Notes for Meudon IACT DL3 meeting

- Meeting documents:
  https://github.com/open-gamma-ray-astro/2016-04_IACT_DL3_Meeting
- Attendees: C. Boisson, K. Brügge, J.L. Contreras, C. Deil, D. Dorner, T. Hassan, B. Khelifi, J. Knödlseder, S. Lombardi, G. Maier, G. Pedaletti, J. Rosado, M. Servillat, R. Terrier, R. Walter,

These notes were taken collaboratively by the meeting participants during the meeting.
A cleaned-up version and summary will be posted to the meeting Github repo later.

Action items are highlighted in red color.

Table of contents:

Observations, GTIs, EVENT-IRF association

Background information (discussion on Github
https://github.com/open-gamma-ray-astro/gamma-astro-data-formats/issues/20#issuecomment-173151492

- OBS_ID = run number for current IACTs
  - HESS and VERITAS: 28 min
  - MAGIC: 15-20 min (flexible)
  - FACT: 5 min
- Array config and pointing position is fixed for a given OBS_ID
- Roland: this definition of observation is confusing to astronomers.
  Most astronomers mean multi-hour exposure as one observation.
- What we have now in HESS (what's currently in the spec)
  - IRF association via OBS_ID
  - One OBS_ID = one consecutive GTI = one IRF
  - IRF association is done via HDU_INDEX table

- What VERITAS does:
  - Multiple GTI per OBS_ID
  - One file = one GTI (for a given OBS_ID the data is split across multiple files)
  - Store EVENTS, GTI and IRF in one file
    (no index files, association by being in the same file)
- Proposals to do better:
  - Change terminology: "observation block" = "run" instead of "observation"
  - Basic unit of analysis for science tools is one GTI
  - OBS_ID, GTI_ID, EVENT_TYPE -> IRF
  - EVENT_TYPE = as used by Fermi-LAT, meaning event quality class.
  - EVENTS:
    - Agreement: use EVENT_CLASS as separate things / files (like Fermi-LAT)
      "Standard" or "hard" cuts
    - Agreement: use EVENT_TYPE int column in event list in a similar way Fermi-LAT does
    - How do we support electron and proton analyses?
      - Unclear if this needs to be supported via DL3 / science tools.
      - Needs particular science tools to do this properly.
      - To be investigated: what additions to the DL3 spec would be needed to support this.
      - Maybe define a dedicated event class for this, like "ELECTRON" or "PROTON"
      - Action item (Jose-Luis, next week) will write a section describing how it's possible with the current info in DL3 and make pull request
  - Jürgen:
    - GTI should be used for quality or user-selected time bins
    - IRF validity time ranges should be separate from GTI.
      Call them TIME. ("TIME_LO" & "TIME_HI" columns)
  - IRF association is done via HDU index table
    - Keys: OBS_ID, GTI_ID, EVENT_TYPE -> IRF filename and HDU name
  - IRF storage could be:
    - Multiple columns in the same HDU
      This would mean having columns like
      - AEFF_TYPE3_GTI4
    - Higher-dimensional arrays for AEFF column
      - Discrete axis where index 0 = (GTI 0, TYPE 0)
    - Multiple HDUs
  - Concerning options for FITS serialisation of time- and event-type dependence of IRFs Jürgen will write down a proposal and then we'll re-discuss.
    (if there's no index files, there might be a second proposal from Christoph)
- Supported observation modes

- Currently fixed pointing and deadtime fraction per event list (key in the header)
- Agreement that we should change the DL3 spec to have a pointing table.
  - This is a simple change for DL3 data producers,
    But implies a big change for science tools (because background models are centered on the pointing position, need time-stepping scheme for binned analysis)
  - See Karl's 2011 document as a starting point … move this over to the current spec.
    - Finally … some TECH3!
    - Which info from the current EVENTS header should be moved over?
    - POINTING table with columns TSTART, TSTOP, RA, DEC, DEADC
      ALT, AZ for the array (this value is used as reference point for IRFs)
  - This might be all that's needed to support important use cases:
    - HAWC
    - IACT convergent pointing and drift scan observations
    - Fermi-LAT?
- Jürgen is proposing a concrete more flexible scheme:
  - EVENTS with multiple GTI (as before)
  - POINTING, IRF with time binning that is completely independent of the GTIs
  - Association of EVENTS to POINTING and IRF done via TIME
  - EVENT_TYPE as extra discrete dimension in existing IRF column.
    Time would also be an extra axis in the IRF with binning TSTART, TSTOP.
    (should we define this as RTI = response time interval so that we can better talk about this?)
  - File-based scheme … the basic unit is one EVENTS table in one FILE.
  - Long discussion about what the motivations are to introduce this scheme (efficiency of IRF filesize, … other ...?) and if it's more or less complex.
    - The main new concept here is that extra time chunkings are introduced. GTI are done for data quality or user selection. The other time bins are used for IRF lookup.
  - For science analysis the IRF time ranges would be the basic unit of processing (not GTI)?
  - The alternative proposal to this scheme is to have GTI_ID which can be used to look up the right IRF. (i.e. for DL3 producers, GTIs are chosen so that it's a good time binning for GTIs)
  - Needs detailed writeup …
    - Why is this extra set of time intervals needed?
    - How does GTI splitting or EVENTS / GTI merging work?

- Provenance information that should be put in DL3
  (even if not required for science tool analyses):
    - TELARRAY, TELMASK, TELID,
    - This should all be optional, i.e. science tools should just access what is really needed for science analysis.
    - Action item (next week, pull request): Catherine and Mathieu will make a first proposal what IACTs should put as header keys (e.g. OBSERVATORY, software, version, chain, config, …)

- Index tables
- Observation definition

Event quality class
- Event type
- Event class

Basic definitions

- FOV coordinates
  - Christoph: proposes to add FOV coordinates as columns FOV_ALTAZ_* and FOV_RADEC_*
  - Jürgen and Roland: use WCS for this?
  - OGIP says DETX, DETY is detector coordinates and X, Y are the projected event position on the sky.
  - What exactly is DETX, DETY?
    - Agreement to make it the ALTAZ
  - Christoph has concerns about using WCS for event lists and IRFs.
    Extra complexity in exporters and science tools.
    Jürgen and Roland think this is a solved problem and we should use the WCS solution.
  - Action item (by next telcon): describe in detail how both solutions work. Re-discuss soon.
    Jürgen, Christoph, Catherine, Alexander will start Github issue to discuss (Monday)
  - Should contact Bill Pence (author of CFITSIO) for comments once we have a concrete proposal on this questions. ADASS meeting in Trieste soon, Catherine and Mathieu might talk to people about issues there.
- Time
  - OGIP standard should be used everywhere (also in observation table)
    I.e. use reference time and TIMESYS.
    Science tools should use TIMESYS correctly.

- ○ Two systems have been proposed
    - ■ TT (Terrestial Time)
    - ■ UTC
- ○ There was a discussion if the TIMESYS in the IACT DL3 spec should be:
    - ■ Required to be TT
    - ■ Recommended to be TT
    - ■ No recommendation given.
      DL3 producers and consumers must respect it.
- ○ Action item: "IACT data storage" should be more clearly labeled as optional, a proposal to organise data. Archive should be encouraged to join the telcons.
    - ■ Make it more clear that index files are optional and in the user domain.
    - ■ Source models and high-level results sections should be split out into a separate document.
    - ■ Jürgen wants to have IACT data storage and OGIP in a section "other stuff", the core of this document is the section on events and IRFs.
- ○ Action item (Jürgen, next week): clean up current event list spec page.
- ○ Concerning the MET zero point: every IACT can choose their own.

IRFs

Serialization

- ● We propose to use adopt the FITS BINTABLE format for now for DL3.
  (If issues arise, we can reconsider that decision and move to a different serialisation).
- ● Action items (people are encouraged to pick this up):
    - ○ Is the factorisation into PSF * AEFF * EDISP good enough?
      (or are there strong correlations that events in the PSF tail have worse EDISP)
    - ○ Could you encode different factorisations in the current format?
    - ○ How large do IRFs get (compared to EVENTS) in the current format if we want to achieve the CTA precision requirements?
        - ■ Christoph: provide some numbers for HESS
        - ■ Tarek: provide first estimate for CTA
    - ○ Do we try to encode IRF error information?
        - ■ How to serialise that info?
        - ■ Is it useful? (i.e. how would it be used by science tools?)
    - ○ Jürgen -- CALDB standard to declare calibration validity ranges (equivalent of safe thresholds we use no) -- might replace THRES_LO, ...
- ● Safe thresholds (e.g. safe energy minimal threshold or maximum FOV offset)
    - ○ Should be computed by CTA pipeline or science tools?
        - ■ Agreement that it's good to give safe threshold info as part of DL3 IRFs.

- Science tools should use it by default, but can provide methods to compute other thresholds or the user can even ignore it if they like (e.g. for detection, not fluxes)
  - How to better propagate IRF error and safe threshold info via DL3 IRFs?
(the options are not exclusive, can be done at the same time)
    - Option 1: more header keys or extra tables (e.g. RAD_MAX and as a function of energy or LO_THRES/HI_THRES as a function of offset)
      - Action item: review what is there now and extend to cover current use cases.
    - Option 2: give an error array (e.g. effective area relative error at a given energy and offset) and then science tools compute safe thresholds from that.
      - Action item: put it as an option for the future now, don't fully spec out.
    - Option 3: give multiple IRF HDU (or modify given ones) so that people can run multiple analyses and compute systematics from the spread in results (like bracketing IRFs in Fermi).
      - Action item: support access to multiple IRFs via HDU index table (HESS is already using this, exporting 3 different PSFs per obs)

Parameterization and factorisation

- See Jürgens presentation on an idea for clever AEFF factorisation
- Needs the experience of existing IACT experts or people working on MC studies.
- No volunteers for this at this meeting. We encourage such studies.

Spec action items

- Action item for all IACT DL3 producers: review current spec and suggest improvements.
- Christoph: Move BINTABLE HDU description to IRF section
- Jürgen: Review axes names / complete?
- Tarek: clean up terminology "format" -> maybe "configuration"
- Jürgen: Propose better naming format for HDU_CLASS "formats" (like aeff_2d)
- Time header keys should follow OGIP standards
(TSTART_STR -> DATE_OBS and TIME_OBS and same for end)

Process

- Action item: Catherine will check if the CTA-internal similar documents from 2011 and 2013 can be put online (for reference of what we're doing now).
- Action item: Jose-Luis will cross-link the CTA-internal DL3 page and this meeting page and the open spec.

- There is agreement that we will work on Github in the open (i.e. not parallel to the same efforts in CTA).
- Jürgen: this should be "proposals", not "spec".
  Conclusion of discussion: they are the same … "proposals for specification" … "working draft".
- Conclusion: stick with one repo / one document approach.
- Action items:
  - for each page, add a box at the top describing maturity / agreement status.
  - Add overview page listing status of all sub-pages.
  - Feedback / discussion on small issues / pull requests happens on Github.
  - Summaries of important controversial questions (like use database on user machine or not) should be put in the document.
- Decision making process:
  - The process described here starts in about 1 month, after making a stable version of what is there now (Christoph wants to clean up the current document and make a version that can be used for HESS).
  - All changes to the repo shall go through pull requests (not commit directly to master).
  - A template will be added for pull requests that describes the change.
  - Christoph: it's important that proposed changes don't sit in pull requests for months. They will not appear in the online document and very few people will see and prototype them.
    For this reason I propose that some proposed changes are merged with a clear label "Provisional, might be removed soon", and then decisions on this are taken before making a stable version (like v2.0) via a RFC period and ultimately a decision by the board.
  - Monthly open telcons
  - Data format decisions are made by a board.
    - Action item: Catherine organises the formation of the board.
    - Should have one representative from each instrument (CTA, HESS, MAGIC, VERITAS, FACT) (Catherine also wants to have a person from Archive, Jürgen disagrees, CTA should have one voice.)
    - The representative from each instrument
    - Question: DL3 consumers (science tools) as stakeholders?
- Scope of this spec:
  - Tarek: MC parameters / DL2 in scope?
- Describe process clearly in the spec
- Action item Christoph: rename "Background information" to "context"
- Action item Christoph: put version "dev" in HDU_VERS for dev versions of the spec.
- Pull requests go in without much review, informal
- Stable versions are circulated in CTA-DATA and open-gamma-astro mailing list with some request for comments process (like PEPs for Python or RFC for FITS changes)

- Example files are important. For 1.0 there shall be one for every format (e.g. aeff_2d, psf_king, …)
- Proposal : ask Bill Pence whether it would be worth to have this spec become an OGIP standard some day (say in a year or a few years once its stable).

- Discussed data challenge.
  HESS will release something in ~ 2 months.
  Agree that it's too soon to do something larger / collaboratively between instruments.
  Maybe in fall or next year this could be done, together with version 2.0 of the spec?

Codes, Reference files

- Produce reference files how?
  - Both contributed files from other codes are welcome.
  - And scripts to produce reference files in this repo are welcome also.
  - Action items:
    - Christoph can put files from HESS for formats we have.
    - Jürgen is volunteering to provide reference files.
- Collaborate on I/O library useful?
  - Some discussion … conclusion is that maybe yes.
  - Tarek is interesting in developing something.
    Christoph suggests a single-header file approach that is super-easy to use in exporters. Could contribute the HESS HAP exporter as a starting point.
  - Kai would like Python implementation or a python wrapper around the C stuff
    This was discussed yesterday by Tarek and Christoph … probably we'll put something in gammapy.irf, but we don't have time in the coming 2 weeks.