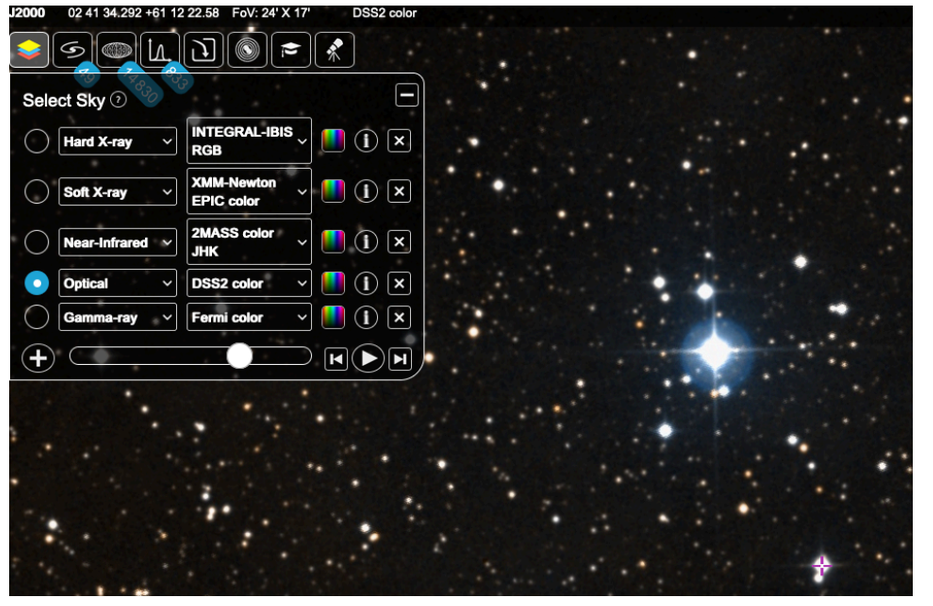


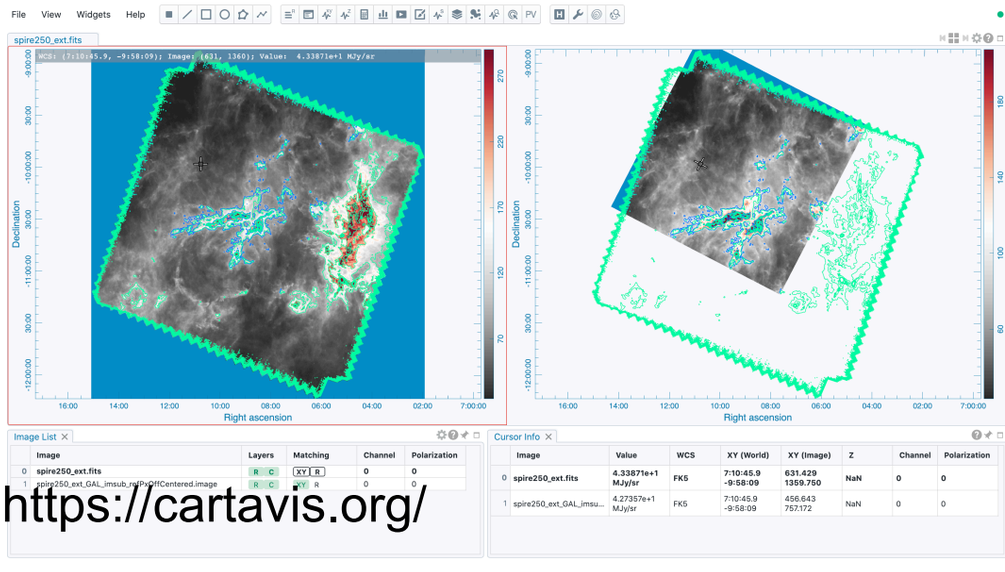




Science Data Platforms

- uniform interface to data for observatory users
  - Jupyter Notebook servers, web-/ command-line interfaces
- sharing of data and computing services
- cross match across data sets of different observatories
- high-performance / access for images with large files sizes
  (e.g., TB large images for ALMA, VLA, SKA)





https://cartavis.org/

Archives

Romaniello et al 2023

PI = scientist proposing observations

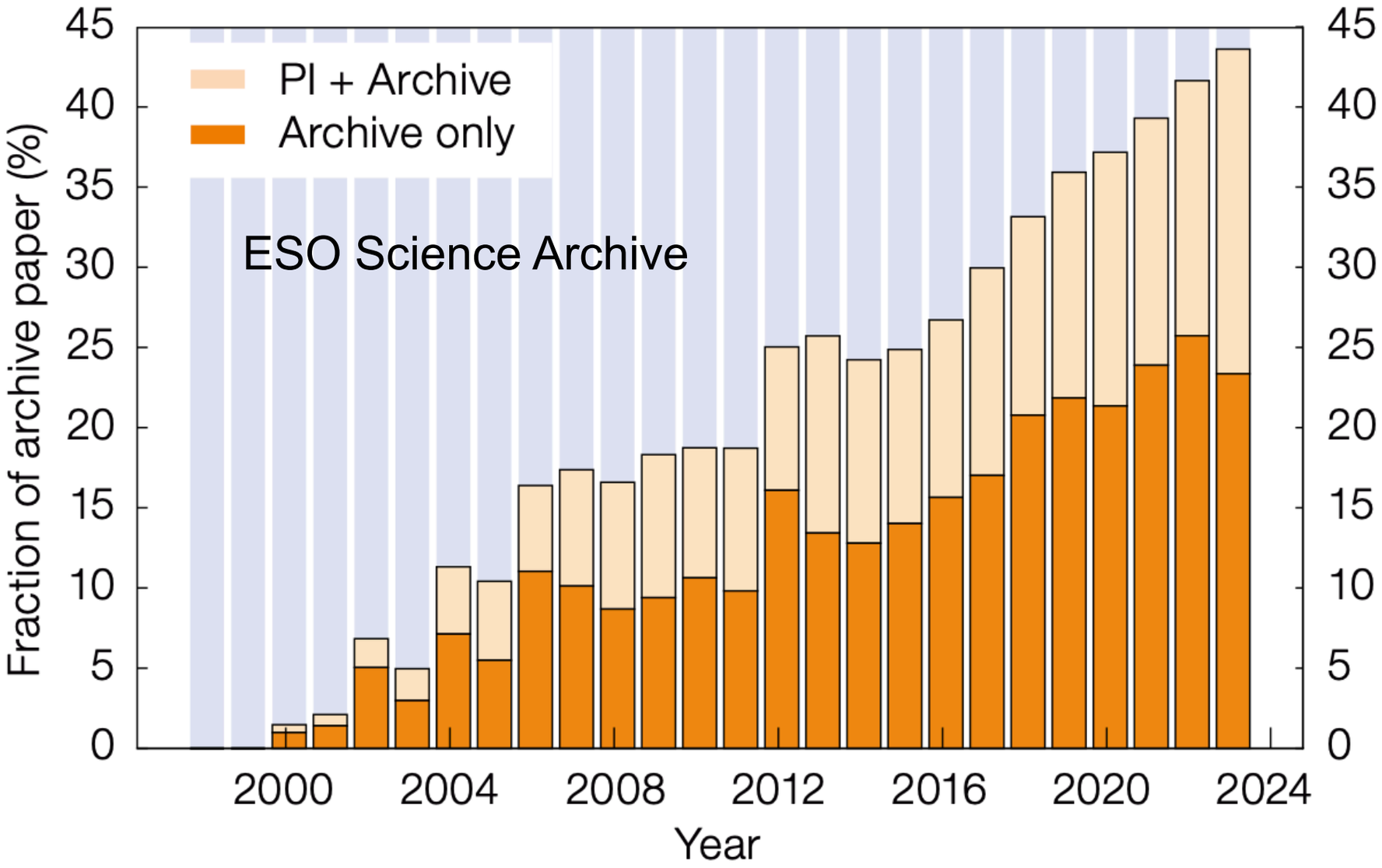
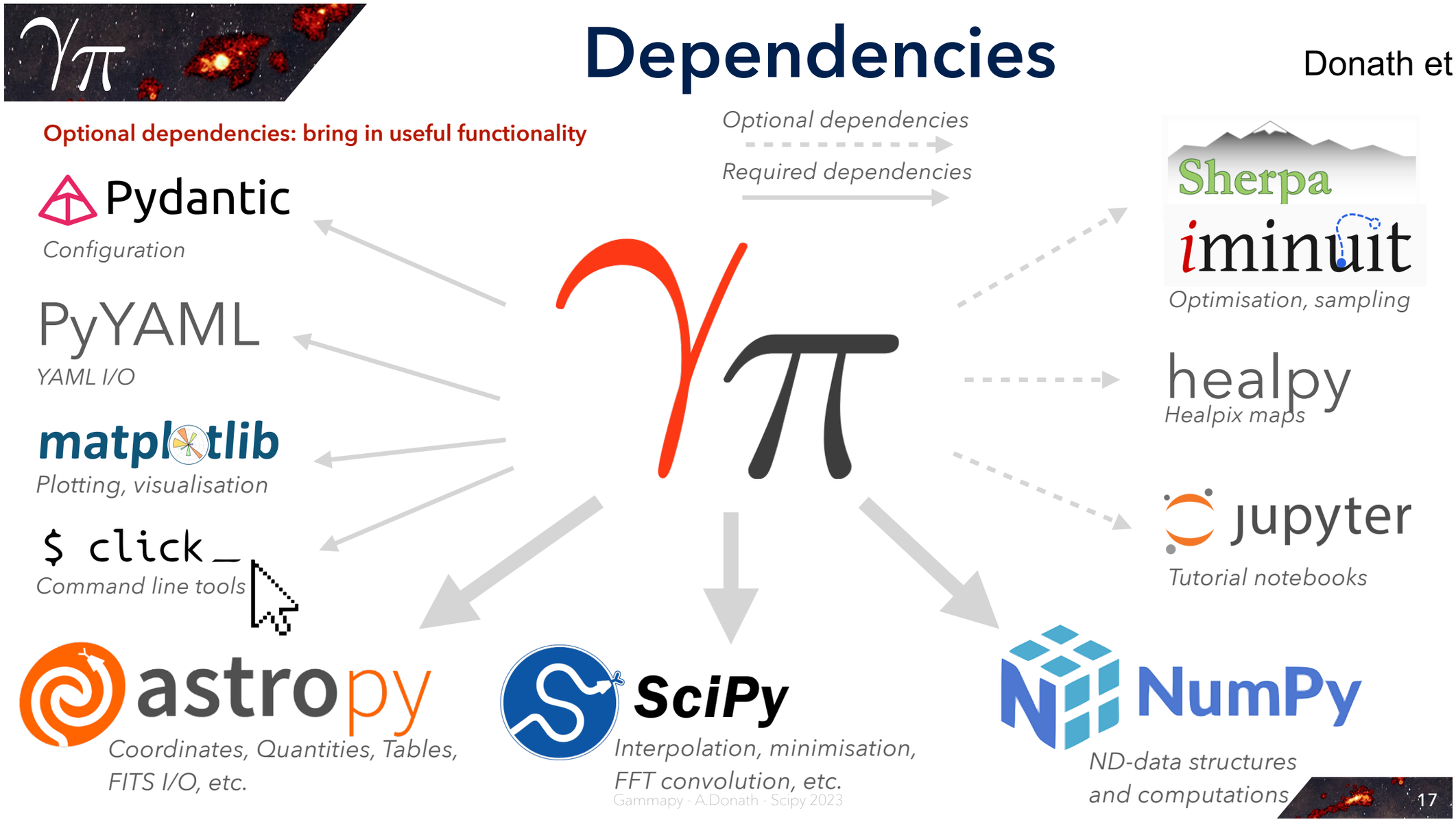


Figure 4. The fraction of refereed publications using La Silla Paranal data that made use of the archive, entirely (dark orange bars) or partially (light yellow bars). Source: ESO Telescope Bibliography[24].

Astronomical data is not getting old.
Long-term preservation and accessibility of data “forever”.
Software preservation.

Open source software development

Typical example of reliance on open-source packages:



Plus workflows (e.g., CWL) to connect data, software, computing resources and generate provenance information

Discussions.