Data diffusion for ground based gamma-ray astronomy

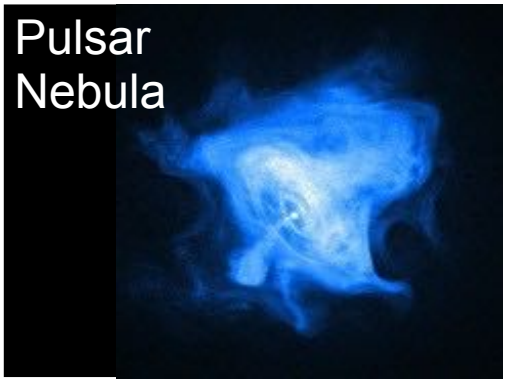
The Cherenkov Telescope Array

Mathieu Servillat

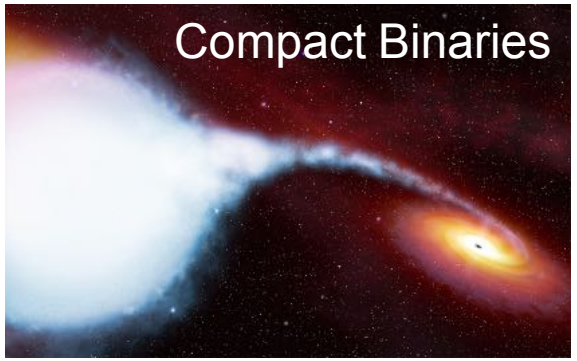
Laboratoire Univers et Théories
Observatoire de Paris
PSL Research University

ASTERICS European Data Provider Forum

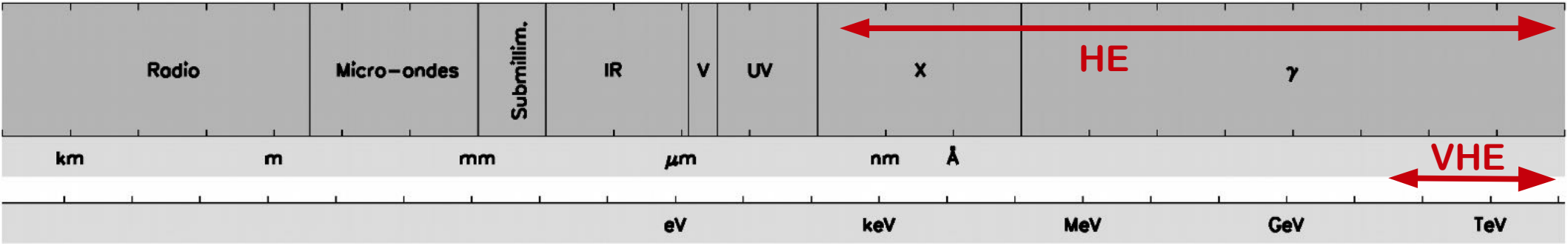
High Energy Astrophysics



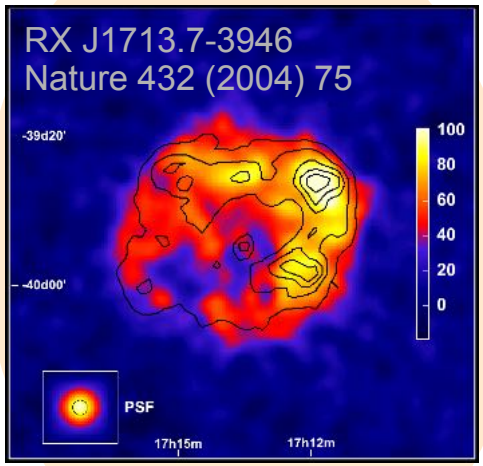
- Violent, transient, non-thermal phenomena
- Matter under extreme conditions
- Particle Acceleration
- Fundamental Physics
- Role of Black Holes in the structuration of the Universe



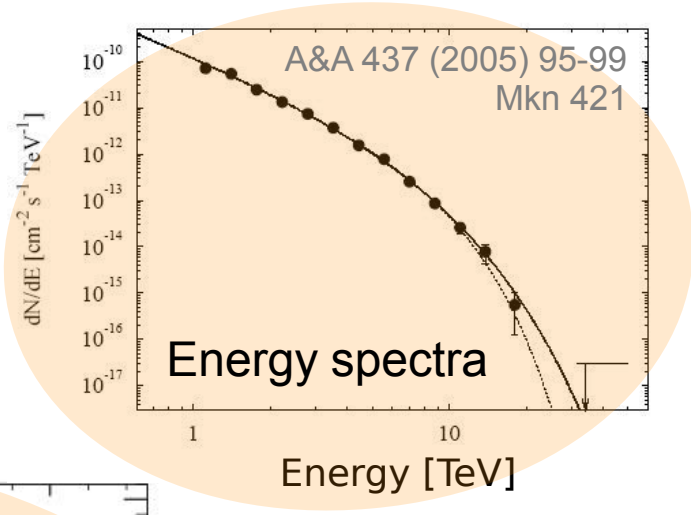
Very high energy (VHE) data

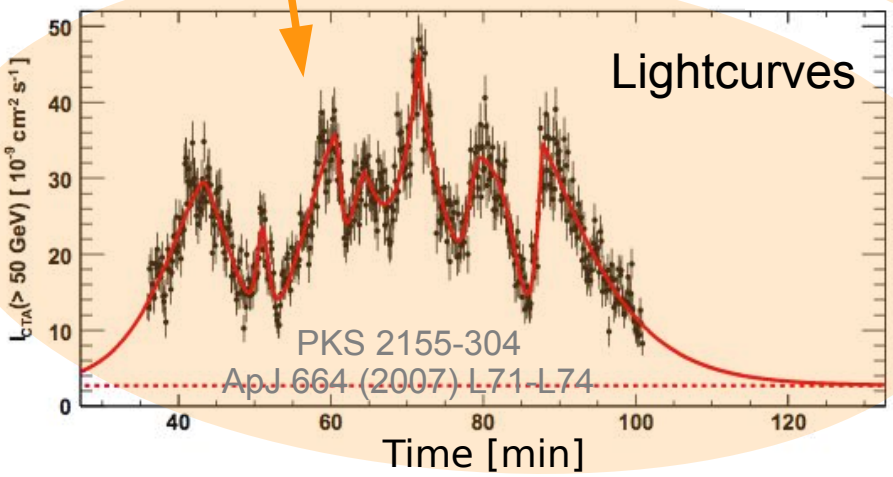


- Several orders of magnitude
- Photon counting
- Low count statistics, high background
- **Event lists** (coordinates, time, energy)

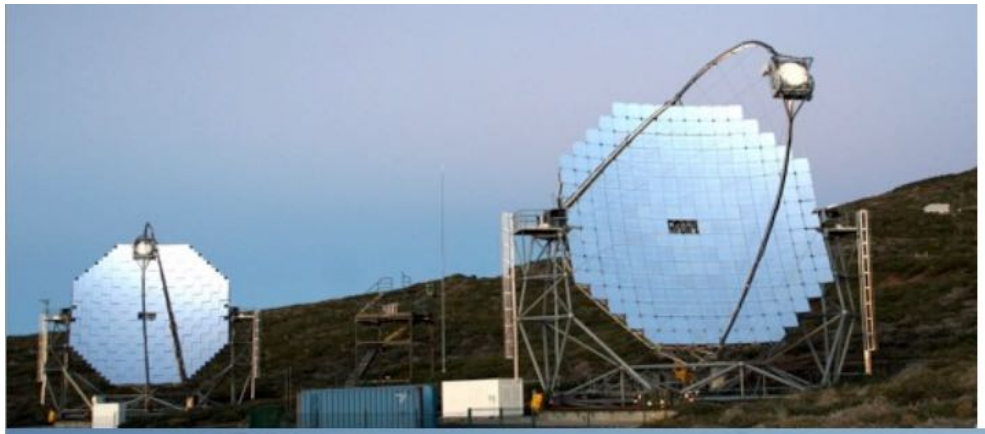


Images





Currently Operating VHE Instruments



MAGIC: located in La Palma, Spain
Since 2004: single 17m telescope
Since 2009: system of two 17m telescopes

VERITAS: located in Mt Hopkins, Arizona
Since 2007: four 12m telescopes
Since 2012: upgraded PMTs

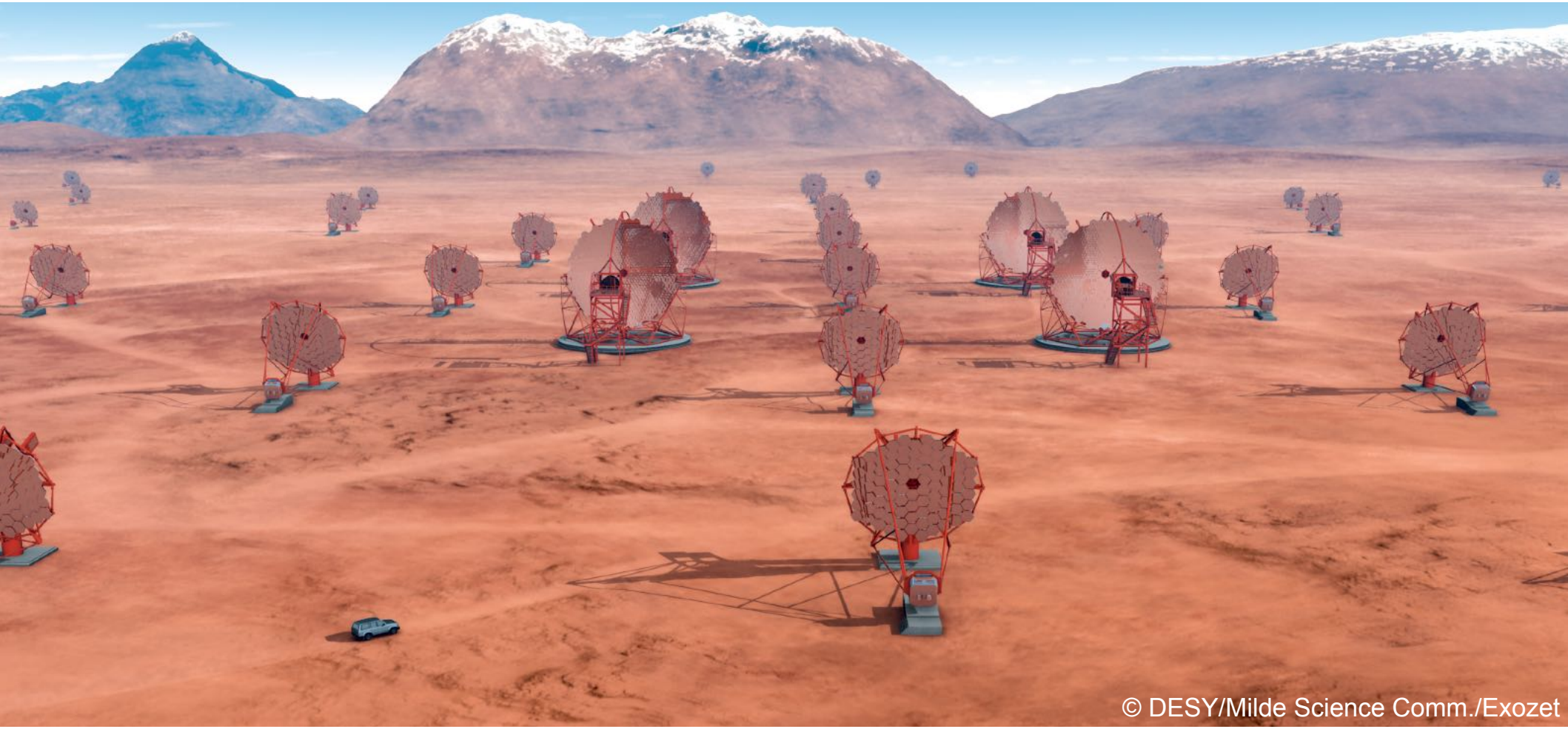
H.E.S.S.: located in Khomas Higlands, Namibia
Since 2002: four 12m telescopes
Since 2012: added 32m by 24m telescope
Since 2015: camera upgrades on 12m telescopes

@ Jeff Grube

cherenkov telescope array

Observatory

- Two arrays of **100 (South)** et **20 (North)** telescopes
- July 2015: sites selection, Chile (ESO) and La Palma
- 2016: pre-production phase
- 2018-2023: production phase
- Observatory **open** to the community



CTA Consortium

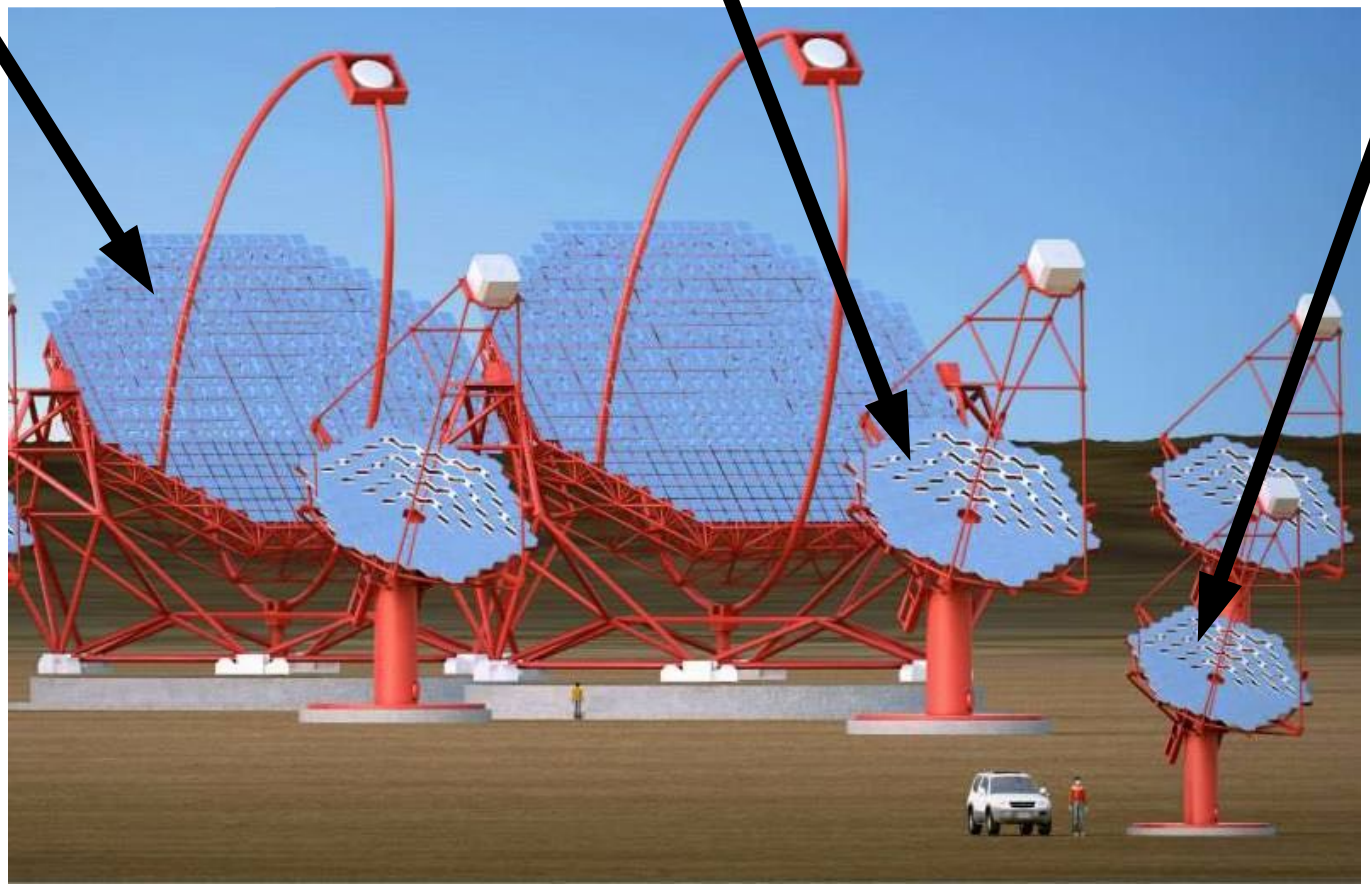
- Over 1200 members
- 200 institutes
- 32 countries

Telescope types

Low-energy section
4 x 23m **(LST)**
FoV: 4-5 degrees
E > 10s of GeV

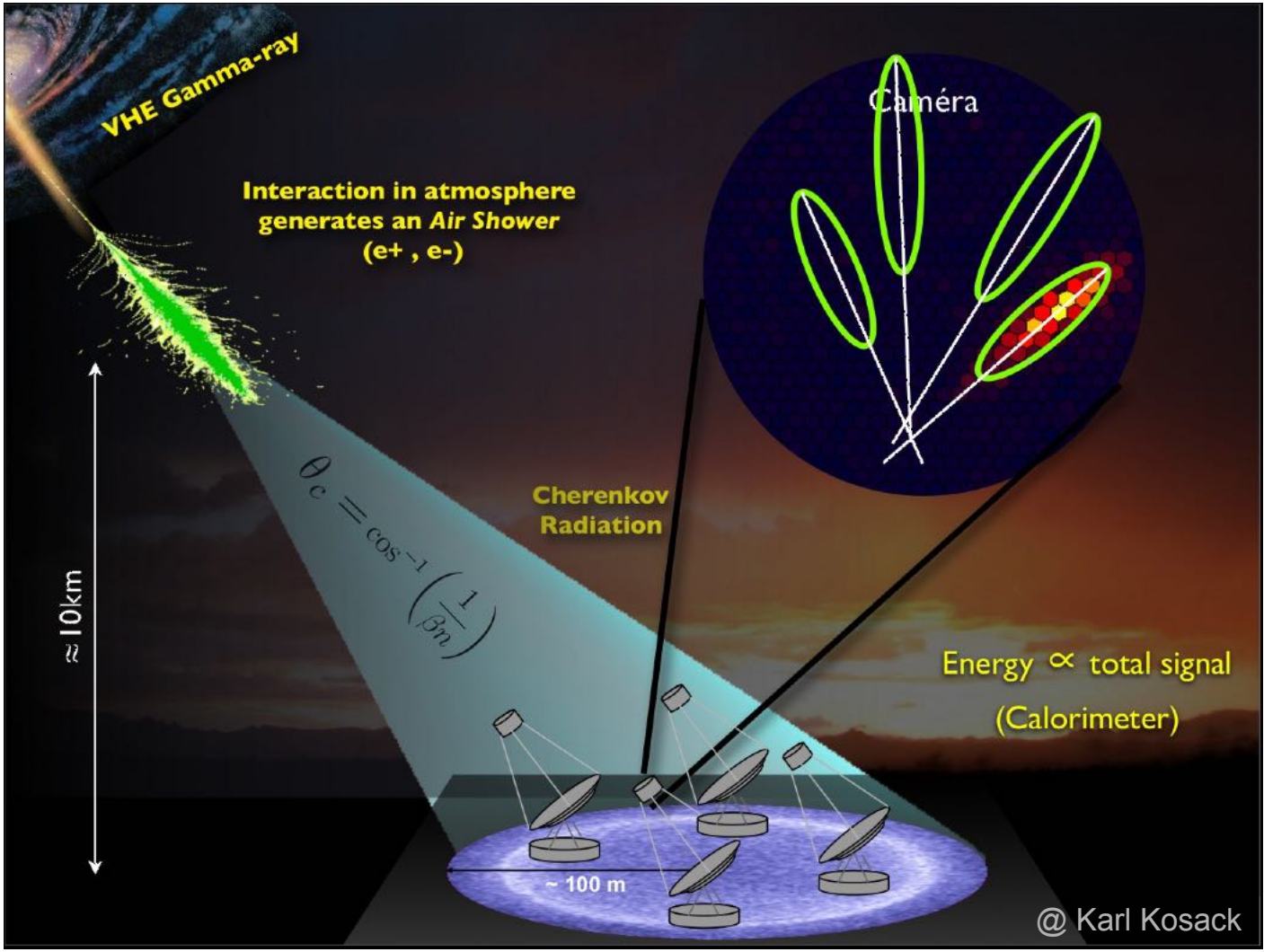
Core-energy section
23 x 12m (MST)
FoV: 7-8 degrees
100 GeV < E < 10 TeV

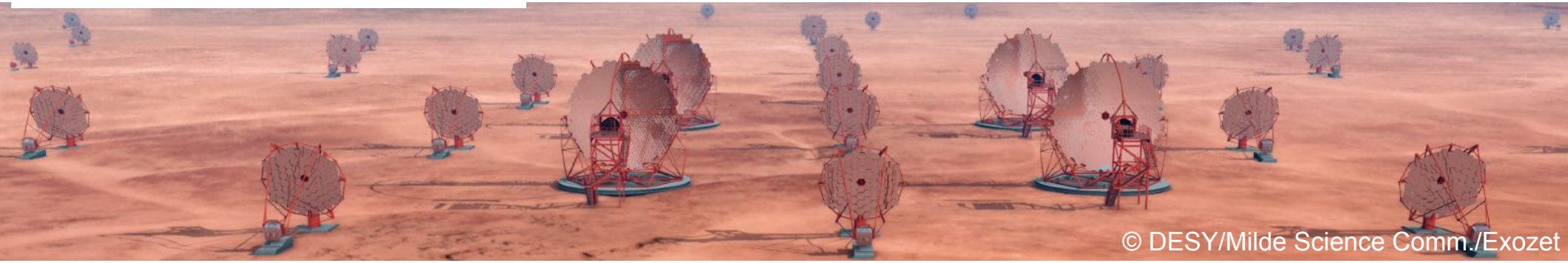
High-energy section
30-70 x 4-6m (SST)
FoV: 10 degrees
Multi-TeV energies



Cherenkov Astronomy Principles

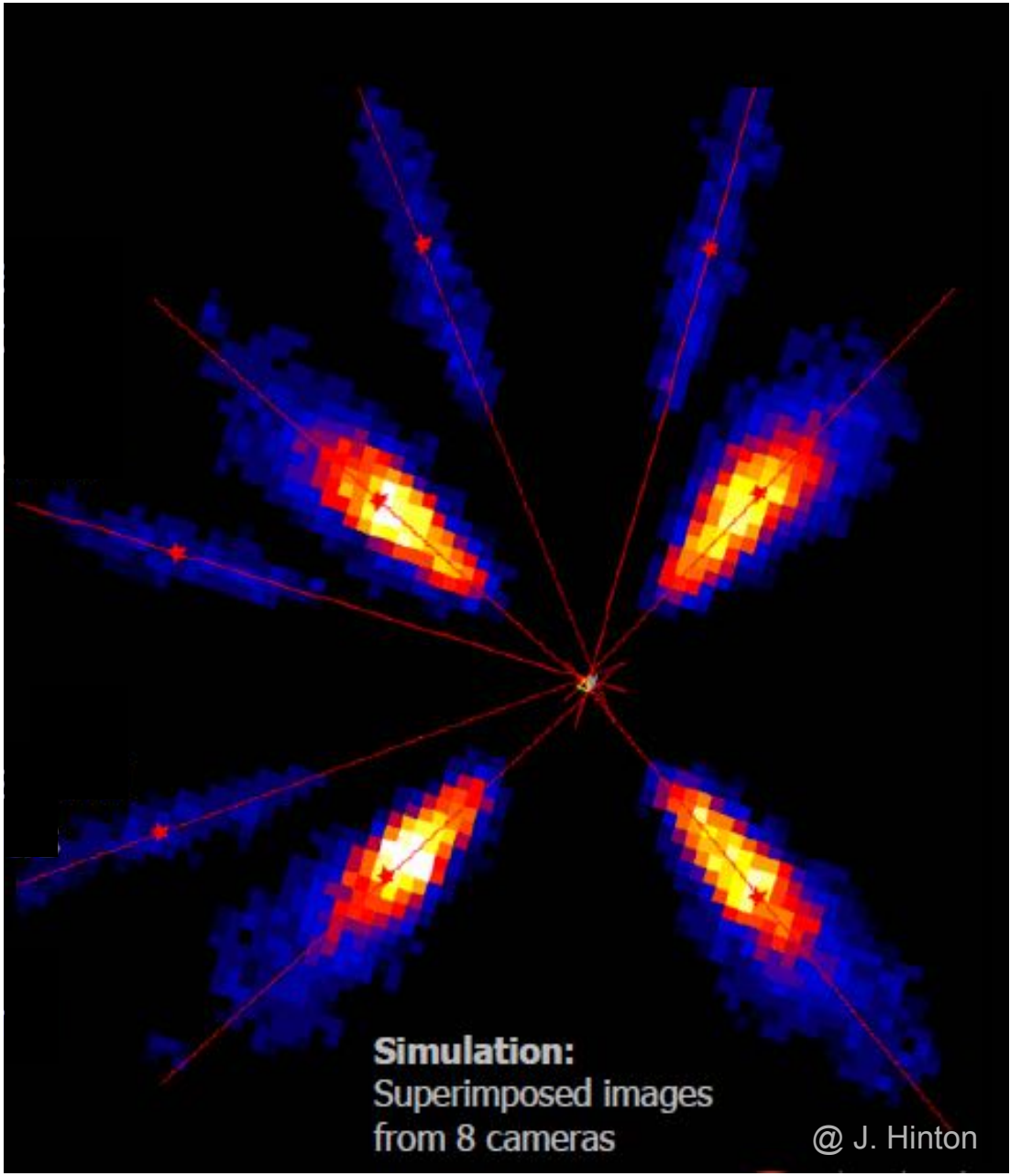
- **Dark nights** (small duty cycle)
- **Event Reconstruction**: photon, particle shower, Cherenkov light (faint, few nanoseconds)
- **Atmosphere** = calorimetre Simulations, assumptions
- **Complex Metadata**, need to be structured





Why more telescopes in an array?

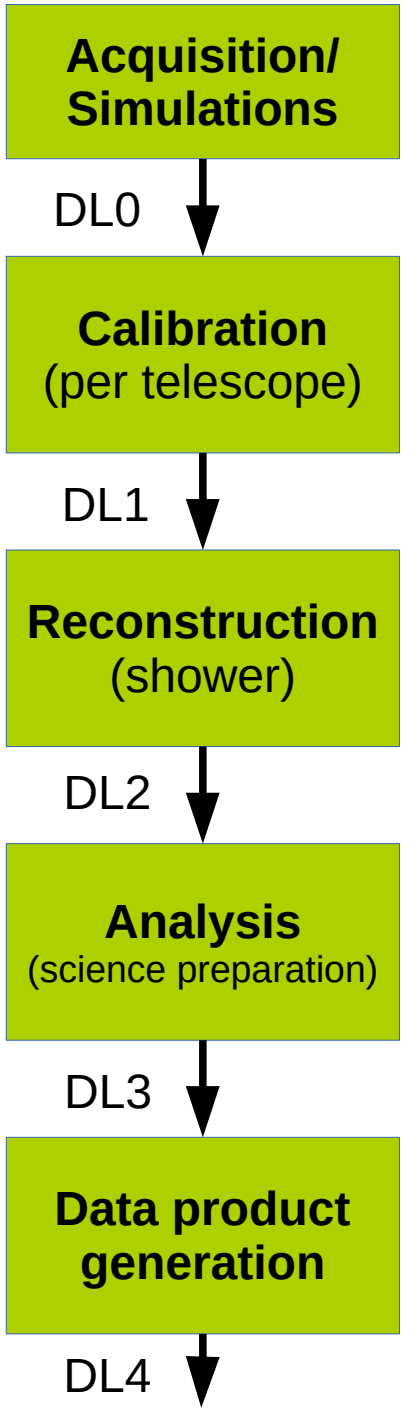
- **Larger** collection area for gamma-rays
  - More events, more photons
  - Better spectra, images, fainter sources
- **Better** events
  - More precise measurements of atmospheric cascades and hence primary gammas
  - Improved angular resolution
  - Improved energy resolution
  - Improved background rejection power



Data levels and workflow

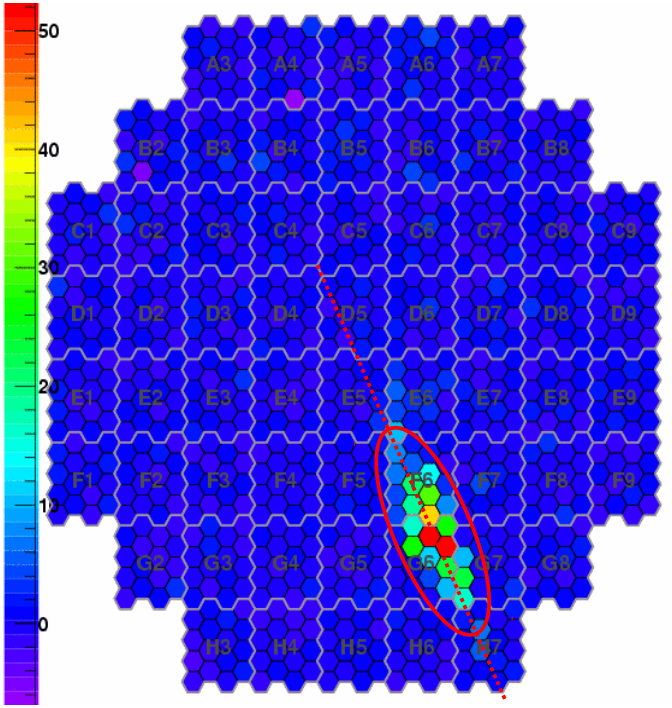
Data Level	Short Name	Description	Data reduction factor
Level 0 (DL0)	DAQ-RAW	Data from the Data Acquisition hardware/software.	
Level 1 (DL1)	CALIBRATED	Physical quantities measured in each separate camera: photons, arrival times, etc., and per-telescope parameters derived from those quantities.	1-0.2
Level 2 (DL2)	RECONSTRUCTED	Reconstructed shower parameters (per event, no longer per-telescope) such as energy, direction, particle ID, and related signal discrimination parameters.	10^{-1}
Level 3 (DL3)	REDUCED	Sets of selected (e.g. gamma-ray-candidate) events, along with associated instrumental response characterizations and any technical data needed for science analysis.	10^{-2}
Level 4 (DL4)	SCIENCE	High Level binned data products like spectra, sky maps, or light curves.	10^{-3}
Level 5 (DL5)	OBSERVATORY	Legacy observatory data, such as CTA survey sky maps or the CTA source catalog.	10^{-5} - 10^{-3}

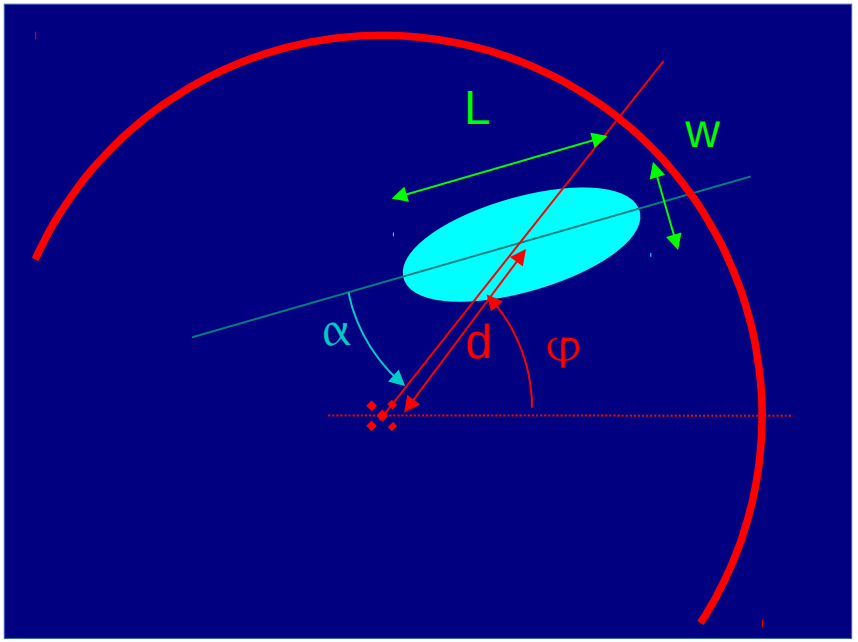
Published (FITS)



Reconstruction

- **Hillas Parameters** (1984)
  - Shower images are **elliptical**
  - A **few** parameters:
    - Length (L) & Width (W)
    - Nominal Distance (d)
    - Azimuthal angle (φ)
    - Orientation angle (α)
    - Amplitude (size)
  - Additional parameters: asymmetry, …
- **Advanced methods**
  - 3D model
  - Fit to simulated images



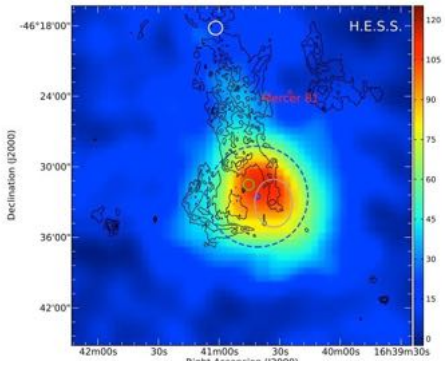


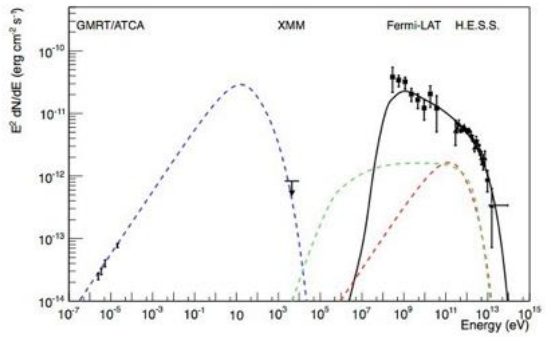
Analysis: producing images, spectra, light-curves

- Event lists (not necessarily gamma-rays!)
- Instrument Response Function
- TECH tables: atmosphere and data-quality measurements, → good time intervals, ...

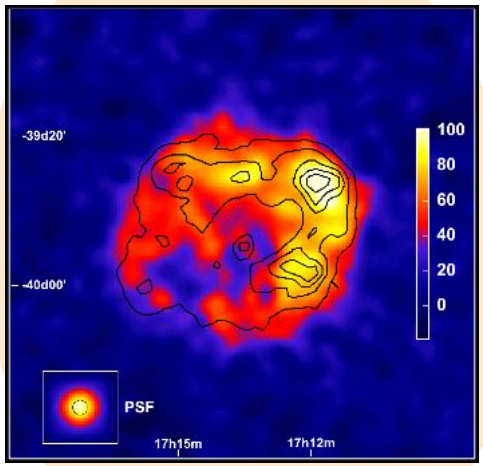
t	RA	Dec	az	alt	E
501	128.2	-37.1	321.1	45	0.112
600	130.23	-36.2	322.1	45.1	0.242
620	120.124	-33.33	312	45.7	0.434
640	121.1	-34.1	323.4	43.0	35.0
720	123.23	-28.1	322.1	43.5	0.401
900	100.0	-31.2	32.2.	45.3	1.23
..	..	..		..	..
..	..	..		..	..

- **Background dominated**
  - Even after stereo-reconstruction, hadron rejection is not 100% efficient
  - Therefore we can only talk **statistically** about gamma rays!
  - Use **likelihood** methods: model background, source assumption, Poisson statistics...
- **Source confusion**
  - Large PSF
  - Many sources are not point-like



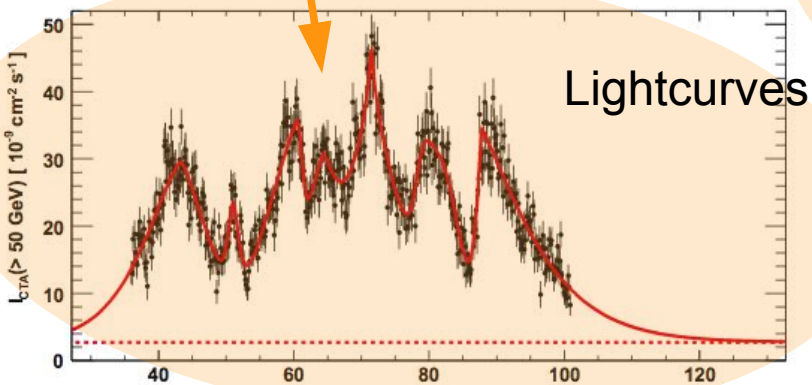


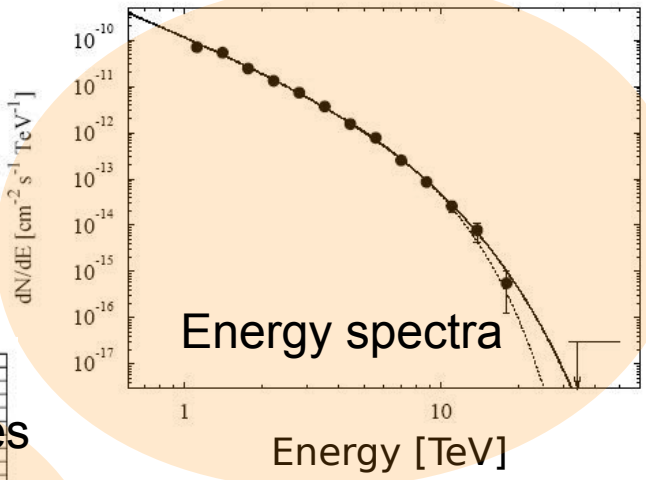
Multi-wavelength analysis

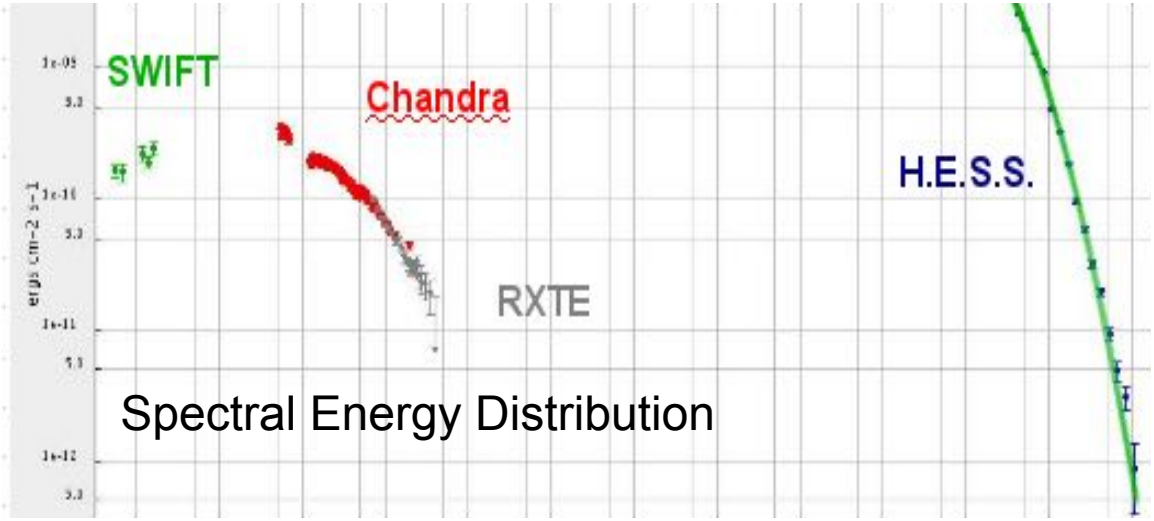


Images

Event lists
(coordinates, time, energy)





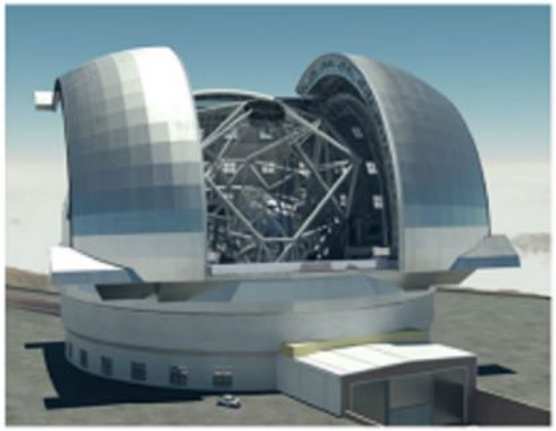
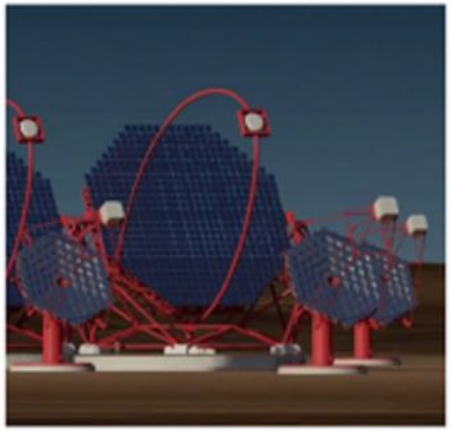
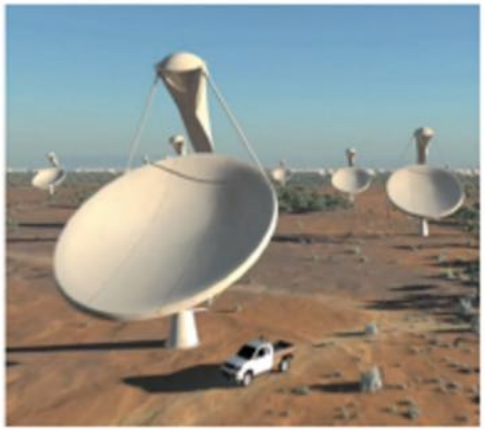
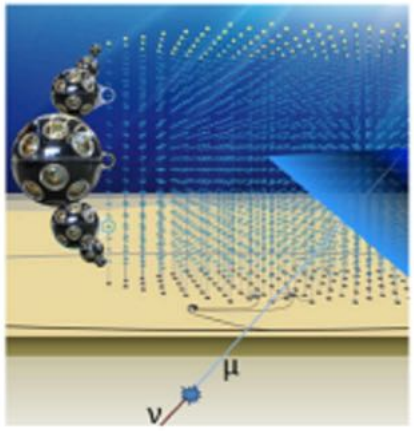


Compatible data at other wavelength?

Simultaneous
Calibrated
Specific Processing?
Context?

Data diffusion for CTA

- European Commission's Horizon 2020 framework
- ASTERICS aims to address the cross-cutting synergies and common challenges shared by the various Astronomy ESFRI facilities (E-ELT, **CTA**, SKA & KM3NeT)
- Work Package 4: Data **Access** and Data **Interoperability** (DADI)

Where CTA meets VO

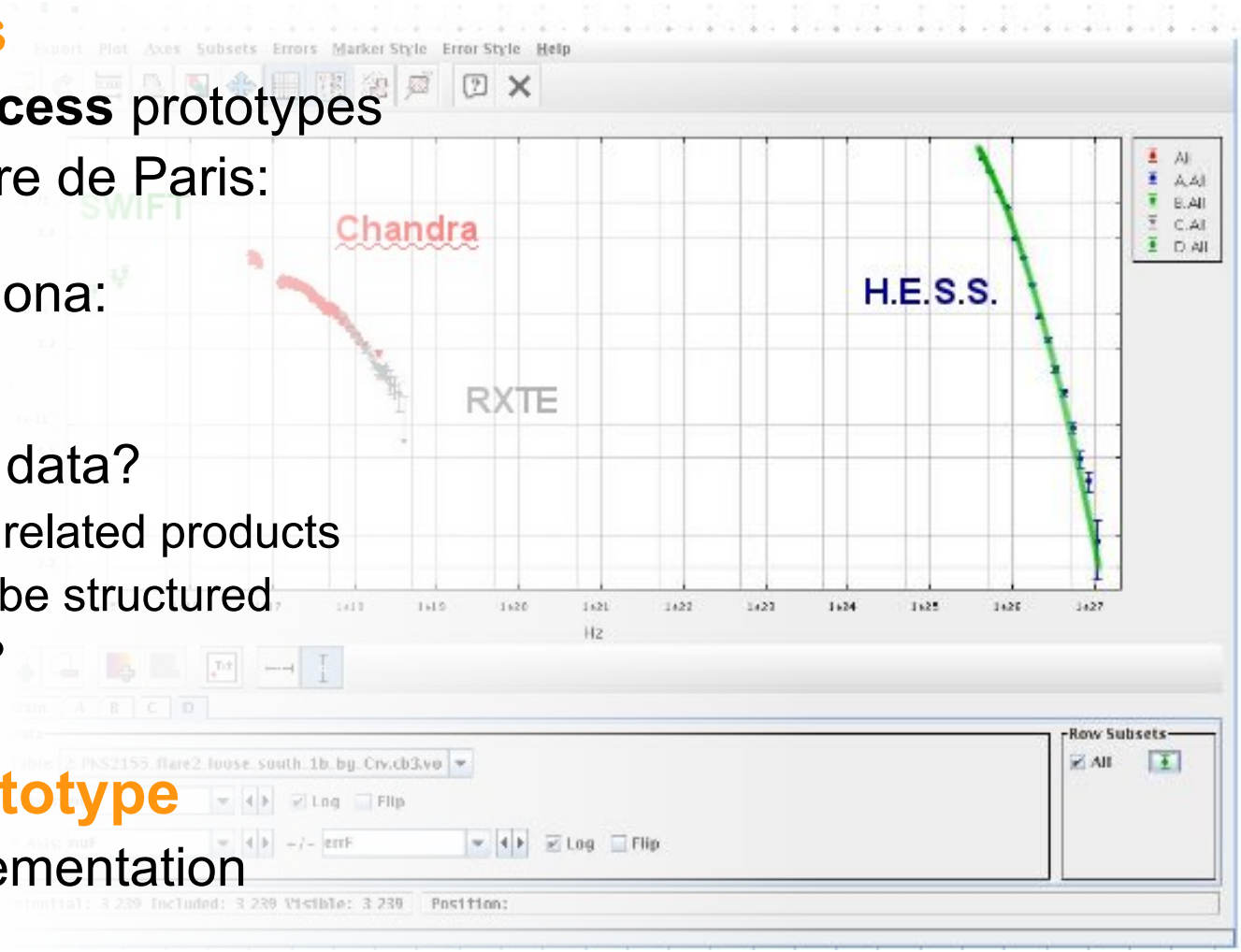
- **CTA Data Model**
  - Compatibility with **DatasetDM**, **ObsCoreDM**…
  - Development of **ProvenanceDM**
- **Data Diffusion**
  - Compatibility with VO protocols (e.g. **TAP**)
- **Data Processing**
  - Use of a work cluster with **UWS** pattern
  - Store/retrieve **Provenance** metadata
- **Data Access**
  - **Authentication** & **Authorization** solutions

VHE Data Access

Initial developments

- High level **VO data access** prototypes
- H.E.S.S at Observatoire de Paris:
  http://hess.obspm.fr/
- MAGIC at IFAE Barcelona:
  http://vobs.magic.pic.es/

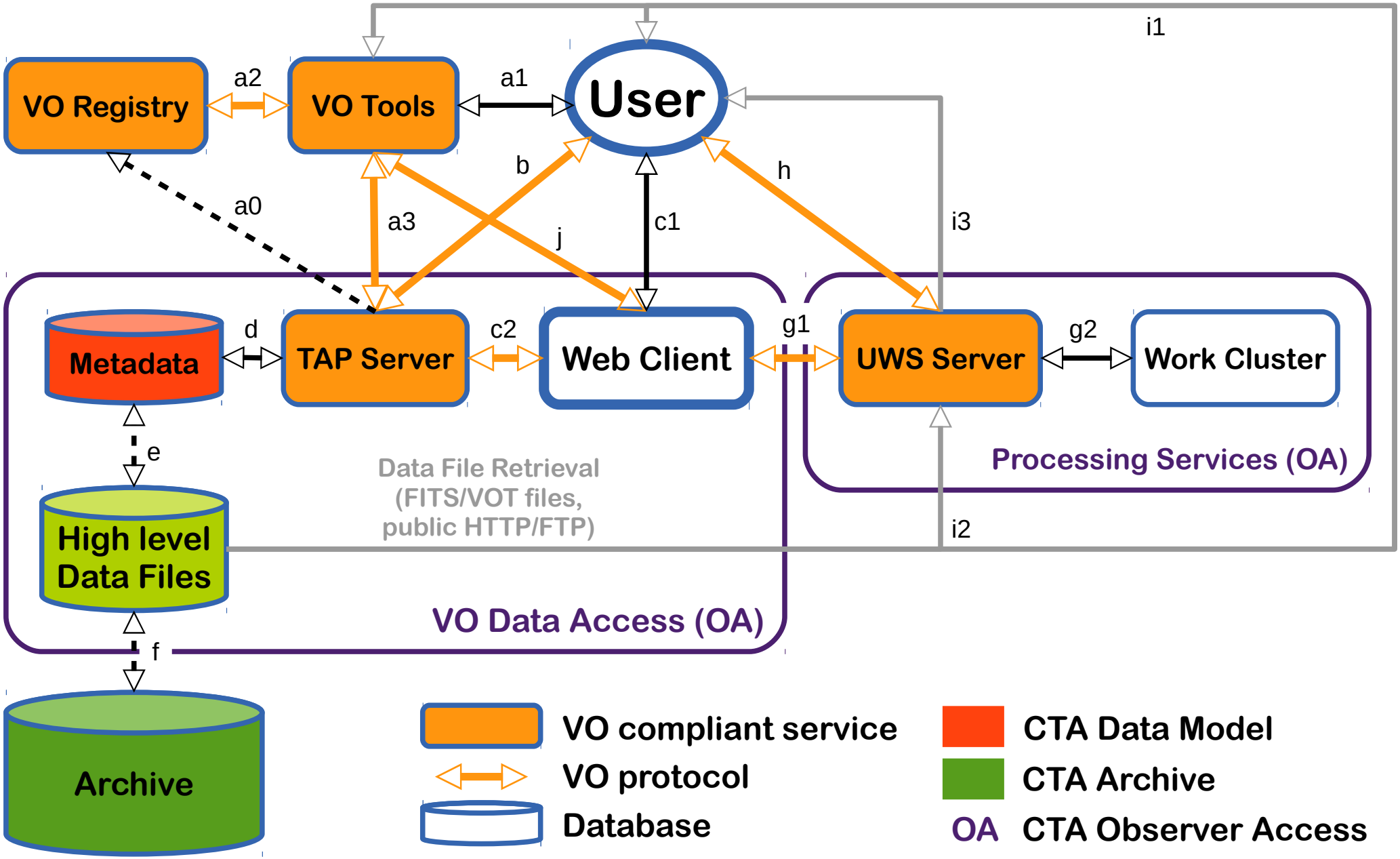
- VO standards vs VHE data?
  - Complex **hierarchy** of related products
  - Complex **metadata** to be structured
  - **Queryable** metadata ?



CTA data access prototype

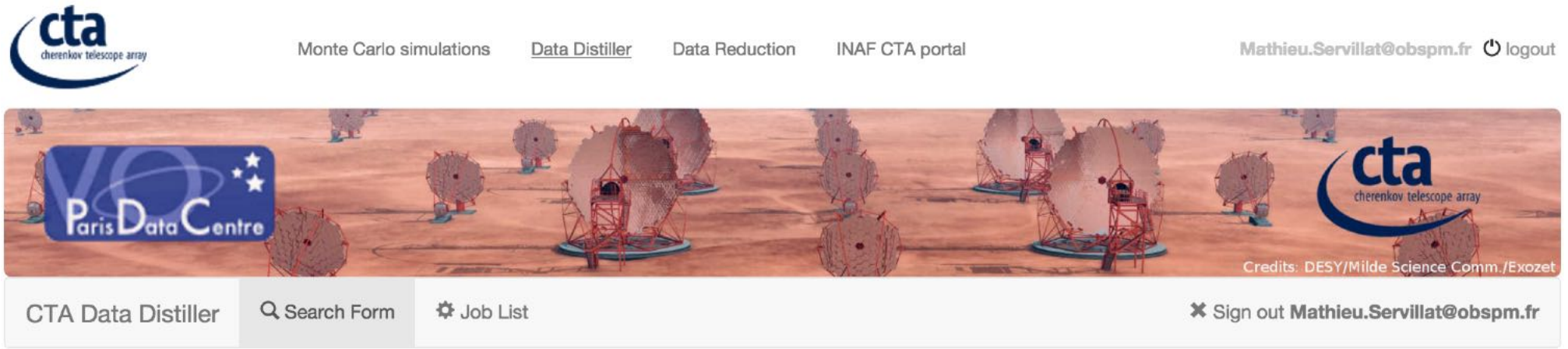
- CTA **data model** implementation
- Test VO compliance
  http://voparis-cta-test.obspm.fr

VO data diffusion prototype



Web client

http://voparis-cta-test.obspm.fr



- Django, jQuery, BootStrap3
- Name resolver
  Simbad through Sesame
- Builds and Sends the ADQL query

Web client

http://voparis-cta-test.obspm.fr

CTA Data Distiller | Search Datasets | Results | Job List | Selected Job | JS9 | Authentication: Sign out **user**

Search | Analyse | Visualisation

SAMP

Interop (SAMP)

Send Result Table

Results

show/hide query

SELECT * FROM cta.vo_obscore as o WHERE 1 = intersects(o.s_region, circle('ICRS', 83.63308333, 22.0145, 0.001))

Send

ADQL query

ObsCore fields

Search

UWS

Analysis tools

Create Count Map(s)

	dataproduct_type	obs_collection	obs_id	target_name	s_ra (deg)	s_dec (deg)
☐	eventlist	1	23592	Crab Nebula	82.01333618164062	22.01444435119629
☐	eventlist	1	23559	Crab Nebula	85.25333404541016	22.01444435119629
☐	eventlist	1	23526	Crab Nebula	83.63333129882812	22.51444435119629
☐	eventlist	1	23523	Crab Nebula	83.63333129882812	21.51444435119629
☐	eventlist	3	5003499	CrabNebula	83.28087615966797	21.784133911132812

Showing 1 to 5 of 10 rows 5 records per page

<< < 1 2 > >>

Plotting tools

TOPCAT

Aladin

VOSpec

SPLAT

Data mining use cases for CTA

Use case	Description
Cone Search	**Search data available for a given Target**
ObsCore search	**Search data available corresponding to ObsCore keywords** (target_name, time interval, ...), e.g.: • search data for a given **target** at a given **time** • search data in a given **region of the sky** • search data that contain events at **energy** higher than 50 TeV
ObsCore optional search	**Search data available corresponding to ObsCore optional keywords** (target_class, data_rights, ...), e.g.: • search **public** data for all **blazars** • search data for a given **proposal_id**
ObsConfig search	**Search data available corresponding to ObsConfig keywords** (sub_array_name, pointing_mode, obs_mode ...), e.g.: • search data that include the **Large Size Telescopes** (LSTs) • search data for a given target, that do not include the divergent **pointing mode**
Provenance search	**Search data available corresponding to Provenance keywords** (calib_version, creation_date ...), e.g.: • search data produced by a given **version of the pipeline** and for a given target • search data produced using a given **reconstruction method** • search data for a given target produced with **loose cuts**

ObsCore fields for CTA

dataproduct_type: has to be one of the following: image, cube, spectrum, sed, timeseries, visibility, event. Set to "event" in the prototype, has it exposes the 1DC DL3 files.

calib_level: one of the following integer values: 0 (instrumental or raw data in a non-standard/proprietary format), 1 (instrumental data in a standard format, e.g. FITS), 2 (calibrated data in standard format, with instrument signature removed), and 3 (more highly processed data product). CTA defines 5 data level, for example DL3 data are calibrated data in scientific units but still include an instrument signature, hence its calib_level would be between 1 and 2.

access_url: to be defined by the Archive, however the CTA 1DC data should not be accessible to the public. We thus include simulated data hosted on `http://voplus.obspm.fr/cta/` and always point to this URL in the prototype. In the VO context, the access URL is generally a public link. To handle data rights, this may point to a retrieval system with the ID of the requested data product.

em_min, em_max: The spectral coordinates are in TeV for us and should be converted to meters to follow the ObsCore standard. This could lead to precision issues in spectral data (though it is not an issue for discovery purposes).

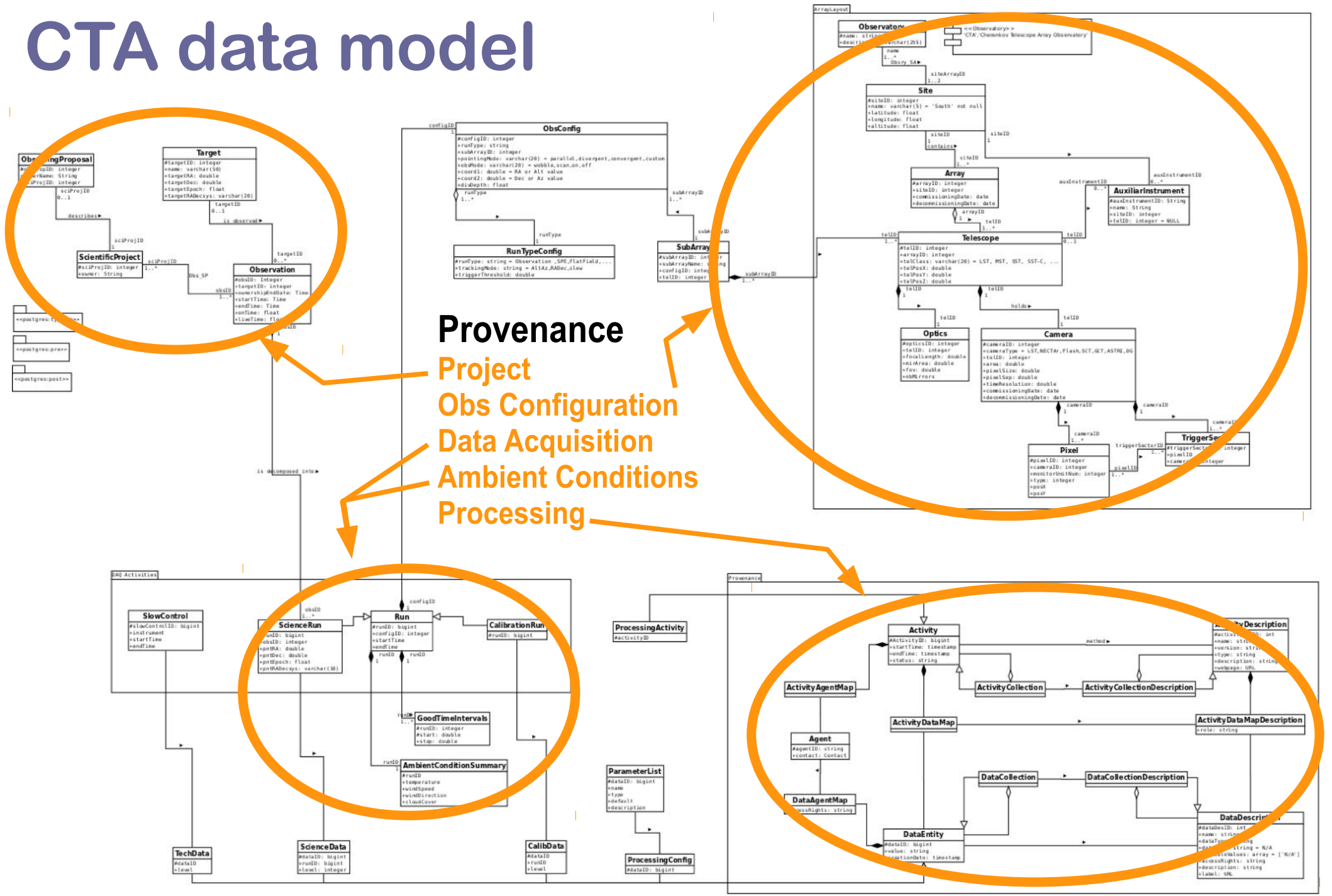
facility_name: we use the observatory name, e.g. "CTA".

instrument_name: As our test data comes from several experiments, we describe them here: HESS, MAGIC, VERITAS or CTOOLS (for simulated data with the ctools). This could be use to expose the CTA SubArray used to acquire the data?

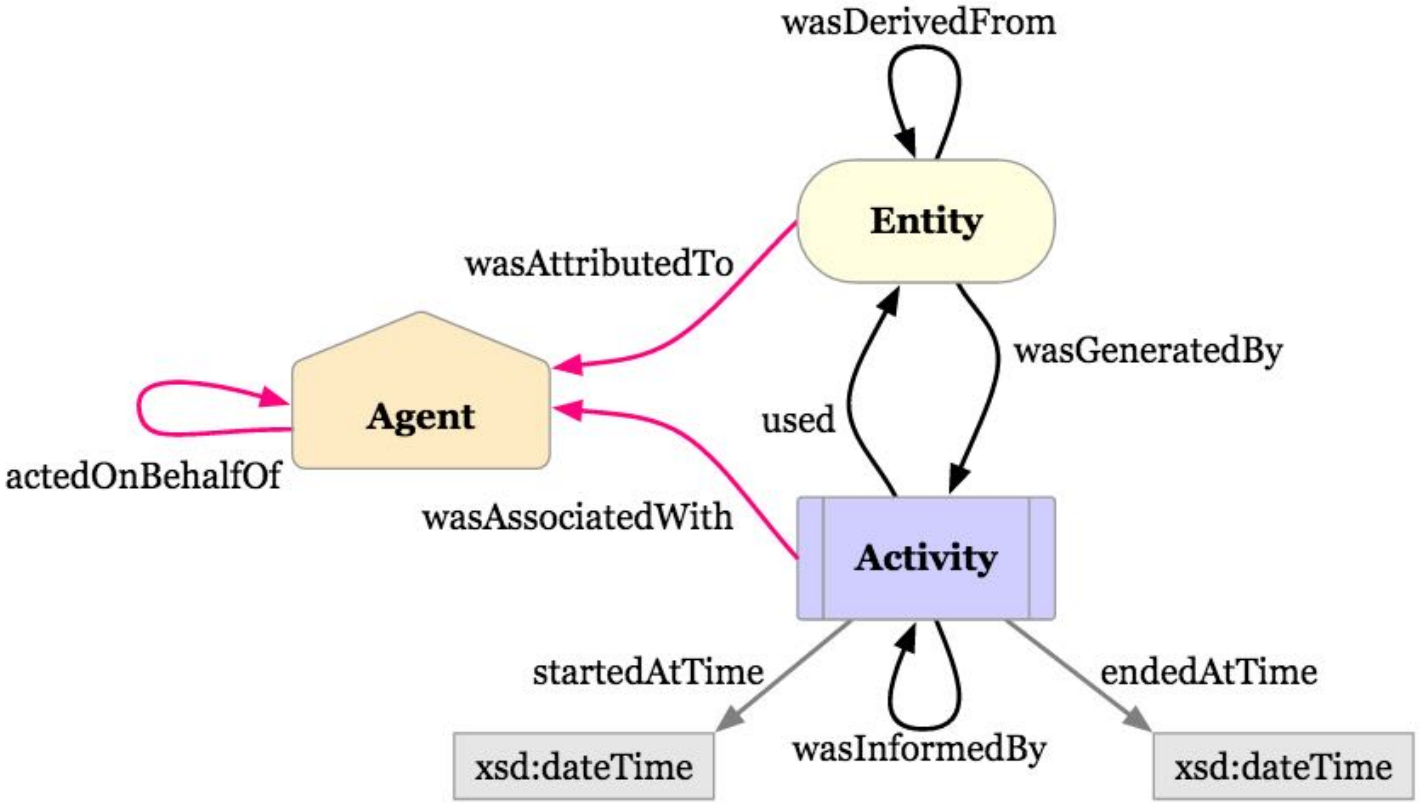
Extended ObsCore fields for CTA

- **Optional ObsCore fields:**
  - **dataproduct_subtype**: show DL0-5?
  - **obs_release_date**
  - **data_rights** (Public/Secure/Proprietary)
  - **s_resolution_min**, **s_resolution_max** (as it is dependent on energy)
  - **proposal_id**
- **ObsConfig:**
  - **site**: North or South site.
  - **sub_array_name** (or directly in instrument_name)
  - **pointing_mode**: parallel, divergent, convergent, custom...
  - **obs_mode**: wobble, scan, on, off
  - **run_type**: flatfield, science, SPE…
- **Provenance:**
  - **data_quality**: flag giving information on the data quality
  - **calib_version**: version of the calibration stage of the Pipeline
  - **reco_version**: version of the reconstruction stage of the Pipeline
  - **reco_method**: reconstruction method used to obtain DL2 data
  - **applied_cuts**: selection criteria used to obtain e.g. a DL3 photon event list
  - **spectral_model**: spectral model assumed to obtain spectrum

CTA data model

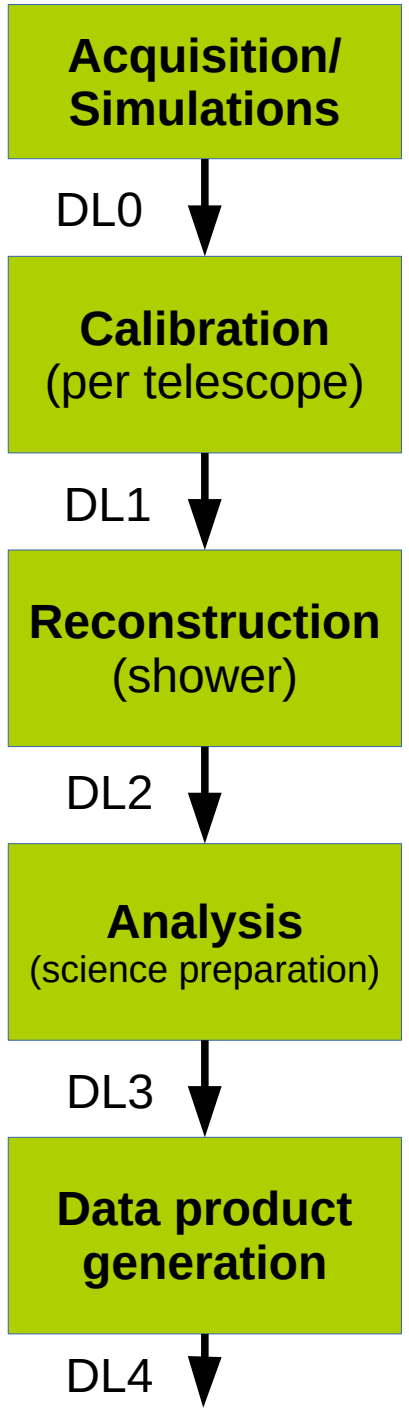


W3C Provenance model

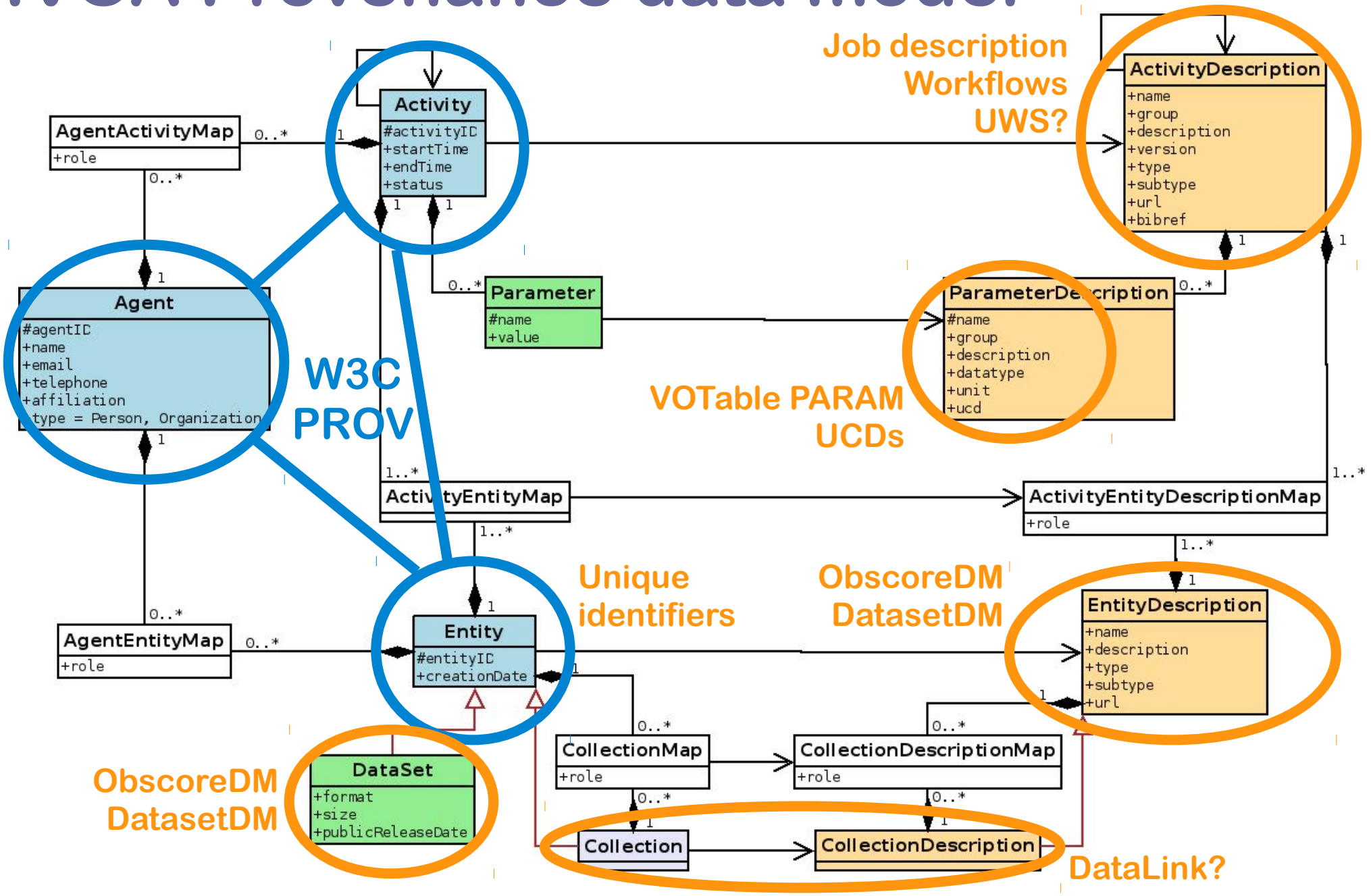


W3C PROV Ontology

https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/



IVOA Provenance data model



Online processing (client)

CTA Data Distiller | Search Form | Job List | Sign out **Mathieu Servillat**

Job List

Refresh Job List | Create Test Job | Job list loaded

Type	Start Time	Destruction Time	Phase	Details	Control
ctbin	2016-04-12 15:08:00	2016-05-12 15:07:30	COMPLETED	Properties Parameters Results	Start Abort Delete
ctbin	2016-01-28 16:06:29	2016-02-27 16:06:29	COMPLETED	Properties Parameters Results	Start Abort Delete
ctbin	2016-01-28 14:57:32	2016-02-27 14:57:32	ERROR	Properties Parameters Results	Start Abort Delete
ctbin	2015-11-05 15:11:09	2015-12-05 15:11:09	COMPLETED	Properties Parameters Results	Start Abort Delete
ctbin	2015-11-05 15:11:06	2015-12-05 15:11:06	PENDING	Properties Parameters Results	Start Abort Delete
ctbin	2015-11-05 14:20:18	2015-12-05 14:20:18	ABORTED	Properties Parameters Results	Start Abort Delete

- **Asynchronous** management of processes
- Job sent on a **work cluster**
- JavaScript library developped at PADC

Online processing (server)

Main features

- IVOA **standard**
  - Universal Worker System (UWS)
- **REST** architecture
  - Python micro-framework: bottle.py
- **Collaborative** development
  - Git server at PADC (gitolite)
  - GitHub: https://github.com/ParisAstronomicalDataCentre/OPUS
- **Tests** and **quality**
  - Unit tests with `unittest` and `webtest`
  - Activity history with `logging`

Prototype available

https://voparis-uws-test.obspm.fr

From UWS to Provenance

UWS Server | Job Definition | Job Manager | Sign out **admin**

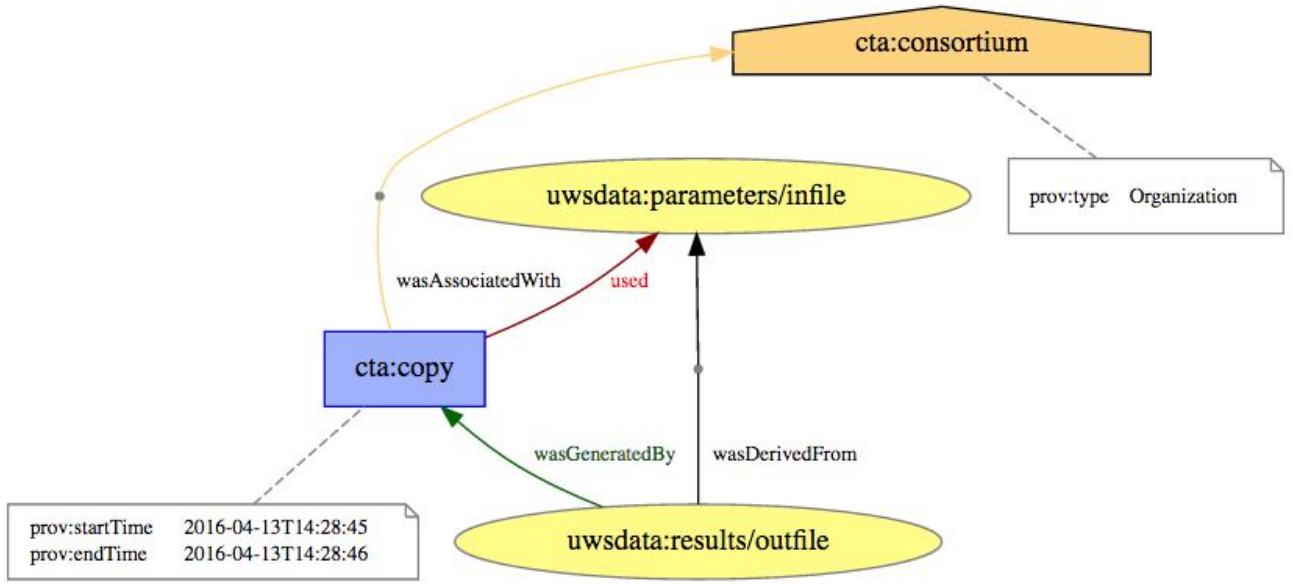
Job Description — Back to job list

Type	Start Time	Destruction Time	Phase	Details	Control
copy	2016-04-13 14:28:45	2016-05-13 14:23:39	COMPLETED	Properties | Parameters | Results	Start | Abort | Delete

- Job Properties
- Job Parameters
- Job Results
- Job Details

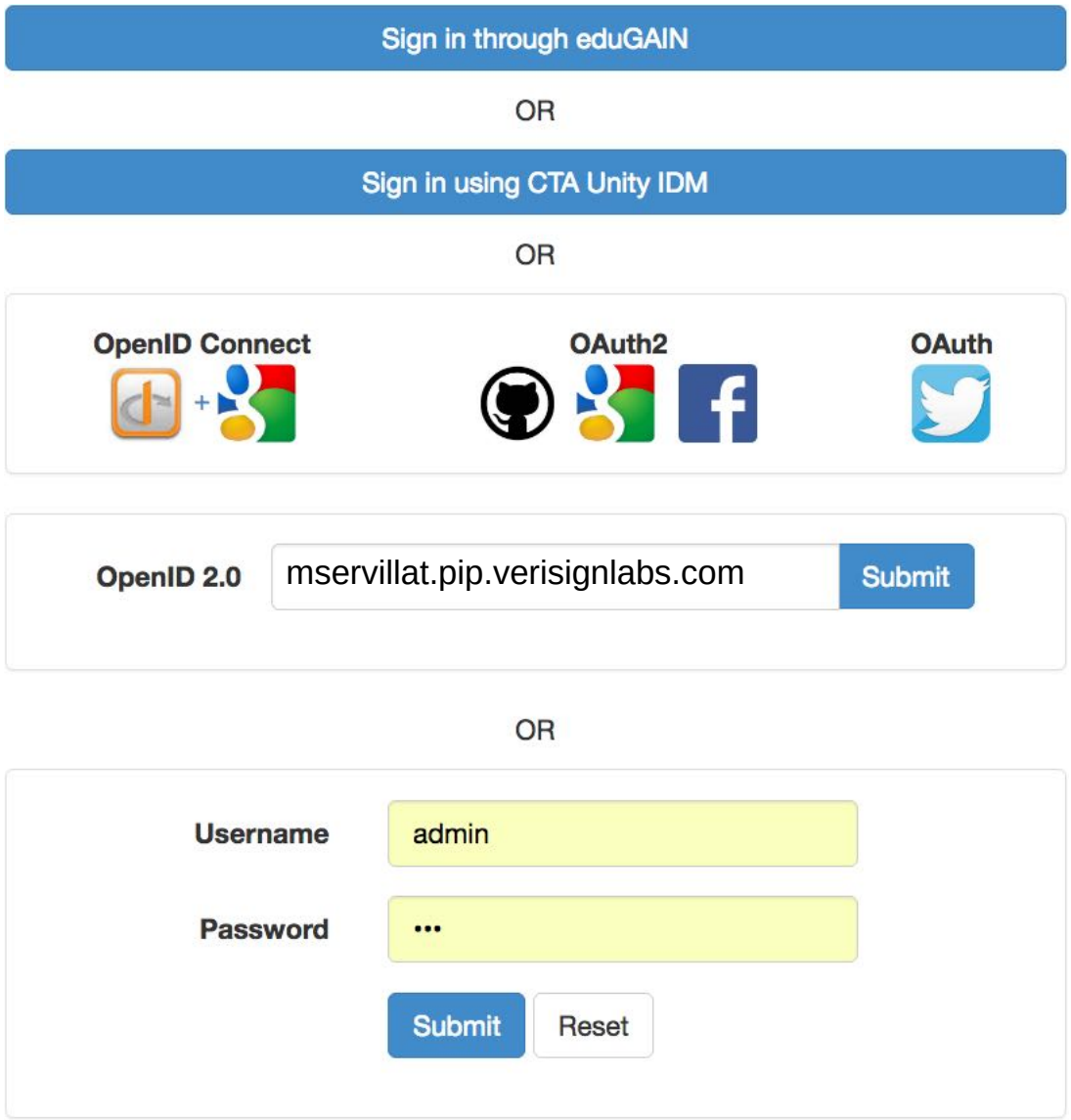
provxml: https://voparis-uws-test.obspm.fr/get_result_file/bc3ac123-82a0-4036-9d06-9880cc196f4f/provxml/provenance.xml

provsvg: https://voparis-uws-test.obspm.fr/get_result_file/bc3ac123-82a0-4036-9d06-9880cc196f4f/provsvg/provenance.svg



provjson: https://voparis-uws-test.obspm.fr/get_result_file/bc3ac123-82a0-4036-9d06-9880cc196f4f/provjson/provenance.json

Authentication & Authorization



- Shibboleth+Grouper
  - EduGAIN federation
  - SAML2
- Unity IDM
  - Uses OpenID Connect
- OpenID Connect
  - Google as an IdP
- OAuth2
  - Github, Google, Facebook, ...
- OAuth
  - Twitter, ...
- OpenID 2.0 (deprecated)
- Local account

Summary

- **CTA: the Cherenkov Telescope Array**
  - An **open** observatory for the TeV domain
  - For the **first time**!
  - **Compatibility** with existing VO standard?
  - **Evolution** of current standards and **new** ones!
- **Related IVOA working groups:**
  - **DM**: Provenance data model
  - **GWS**: UWS pattern, A&A solutions
  - Semantics, DAL, Applications…
- **ASTERICS H2020 project**
  - Meetings, schools, working groups on those topics