IVOA Metadata for high energy astrophysics

WP4 ESCAPE preparatory meeting
How does high energy metadata connects with IVOA standard data models?
https://indico.in2p3.fr/event/24652/
Friday 25 Jun 2021, 09:30 → 17:30 Europe/Paris
LUTH - Observatoire de Paris, Meudon

Participants:

- Catherine Boisson
- François Bonnarel
- Bruno Khélifi
- Mireille Louys
- Mathieu Servillat

Joined via Zoom at 15h:

- Matthias Fuessling
- Gareth Hughes

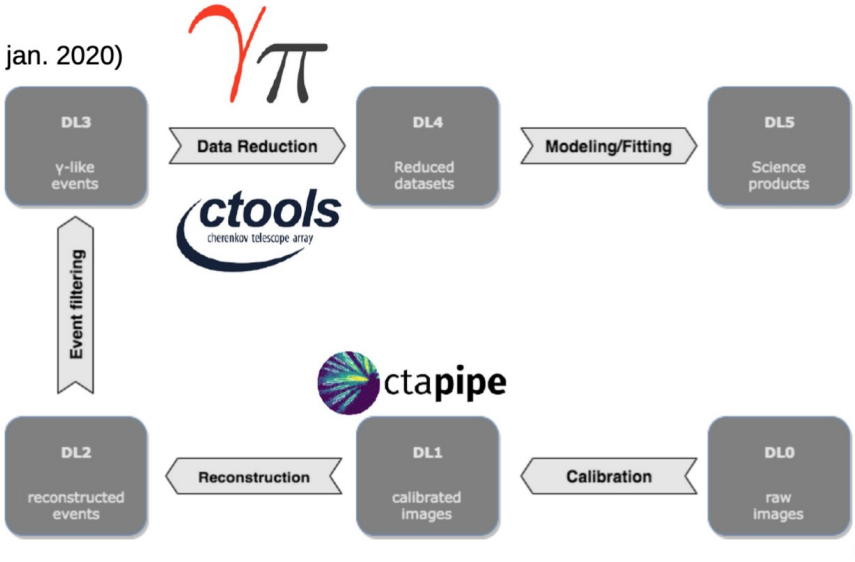
1/ Metadata for Cherenkov astronomy

Bruno:

- Data format
    - inspired from Fermi LAT and adapted
- Fixed observation duration (25-30 min runs for H.E.S.S.)
- Data Levels
    - DL0 : camera data
    - DL1 : calibrated signals from cameras
    - DL2->DL3 : reconstructions and cuts
        - different options (reco method, applied cuts)
        - 1 file with the detected events
        - 1 set of files for the instrument response functions(IRF) corresponding to the observation

Archive – data level

CTA terminology
(from Top-level Data Model jan. 2020)



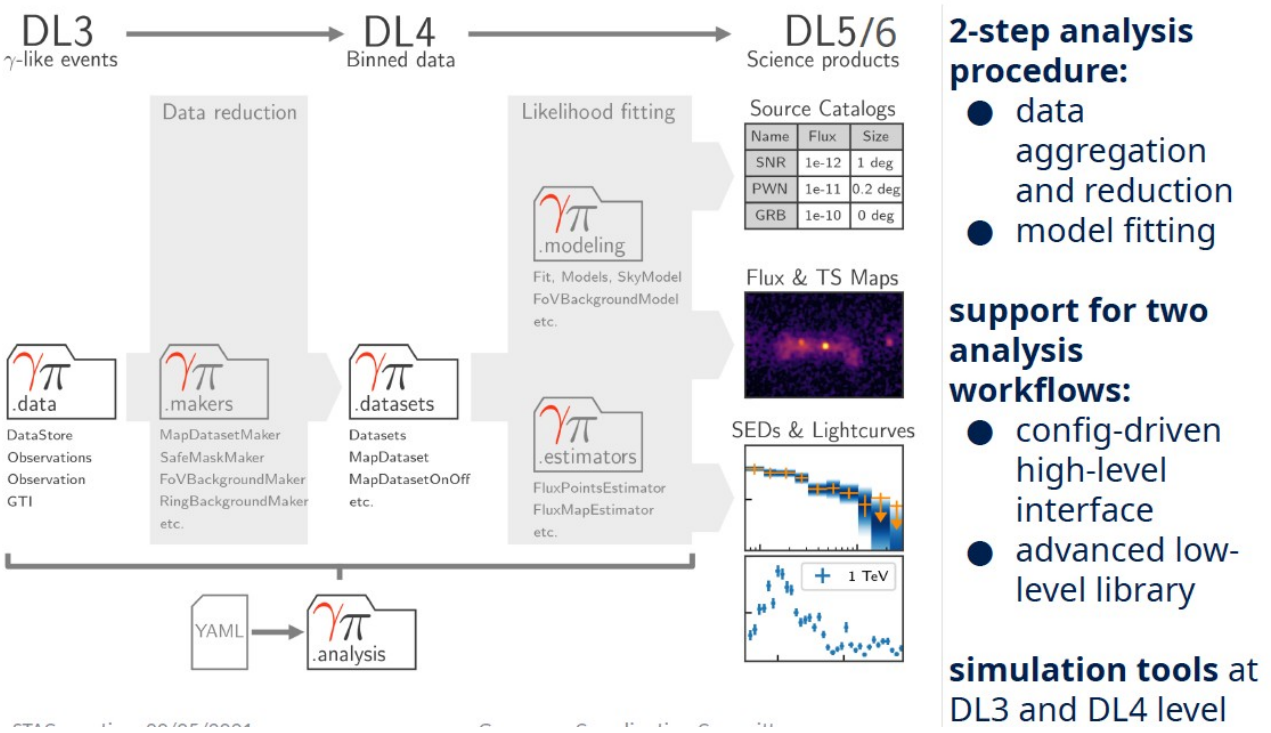
H.E.S.S.

2

- File with DL3 in FITS format
  - EVENTS frame (time, sky position, energy)
  - Additional frames
    - TELARRAY: list of telescopes (e.g. 4 small H.E.S.S. telescopes)
    - GTI: Good Time Intervals (may be just one line, start and stop)
    - POINTING: time stamp with pointing position in different coordinate systems
  - Some metadata
    - obs_id (runid)
    - telescope type, list
    - start, stop times
    - mean pointing
    - ...
- IRF files
  - From simulations:
    - aeff: effective area
    - edisp: energy dispersion
    - psf: point spread function (after reconstruction)
  - From observations:
    - bkg : estimation of the hadronic background noise
- Index of observations :
  - Collections are stored on disk for gammapy ("DataStore")
  - obs-index et hdu-index tables
  - Contain characteristics for requests and data selection

    - Remarks :
        - If some data are downloaded from an archive, obs-index and hdu-index should be specifically generated for this data subset, and all fiels packed together (e.g. tar)
        - Separate datasets in distinct directories
        - Should we provide a way to combine several obs-index and hdu-index?
- Total DL3 Archive DL3
    - 12 GB for H.E.S.S. (but for 1 type of analysis)
    - x10 for CTA
- DL3 -> DL4 -> DL5/6
    - gammapy (official tool for CTA)
    - provides a specific workflow
    - DL4-DL5 formats are not yet defined
    - DL6: catalogs, but scope is not clear (per project catalogs, full catalogs with releases...)

Data workflow and package structure



- Notebook example:
    - Spectral analysis of the Crab with gammapy
    - 1D_systematics_Crab_he
    - See also https://docs.gammapy.org/0.18.2/tutorials/index.html

2/ VO standards

- Relevant standards :
  - MOC
    - HEALPix Multi-Order Coverage map
    - https://www.ivoa.net/documents/MOC/
  - ObsCore
    - Observation Data Model Core Components
    - https://www.ivoa.net/documents/ObsCore/
  - DataLink
    - Linking of data discovery metadata to access to the data itself, further detailed metadata, related resources, and to services that perform operations on the data
    - https://www.ivoa.net/documents/DataLink/
  - SODA
    - Server-side Operations for Data Access
    - https://www.ivoa.net/documents/SODA/
  - TAP
    - Table Access Protocol
    - https://www.ivoa.net/documents/TAP/
  - ADQL
    - Astronomical Data Query Language
    - https://www.ivoa.net/documents/ADQL/
- Example at CADC with Aladin
  - Cone search on the image
  - List of ObsCore records in the selected region
  - access_url = Datalink, with several options proposed (via a 'drop down' list)
    - preview
    - fits
    - cutout service
- ASKAP with Aladin
  - MOC = coverage on the sky
  - Cone Search -> ObsCore table
  - access = Datalink, including a cutout service SODA
  - SODA
    - small image overlapped on the sky map, with the requested size and binning
    - useful if main data file is too large (100s of MB)
- Radio Visibilities with TAP Handle
  - TAP Request (using ADQL)
  - Datalink
  - There may be similarities in the handling of radio visibilities and high energy IRF

3/ VO for Cherenkov data

Data Access and ObsCore fields content

- access_url
    - Is there an interest in having juste the event file (no IRF)
        - FITS is more interoperable than tar
        - but no science or preview can be easily done with the event list of one observation (except for the Crab maybe...)
    - Concatenate in 1 FITS file only ?
        - FITS is more interoperable than tar
        - may not be better anyway...
    - GTI (Good Time Intervals)
        - There may be several IRF for one observation: one per GTI
    - by default, a tar file with events + IRF
        - structure of directories in tar file (to respect gammapy DataStore)
        - add obs-index and hdu-index for the files accessed
    - Several options shown as a Datalink
        - TAR of all files
        - Each file
        - Service to show preview, processing capabilities...
- content_type
    - application/x-fits+tar+gz (to be checked)
    -
- dataproduct_type
    - "event": event counting, typically (ObsCore definition)
    - How can we indicate the presence of IRFs ?
    - Use dataproduct_subtype ?
    - Datalink for access_url
- calib_level
    - between 2 and 3 (see ObsCore definition)
    - EVENTS: instrumental effects are not completely removed, but data is calibrated
    - EVENTS+IRF: instrumental effects removed, but after some processing
    - DL3 -> calib_level = 2
    - DL5 -> calib_level = 3
- obs_collection
    - could distinguish between different types of analysis
    - ex : hess_dl3_dr1
- obs_publisher_did
    - example for hess_dl3_dr1:
        - <ResourceID>
          <AuthorityID>padc</AuthorityID>
          <ResourceKey>hess_dl3_dr1/23523</ResourceKey>
          </ResourceID>
        - ivo://padc/hess_dl3_dr1/23523

- s_ra, s_dec
    - coordinate of the mean pointing (not always the target)
- s_fov
    - circle with the mean radius of the field of view ?
    - compute the distance of the most external event ?
    - sensitivity map contours ?

Use cases for data access

- Download DL3 data
    - All (experts would have all the data on disk, disk space is not a constraint)
    - Subset
    - in all cases, include obs-index and hdu-index to make it usable with gammapy
- Preview the data
    - Show each event ? no real interest
    - Preview of DL5 data corresponding to DL3 data
        - pre-generated, with specific options (reco method, cuts)
        - SODA: request a map with the relevant options
- Download DL5 data (reverse of the previous use case)
    - direct access to image, spectrum, lightcurve, SED...
    - datalink to #progenitor : DL3 with IRF in FITS.tar.gz
- See what is prepared for SVOM (L. Michel)
- See what X-ray observatories do
- See what Radio observatories do

Possible ObsCore extensions for DL4/5 data

- t_support (T-MOC format ?)
- snr
- binning_strategy
- model_fit_name
- model_fit_url
- model_fit_minimiser_method
- ...

MOC and HiPS

- Such products are relevant for a given collection

Prototype: proposed evolutions

- HESS_DR via ObsTAP
    - https://hess-dr.obspm.fr/
- Update ObsCore metadata
- Should provide a tar file with all IRF
- Should include obs-index and hdu-index for the selected obs

- The tar file would thus be directly readable by gammapy
- Datalink
    - Complete DL3+IRF (default)
    - DL5 preview for this observation
    - Provenance (last_step_provenance record? ProvSAP?)
- Additional DL5 may be found in another ObsTAP service (or another collection on a service)
    - relevant DL5 are generally a combination of several DL3 (wobble, increased exposure time, several nights)
    - thus the identifiers are different (but mapping DL5 -> DL3 is kept)
    - Service to locate DL5 proposed via the Datalink service
    - or reverse system : find DL3 from DL5 -> table for mapping